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

Scientometrics,
Vol. 44, No. 3 (1999) 323–345

INDICATORS IN A RESEARCH INSTITUTE: A MULTI-LEVEL CLASSIFICATION OF SCIENTIFIC JOURNALS

E. BASSECOULARD,[1] M. ZITT[1,2]

[1]*LERECO, INRA, BP 71627, F-44316 Nantes Cedex 03 (France)*
[2]*Observatoire des Sciences et des Techniques (OST), 93 rue de Vaugirard, F-75006 Paris (France)*

(Received January 13, 1999)

Indicators in a research Institute ought to be readable at several decision levels, and particularly with different break-downs of the publication set chosen as reference. Citation transactions between journals have been widely used to structure scientific subfields in ISI databases. We tried a seed-free structuration of SCI/CMCI journals (a) to test convergence of pure citation-built specialties (roughly 150) on SCI/CMCI journals with existing classifications at the subfield level (b) to explore the interest and the limits of this approach for upper levels of aggregation (roughly 30 fields). A few limits of journal-level classification are addressed. At the subfield level, the convergence is large with some discrepancies worth noticing. At the subdiscipline level, the method is not sufficient to achieve a satisfactory 30-level delineation, but gives a good basis for informed expert validation.

Introduction

The present work has been elaborated as a part of a joint INRA/OST project aiming at the design of science output indicators for INRA, the French National Institute of Agronomic Research, one of the world largest public research organisations in the agronomic sector. Academic output represents only one dimension of its activity. However it is an important one and INRA was to be positioned in international science.* As citation studies were foreseen, we used at first ISI sources SCI and CMCI (social sciences have been excluded, because of the various shortcomings of SSCI).

---

*INRA missions include academic "finalised" research, decision support and expertise, innovation and technology transfer, diffusion of knowledge, scientific training etc. To prevent misuses of indicators in term of "productivity", one should not forget the variety of ouputs in an Applied Research Institute. Besides, the representativity of SCI/CMCI for INRA academic activity is another key point: an estimation of "what is left out" requires a combination of expert advices and scientometric characterisations.

0138–9130/99/US $ 15.00

For comparisons in national or international perspectives we chose to use a "universal" grid, likely to add some information to "in-house" knowledge. The breakdown into broad academic disciplines (such as the 8-level pattern* used by OST for macro-indicators) is little controversial. In our case a more precise and operational view was looked for: methodology of indicators had to be adapted to users at various decision levels. A classical granulometry, with ca. 150 subfields and a reaggregation into ca. 30 fields or "subdisciplines", seemed a reasonable target. A first attempt to aggregate existing ISI subject categories** was disappointing. To avoid black-box effects or biases induced by a-priori settings we tried a full "seed-free" reconstruction on the whole SCI-CMCI set, as INRA activities integrate many subfields. We started from the journal level and used inter-journal citation transactions, so that the aggregation process could be explicit and controlled. At the end of the process, the compatibility with ISI classification was addressed.

In this paper, advantages and limits of the protocol are discussed. Section I outlines the general context, section II describes the clustering method, section III shows examples of results and discussion. Acronyms of institutions are detailed in an appendix.

Context

Science is often viewed as a self-organizing system submitted to various irreversibilities and chaotic changes. In practical contexts such as the building of indicators, classification of science is a recurrent question.[1] Broad academic disciplines are still a common grid suitable for macro-studies but at a high degree of generality. At lower levels, no detailed science classification is commonly accepted. Academic research, governments, international institutions or private macro-indicator producers elaborate their own classifications. Disciplinary nomenclatures exist in specialised databases (Chemical Abstracts, Inspec, Compendex, etc.), usually at the document

---

*Fundamental Biology, Medical Research, Applied Biology-Ecology, Chemistry, Physics, Earth and Space Sciences, Engineering, Mathematics; Multidisciplinary; the general pattern goes back to CHI research work for NSF with further adaptations.[24] The breakdown of SCI-CMCI into sub-areas, a seemingly purely technical issue, has obvious political implications in an evaluation or decision-support context : a high performance of an actor on a field A can be completely obscured when aggregating A and another field B; more generally, different grid designs may depict an actor's activity as specialised or diversified. In both examples, the image of the actor may be deeply modified.

**Bibliometric aggregation through inter-categories citations amplified some delineation problems observed at the subfield level: ISI classification uses « a combination of journal-journal citation patterns, keyword analysis and user feedback » (*Katz*[5] quoting Henri Small, ISI Philadelphia).

level. These databases are well known by individual researchers and allow further enriched treatments in specific fields.[2] But harmonisation problems are severe if the activity spectrum is broad (in the extreme whole science). Another option is to rely on informetric studies of science networks in multidisciplinary databases such as ISI-SCI. The purpose of informetricians, far from the positivist criteria *à la* Auguste Comte, is rather to obtain detailed and temporary frameworks, reflecting scientists' "collective" view of science. Some trade-off between the minimal stability needed for longitudinal indicators and the proteiform and chaotic nature of scientific advances is unavoidable.

Classification of SCI for indicators

The most usual current classifications applied to SCI are based on journals lists aggregates, using some mix of informetric analysis (citation between journals) and other approaches, with expert validation. The best known basic systems are Subject Category Classes defined by ISI (currently 170-level), and CHI-NSF pattern (roughly 120-level). Some of the CHI subfields and ISI subject categories are close to each other and often bear the same name. However the way these classifications keep pace with the changing structure of science strongly differs. CHI works on a constant journal set: short-term comparativity is favoured for science indicators, the drift over years being managed by deep long-interval updates (the last one being on 1993 data). ISI manages an annual turn-over, with poor short-term comparativity for derived indicators but natural adaptation in the long run. The way the two systems take care of large-spectrum journals also differs: both patterns include a multidisciplinary subfield, and lower level disciplinary classes (general biology, etc.); but in addition subject category classes allow overlaps, while new CHI classification does not (ISI also developed other classifications, such as Current Contents non-overlapping classes).

Either patterns (ISI/CHI) have been used as starting points by indicators producers or S&T observatories in various countries, sometimes with particular aggregations at the discipline level: for example FHG-ISI, that proposed an intermediary 27-level[3] inspired from ISSRU works after CHI/NSF classification. Another disciplinary setting, based on subject category classes, has been proposed by SRI.[4] The SPRU team, starting from a discussion of the Australian Standard Research Classification (ASRC), reaggregated ISI subject categories into a hierarchical system taking ISI multiassignments of journals into account, with 3 superdisciplines* – natural sciences, life sciences, applied sciences – and their overlaps, inter-field and multi-disciplinary.[5] In the Australian Research Evaluation and Policy Project, *Bourke* and *Butler*[6] combined this pattern with the basic structure of the ASRC. Lately ISSRU[7] proposed an improvement of ISI's classification with an assigment of individual documents in multidisciplinary and general journals.**

Many methods have been used by informetricians to structure scientific activity. They differ in several respects (a) basic data (document, journal) (b) starting point: databases classification and co-classification, co-activity of institutions, citation transactions... (c) statistical methods used to reveal the underlying structures.

Citation based methods

Journals remain the main substrate for macro-level classifications. Moreover such nomenclatures are easy to use in practical contexts. Citation-based structuration methods root in the skew distribution of citations patterns and the sparsity of citation matrices. As for documents, citation exchanges between journals are likely to reflect hierarchies, levels of application, thematic proximities. Since the pioneering work of CHI Research on journal-to-journal links[8] and clustering,[9] citation transactions have been widely used to position journals in journal networks[10] or group them by a variety of methods (for a review see Ref. 11), from "quasi-correspondence" analysis[12] to graph approach for small sets like the clique detection by *Burton.*[13] *Leydersdorff* and *Cozzens*[14] particularly investigated continuous methods, based on iterative factor analysis for uncovering latent structures and the dynamics of the system, for example through the resistance of new journals to previous classification patterns.

However the strong power of inter-journal citations is not without limits. First, though the aggregation at the journal level has a smoothing effect, usual warnings about citation uses apply by and large to inter-journal transactions. The cognitive and sociological nature of citation networks (for recent surveys see *Luukkonen*[15] and the topical issue of *Scientometrics*, September 1998) must be taken into account. In practice, the multicriteria synthesis offered by citation linkages is often disconcerting for users (researchers, policy makers etc.), who are unfamiliar with scientometrics and

---

*A former SPRU work[25] proposed a research classification in 7 disciplines for exact sciences (including multidisciplinary and psychology) and 3 for social sciences and humanities. The 7 exact sciences were splitted into about 35 titles and more than 250 illustrative specialties. At high level, all physical sciences were grouped, geosciences were part of environmental sciences and all life sciences were together.

**ISI regularly realises a similar process, but limited to *Nature*, *Science*, *PNAS*, and with *Current Contents* classification codes.

accustomed to one-dimension typologies (for instance, when one defines a medical specialty after the biological system, or the pathology, or the method).

Then, classification and scale problems are closely related. Several authors have stressed scale-invariance phenomena in science networks, attributing it to self-organization mechanisms. The scale-invariance problems in science mapping are intricate because statistical metric options are largely arbitrary. *Van Raan*[16] stressed, in its study of "fractal dimension" of ISI co-citation fronts, that it may be difficult to disentangle artefacts and results, when a classification algorithm is involved.

Last but not least, possible artefacts have to be dealt with:

- *general journals*: exceptions can be found but a majority of journals exhibits a very skewed and consistant topic distribution and easily respond to small-grain classification. General journals should be excluded from the core of the classification process for two related reasons: they create massive spurious linkages jeopardizing fine-grain classes, and they somewhat artificially favour high-level disciplinary groupings. But, with the exception of well-known "multidisciplinary journals", it may not be easy to determine which journal should be considered as "general" at its disciplinary level. A kindred problem sometimes appears with journals dealing with pervasive techniques (in genetics or biotechnology for instance), linked to many applications. We chose to discard general journals in biology and medicine mainly after their qualification in ISI classification; a similar treatment could not be operated for physical science, with some consequences on the results (see below). The discriminating power of citation linkages obviously depends on the field, according to the importance and treatment of general journals.
- *classification artefacts*: the outcomes of classifications are highly dependent on the initial metric first, then on the aggregation method (partition/ hierarchical, hierarchical descending/ ascending, static/ dynamic algorithms etc.). A huge literature is devoted to classification in statistical sources[17] and many applications are reported in scientometrics and information retrieval. Some methods are biased towards particular cluster characteristics (equal size or variance), others have adverse effects (chaining, excessive sensitivity, etc). In addition statistical significance of partitions is a controversial subject. Divergences between methods can be severe, and robustness is to be favoured. Interpretation must take technical choices into account (for instance, in average linkage, a close contact between specialties through a few journals is not enough to make those specialties close to each other). Some specific effects of ascending algorithms are particularly visible when approaching the root of the tree: the

concentration of items in the major cluster(s) for uncorrected single linkage is a well-known example. The fluctuations of weak citation linkages, that may determine some unstable high-level arrangements, remain another source of concern at large scale.

In spite of the drawbacks of discontinuous methods stressed by *Leydesdorff*[18] – some of them can be coped with – hierarchical approaches remain appropriate on theoretical grounds (the embedded nature of science networks) and on practical ones (a multi-level framework for various levels of decision).

Clustering methodology

Sources and protocol

For this first study, the source of data for transactions is the *Journal Citation Report* (JCR) 1993. The classification process was carried out independently of previous nomenclatures, except for prior eliminations of multidisciplinary journals (see below). We used the average of annual citations in the window 1990-93. This range is an acceptable trade-off for an all science approach, with a mix of short and medium term citation behavior. Impact figures used for preselection of journals are taken from OST on primary ISI sources (Integrated Citation File extract, ICF). For a journal, impact is calculated here on the most citable types of documents (articles, notes, reviews, letters) in the citation window. Taking into account the constraints of JCR of longitudinal management of journals, unification of journal names between JCR and ICF was necessary in many cases.

The protocol comprises 3 steps (1) the selection of a highly-cited subset of SCI-CMCI journals (2) the building of an appropriate similarity matrix of these "core" journals and its hierarchical clustering into roughly 150 specialties (3) the re-assignment of all journals to these categories. Technicalities of stages 2 and 3 resemble much those developed for document-cocitation by *Zitt* and *Bassecoulard.*[19]

1) *Preselection of core cited journals: a "seed-free" process, based on locally normalized impact.* We started from the citation matrix of the *Journal Citation Report* 1993. It is useful to classify only higher impact journals to lower the noise level. However, as the citation behaviour varies among fields,[20] a direct selection on absolute impacts would lead to a distortion in favour of highly cited fields. On the other hand using a preexisting classification (e.g. ISI subject categories) to normalize impacts would bias final classes towards this pattern. Fortunately JCR data provide internal means of normalisation: a journal impact can be normalised "locally", i.e., after the average impact of its neighbours. Neighbours have been detected by a similarity index

(see below) and up to 30 neighbours have been used for normalisation. Roughly 2,000 journals with higher locally normalised impacts have been selected. They capture a very large part of all citation transactions. The above process provides a seed-free selection for the clustering. The only exception is the compulsory elimination of multidisciplinary and general journals from the first stage of classification. We used existing lists of multidisciplinary journals (mainly ISI) and created three multidisciplinary blocks: general, general biology, general medicine. General journals may be reassigned in a second stage, using their preferential linkages.

2) *Similarity matrix and first classification of the core.* Various approaches may be defined to explore the intellectual framework revealed by the citation linkages. For instance, the distance (or dissimilarity) between journals may be assessed according to their direct transactions. Such simple direct measures have been used for example by *Pudovkin*[21] on aquatic biology journals. An alternative to direct transaction analysis would be to work one step further, on citation patterns of A and B (cited and/or citing) to assess their proximity. The rationale is somewhat kindred either to co-citation or bibliographic coupling, with journal as macro-documents. This approach is powerful and may be used as a natural starting point for factoral analyses, with some difficulties for the treatment of journal self-citations appearing on the diagonal.* However, the distinct advantage of giving high similarity to items that do not communicate directly but are equivalent in the network, is less decisive when a classification follows up: in most cases these "quasi-synonyms" will be clustered together, without adding a risk of spurious synonymies and uneasy interpretations. For this reason only direct transactions have been used in this experiment.

Since we are looking for field identification rather than impact/influence stratification, a symmetrical view is adopted, taking into account both flows A–>B (noted AB) and B–>A (noted BA) for the proximity of journals A and B. Citations and references are restricted to classified journals: A_e represent the total of references of A to journals within the core, A_r the total of citations received by A from the core. It is pretty natural in this context to use similarities index. Since a two-way relationship is involved, symmetrical indexes are used with a normalisation by the total inter-journal transactions of both A and B (self-citations excluded). Symmetrical indexes take their maximum value in case of exclusive linkage of A and B for their external citations.

Ochiai form: $OCH(AB) = AB / sqrt(A_E * B_R)$

Symmetrical index: $SYMOCH(AB) = sqrt(OCH(AB) * OCH(BA))$

Jaccard form: $JAC(AB) = AB / (A_E + B_R - AB)$

Symmetrical index: $SYMJAC(AB) = (1/2)*(JAC(AB) + JAC(BA))$

---

*Several options have been proposed, for instance by *Price*[26], *Noma*[27] or *Tijssen*[12].

An ascending hierarchical classification was carried out on the symmetrical Ochiai index. The hierarchical process provides a consistant representation in a wide range of scale, even though partition at particular cutting levels may not be optimum. For clustering, we chose group average algorithm rather than extreme algorithms of single or complete linkage. Average linkage is generally considered as a low-risk method,* but it is slightly biased towards spherical equal-variance clusters. An example of comparison between average proximities and more local measures is given in Section III.

To delineate specialties, the hierarchical tree was cut off at a local optimum of stability, with a modal setting allowing peripherical items to join their next cluster. Central/peripherical position of each journal in its cluster is characterised using the tree structure, so that items can be ranked after their relevancy within a cluster.

3) *Final multi-assignment of citing journals.* The rationale of this second stage is twofold. It allows:

- the assignment of journals outside the core (i.e., the low impact journals) not classified in the first round, which are assigned to the cluster(s) with which they exchange most.
- the multi-assignment for all journals (core and not-core) as the mono-assignment carried out by most hierarchical algorithms** may appear as impoverishing.

Technically, symmetrical Ochiai and Jaccard indexes were used for this second stage. Jaccard index does not penalise unidirectional flows; it is useful to retrieve peripherical or recent journals that give many references to the core but do not receive much citations from it. Up to 3 assignments were allowed at the 150 subfields level.

The process minimizes some of the shortcomings of discontinuous approaches (classification): chain effects, non-overlapping assignment, sensitivity to thresholds, multidisciplinary problem. The two first ones are addressed by appropriate protocols. Threshold problems are of a very different nature: cut-off thresholds in the classification tree may be modulated to be robust and optimised. A perhaps more decisive option concerns the selection of core versus complementary items (journals), but it is not typical of discontinuous approaches. The treatment of multidisciplinary journals, the convergence between methods, are more fundamental ones – and it is dubious that any journal-based approach can be fully satisfactory from this point of view.

---

*See *Zitt* and *Bassecoulard* [19] in a bibliometric co-citationist context. Both single linkage and maximum linkage have shortcomings: chain effects in the first case, zero-transaction management in the second. A good alternative could be density clustering.

**With exceptions of "pyramidal" classifications. Several non-hierarchical algorithms allow overlaps.

Cluster documentation

An important additional problem that occurs when trying to make a constructive discussion with scientists-experts is the wording of these specialties. Cluster titles were automatically built after journal titles and submitted to experts. In some cases their terminology slightly differed from the automatic one and titles have been modified (e.g. biotechnology replaced by bioprocesses).

Table 1 gives an extract of cluster information in our classification. Core journals assigned to the cluster in the first stage are ranked after their position in the hierarchical tree, taking into account the level of aggregation and the subcluster size (Journal rank); the first two journals in the list are linked at the lowest level of the aggregation process for the cluster. In this type of ranking, the order relation between core-items is preserved throughout the aggregation process.

For each journal three other characteristic are recorded:

- possible multi-assignments (Cluster Rank); for instance 1.3 means that the journal is assigned in three specialties, and the present is the most relevant.
- contribution to the core (Primary Core); equals 1 for core journals clustered in the first stage and assigned first to the cluster, -1 for core journals assigned first to another specialty, 0 for non-core journals.
- sub-structures (Sub-Cluster); subclasses inside the cluster are numbered 1, 2, 3, etc. according to their size (number of journals). This shows what happens at a lower cut off level (roughly 300-level), and hence gives an idea of the homogeneity of cluster.

Let us give a few examples. Food Science is a small specialty of 50 journals, 33 exclusively belonging to it. The cluster clearly has 3 sub-classes: (1) General Food science, (2) Cereal and Carbohydrates and (3) Milk and Dairy. 4 of the first 25 core journals (assigned to Food Science in the first clustering stage) received a second assignment in the second stage: *Meat Science* was assigned to Animal Science, the *Glycoconjugate Journal* and the *International Journal of Biological Macromolecules* to Biochemistry/Molecular Biology, the *Journal of Carbohydrate Chemistry* to Chemistry. Conversely other core journals have been assigned first to other specialties, for instance Animal Science for the *Journal of Dairy Science*, Lipids for the *Journal of The American Oil Chemists Society*. Non-core journals are characterised in the same way.

Table 1
Extracts of clustering results for the specialty food science

Journal Rank	Journal title	Cluster Rank	Primary Core	Sub Cluster
1	*Journal of Food Science*	1.1	1	1
2	*Meat Science*	1.2	1	1
3	*Food Technology*	1.1	1	1
4	*Journal of the Science of Food and Agriculture*	1.1	1	1
5	*Food Chemistry*	1.1	1	1
6	*Journal of Agricultural and Food Chemistry*	1.1	1	1
7	*Zeitschrift fuer Lebensmittel-Untersuchung und -Forschung*	1.1	1	1
8	*Critical Reviews in Food Science and Nutrition*	1.1	1	1
9	*Cereal Chemistry*	1.1	1	2
10	*Journal of Cereal Science*	1.1	1	2
11	*Starch-Staerke*	1.1	1	2
12	*Milchwissenschaft-Milk Science International*	1.1	1	3
13	*Journal of Dairy Research*	1.1	1	3
14	*International Journal of Food Science and Technology*	1.1	1	1
15	*Lait*	1.1	1	3
16	*Carbohydrate Polymers*	1.1	1	2
17	*Food Hydrocolloids*	1.1	1	2
18	*Carbohydrate Research*	1.1	1	2
19	*Journal of Carbohydrate Chemistry*	1.2	1	2
20	*International Journal of Biological Macromolecules*	1.2	1	2
21	*Netherlands Milk and Dairy Journal*	1.1	1	3
22	*Glycoconjugate Journal*	1.2	1	2
23	*Journal of Texture Studies*	1.1	1	3
24	*Food Structure*	1.1	1	3
25	*American Journal of Enology and Viticulture*	1.1	1	1
26	*Journal of Dairy Science*	2.2	-1	3
27	*Cereal Foods World*	1.1	0	2
28	*Journal of Food Engineering*	1.1	0	1
29	*Journal of the American Oil Chemists Society*	2.2	-1	1
30	*Food Science and Technology-Lebensmittel-Wissenschaft und Technologie*	1.1	0	1
31	*Australian Journal of Dairy Technology*	1.1	0	3
32	*Nahrung-Food*	1.1	0	1
33	*Journal of the Institute of Brewing*	1.1	0	2
34	*Trends In Food Science and Technology*	1.1	0	1
35	*Food Research International*	1.1	0	1
36	*Journal of Food Biochemistry*	1.1	0	1
37	*Food Reviews International*	1.1	0	1
38	*Deutsche Lebensmittel-Rundschau*	1.1	0	1
39	*Journal of the Society of Dairy Technology*	1.1	0	3
40	*Journal of Aoac International*	2.2	0	1
41	*Acs Symposium Series*	2.3	0	1
42	*Bioscience Biotechnology and Biochemistry*	1.3	0	1
43	*Archiv fuer Lebensmittel Hygiene*	2.3	0	1
44	*Bioorganicheskaya Khimiya*	3.3	-1	2
45	*Agricultural and Biological Chemistry*	2.3	0	1
46	*Ecology of Food and Nutrition*	3.3	0	1
47	*Swedish Journal of Agricultural Research*	2.3	0	1
48	*Agribiological Research-Zeitschrift fuer Agrarbiologie Agrikulturchemie Oekologie*	3.3	0	1

Journals at the end of the list are generally more peripherical: the specialty Food Science appears as the second or third assignment for most of them. The last two journals have been discarded because of their very poor citation transactions.

Results and discussion

In this section, we will discuss the outcomes of the classification at two levels: 141 specialties, with 17 very small residual clusters, and 32 subdisciplines, shown on the general map of Fig. 1. A convergence analysis with two other classifications, ISI subject categories and CHI subfields, is sketched.

Specialty level

Comparison with existing patterns is based on our "raw" original classification (IBIS), without any expert reclassification. As the problem is not to discuss the overall coverage but the classification patterns of the bases, we only consider the journal set common to the 3 bases and specialties with more than 6 journals in each classification. For each IBIS specialty, a "corresponding" specialty is looked for in CHI and ISI subfields, i.e., the one with the maximum number of common journals. Mutual inclusion index (Ochiai) and best unilateral inclusion with ISI, respectively CHI, and corresponding IBIS specialties are computed. Table 2 shows examples of the convergence analysis of the three classifications (CHI/ISI/IBIS) on 1993 data.

In Table 2, the code "+++" indicates excellent ratios >85%, "++" good >75%, "+" fairly good >65%. > and < signs show the direction of inclusion. For example CHI and ISI "Astronomy-Astrophysics" are included in the corresponding IBIS specialty, which is only sligthly larger. Blank indicates lower ratios, with different choices of delineation, e.g. for IBIS and ISI "Microbiology". A strong "best inclusion" whith a low "mutual inclusion", indicates different choices in cutting level: for instance IBIS "Parasitology & Tropical Medicine" aggregates the two separate corresponding CHI/ISI subfields. Conversely, IBIS "Soil Science & Agronomy" is only a part of CHI large "Agriculture & Food Science" category. Generally, CHI specialties tend to be smaller despite their lower number, because of the unique assignment of journals.

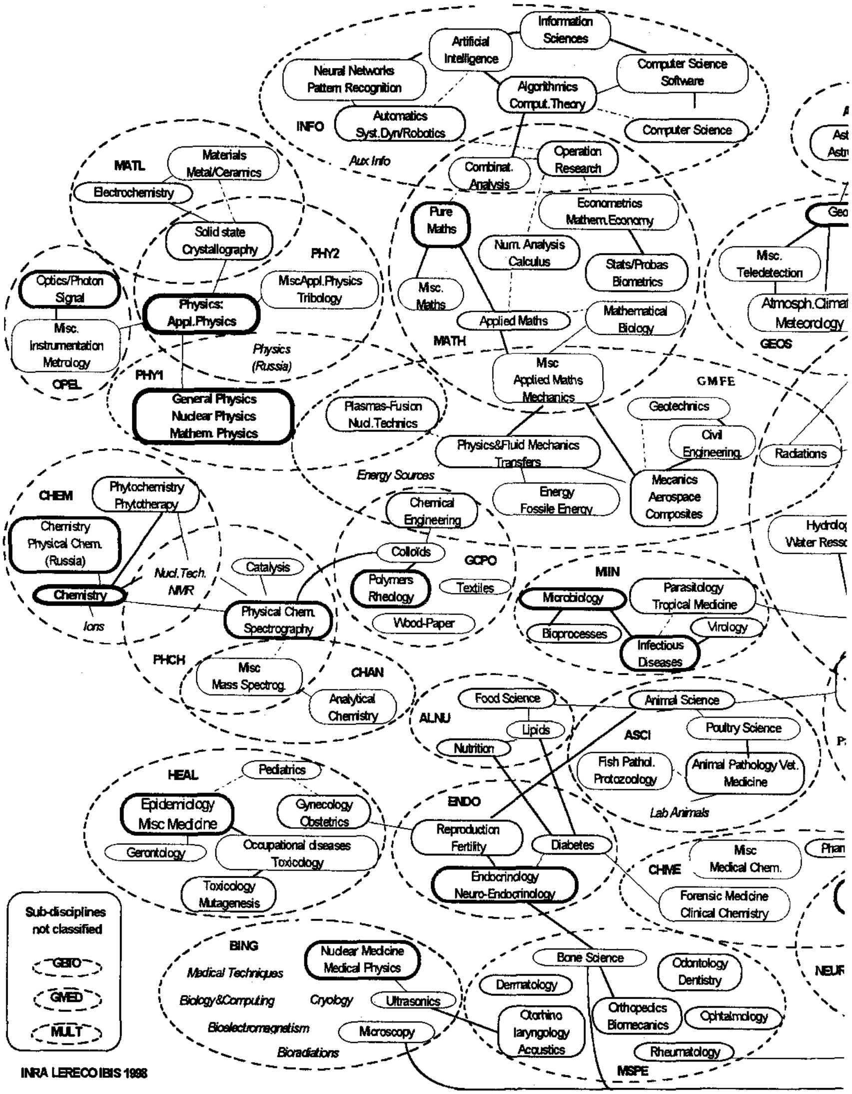
Information Sciences
Artificial Intelligence
Neural Networks Pattern Recognition
Computer Science Software
Algorithmics Comput. Theory
Automatics Syst.Dyn/Robotics
INFO
Computer Science
Aux Info
MATL
Materials Metal/Ceramics
Electrochemistry
Combinat. Analysis
Operation Research
Pure Maths
Econometrics Mathem.Economy
Solid state Crystallography
PHY2
Num. Analysis Calculus
Stats/Probas Biometrics
Optics/Photon Signal
MiscAppl.Physics Tribology
Misc. Maths
Physics: Appl.Physics
Applied Maths
Mathematical Biology
Misc. Instrumentation Metrology
Physics (Russia)
MATH
OPEL
PHY1
Misc Applied Maths Mechanics
GMFE
General Physics Nuclear Physics Mathem. Physics
Plasmas-Fusion Nucl.Technics
Geotechnics
Civil Engineering
Physics&Fluid Mechanics Transfers
Radiations
Energy Sources
Energy Fossile Energy
Mecanics Aerospace Composites
CHEM
Phytochemistry Phytotherapy
Chemistry Physical Chem. (Russia)
Chemical Engineering
Colloids
GCPO
Catalysis
Nucl.Tech. NMR
Polymers Rheology
Textiles
MIN
Chemistry
Microbiology
Parasitology Tropical Medicine
Physical Chem. Spectrography
Wood-Paper
Ions
Bioprocesses
Virology
Infectious Diseases
PHCH
Misc Mass Spectrog.
CHAN
Analytical Chemistry
Food Science
Animal Science
ALNU
Lipids
ASCI
Poultry Science
Nutrition
HEAL
Pediatrics
Fish Pathol. Protozoology
Animal Pathology Vet. Medicine
Epidemiology Misc Medicine
Gynecology Obstetrics
ENDO
Lab Animals
Gerontology
Occupational diseases Toxicology
Reproduction Fertility
Diabetes
Misc Medical Chem.
CHME
Toxicology Mutagenesis
Endocrinology Neuro-Endocrinology
Forensic Medicine Clinical Chemistry
Sub-disciplines not classified
GEIO
GMED
MULT
BING
Nuclear Medicine Medical Physics
Bone Science
Odontology Dentistry
Medical Techniques
Dermatology
Biology&Computing
Cryology
Ultrasonics
Orthopedics Biomecanics
Ophtalmology
Bioelectromagnetism
Microscopy
Otorhino laryngology Acoustics
Bioradiations
Rheumatology
MSPE
INRA LERECO IBIS 1998
Geo
Misc. Teledetection
Atmosph.Climat Meteorology
GEOS
Hydrolo Water Resso
NEUR

Fig. 1 Left

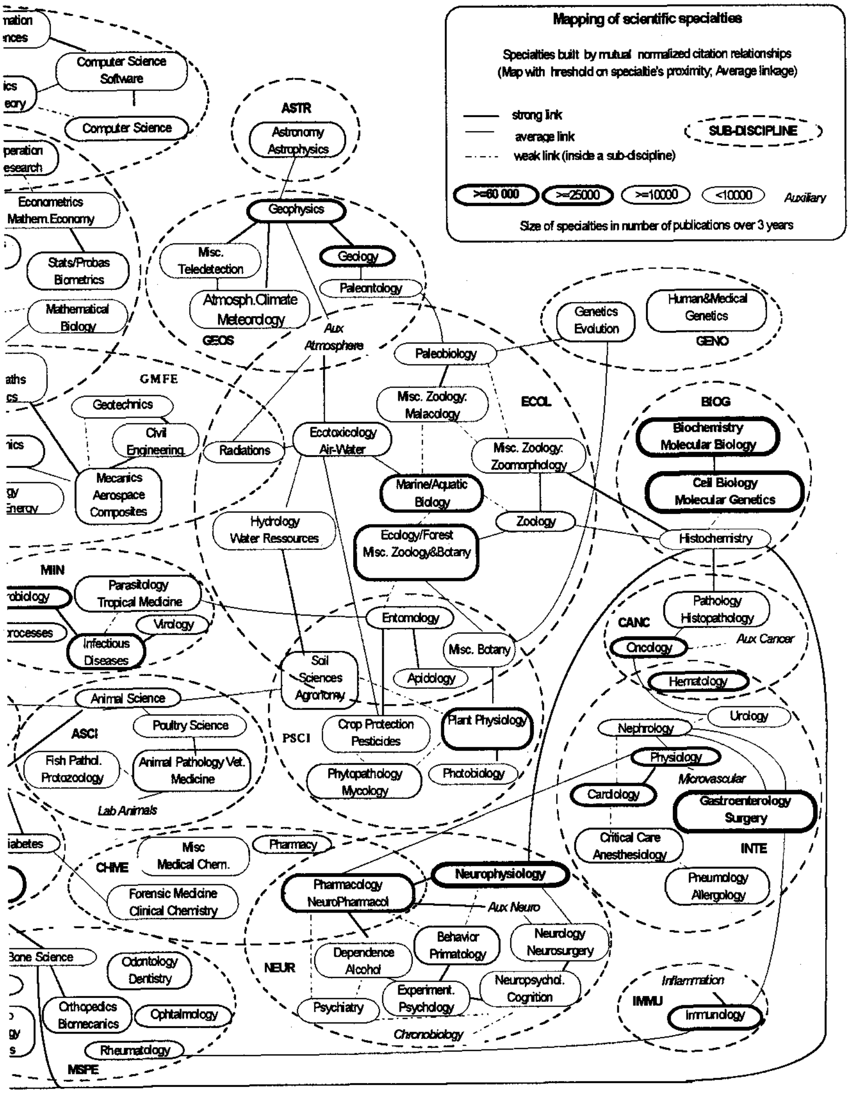
Mapping of scientific specialties
Specialties built by mutual normalized citation relationships
(Map with hreshold on specialtie's proximity, Average linkage)
strong link
average link
weak link (inside a sub-discipline)
SUB-DISCIPLINE
>=60 000
>=25000
>=10000
<10000
Auxiliary
Size of specialties in number of publications over 3 years
nation
nces
Computer Science Software
ics
eory
Computer Science
peration
esearch
Econometrics Mathem.Economy
Stats/Probas Biometrics
Mathematical Biology
aths
cs
ics
gy
nergy
ASTR
Astronomy Astrophysics
Geophysics
Misc. Teledetection
Geology
Paleontology
Atmosph.Climate Meteorology
GEOS
Aux Atmosphere
GMFE
Geotechnics
Civil Engineering
Mecanics Aerospace Composites
Radiations
Ecotoxicology Air-Water
Hydrology Water Ressources
Paleobiology
Misc. Zoology. Malacology
ECOL
Misc. Zoology. Zoomorphology
Marine/Aquatic Biology
Zoology
Ecology/Forest Misc. Zoology&Botany
Genetics Evolution
Human&Medical Genetics
GENO
BIOG
Biochemistry Molecular Biology
Cell Biology Molecular Genetics
Histochemistry
MIN
Parasitology Tropical Medicine
obiology
Virology
rocesses
Infectious Diseases
Entomology
Misc. Botany
Soil Sciences Agronomy
Apidology
Animal Science
Poultry Science
ASCI
Fish Pathol. Protozoology
Animal Pathology Vet. Medicine
Lab Animals
PSCI
Crop Protection Pesticides
Plant Physiology
Phytopathology Mycology
Photobiology
CANC
Pathology Histopathology
Oncology
Aux Cancer
Hematology
Urology
Nephrology
Physiology
Microvascular
Cardiology
Gastroenterology Surgery
Critical Care Anesthesiology
INTE
Pneumology Allergology
iabetes
CHME
Misc Medical Chem.
Pharmacy
Forensic Medicine Clinical Chemistry
Pharmacology NeuroPharmacol
Neurophysiology
Aux Neuro
Behavior Primatology
Neurology Neurosurgery
NEUR
Dependence Alcohol
Neuropsychol. Cognition
Psychiatry
Experiment. Psychology
Chronobiology
Inflammation
IMMU
Immunology
Bone Science
Odontology Dentistry
Orthopedics Biomecanics
Ophtalmology
Rheumatology
MSPE

Fig. 1 Right

At the specialty level, in many cases the process naturally retrieves groupings of journals kindred to ISI or CHI ones: it seems natural since they both involve some kind of journal citation analysis. A striking result is that by and large, the three patterns agree. Roughly one hundred IBIS specialties out of the 124 specialties of reasonable size have at least a fairly good overlap with ISI or CHI subfields. A very high level of convergence was expected and obtained in dense and strongly identified subfields such as "Astronomy-Astrophysics", "Statistics-Probability", many medical specialties (especially "Gerontology", "Ophtalmology", "Rheumatology", "Odontology"), "Veterinary medicine", "Polymer" etc. Many other subfieds show a good convergence.

Table 2
Examples of convergence analysis with ISI and CHI nomenclatures at the subfield level

Specialties with a high INRA activity

IBIS	CHI 1993			ISI 1993		
	'<' if CHI<IBIS	mutual	best	'<' if ISI<IBIS	mutual	best
Analytical chemistry	Analytical chemistry<	+	+++	Chemistry, analytical		++
Animal Science/Dairy				Agriculture, dairy&animal <	+	+
Bioprocesses				Biotech/Applied microbio <	+	+++
Food science				Food science&technology <	++	++
Microbiology	Microbiology<	+	++	Microbiology>		
Nutrition & Dietetics	Nutrition & Dietetics <	++	+++	Nutrition & Dietetics <	+	+
Parasitology/Trop Medicin	Parasitology <		++	Parasitology <		+++
Parasitology/Trop Medicin	Tropical Medicine <		+++	Tropical Medicine <		+++
Soil science/ Agronomy	Agric & Food Science >		++	Agriculture <		+++
Plant physiology	Botany>		+	Botany <		+
				Horticulture		+++
Animal Path/Vet. Medicine	Veterinary Medicine <	+++	+++	Veterinary Medicine >	++	+++
Virology	Virology <	+	+++	Virology <	++	+++

Other specialties

IBIS	CHI 1993			ISI 1993		
	'<' if CHI<IBIS	mutual	best	'<' if ISI<IBIS	mutual	best
Astronomy-Astrophysics	Astronomy-Astrophysics <	+++	+++	Astronomy-Astrophysics <	+++	+++
Gerontology	Geriatrics<	+++	+++	Geriatrics & Gerontology	+++	+++
Hematology	Hematology <	++	+++	Hematology <	++	+++
Polymer Rheology	Polymers<	++	+++	Polymer Science <	+	+++
Stats/Proba/Biometrics	Probabilities & Statistics <	++	+++	Statistics & Probability	+++	+++

A few of our citation-built specialties do not retrieve conventional groupings. In some cases, the delineations do not follow the traditional borders. For instance, "Pharmacology" is clearly distinct from "Pharmacy" and attracted by neurosciences. The way citation analysis operates clearly appears in limit cases: the specialty "Russian physics" is an artefacts from the thematic point of view,* but stresses the low internationalization of russian journals dominating these clusters. The social dimension of citations also separates "Apidology" from "Entomology", as the institutional and historical networks are not identical. In the subdiscipline "Food Science and Nutrition" (ALNU, see below), lipidists are set apart from other food scientists because they are more concerned by nutrition and metabolism and seem to build a different scientific community. Another striking point is that citation analysis, even built on symmetrical indexes, sometimes reconstructs vertical channels, such as for the block optics/photonics/signal which aggregated applied physics journals and engineering journals.

But the main problem is the unability to split a few large sets even with an algorithm that does not favour exceedingly skew cluster size distribution. Citations between journals – at least with our methodology – cannot split a few big specialties:

- the dense core of chemistry (except analytical chemistry and materials).
- two major subfields emerge in physics, each one hardly splittable; moreover they share generalist journals.
- the same is true for pure biology with two major subfields "Biochemistry & Molecular Biology", "Cellular Biology-Molecular Genetics".

The classification tree shows a very high stability for all these big sets: trying to cut at a smaller scale would ripe off small satellites out of the core ("artichoke" model), instead of breaking it down into balanced subsets. This may be interpreted in two ways: either these subfields are very dense and hardly decomposable, either the limits of a method based on journal transactions are reached, probably because of the weight of generalist journals in these areas.

Higher levels of aggregation: subdisciplines

Now let us have a look at the relationship between specialties. The threshold maps of Fig. 2 give a landscape of the areas of science where INRA is the most active. Both maps start from the same specialties built as mentioned by average linkage algorithm. Map 2A displays the original inter-specialty similarities after symmetrical Ochiai indexes. These indexes are multiplicative: if say "Food Science" refers to "Analytical Chemistry" literature but "Analytical Chemistry" journals never refer to "Food Science", the link between both specialaties is zero. To draw attention on some connections between classification method and interpretation, a variant is shown on Map 2B.

**Leydesdorff* and *Cozzens*[14] mentioned a similar phenomenon p.152.

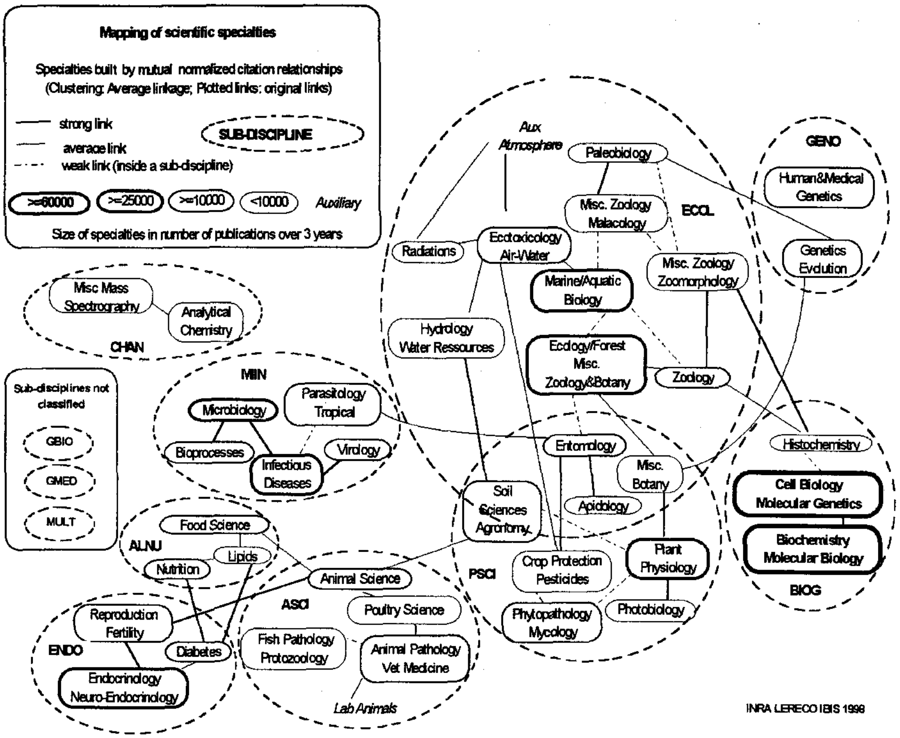

Fig. 2A Original Links

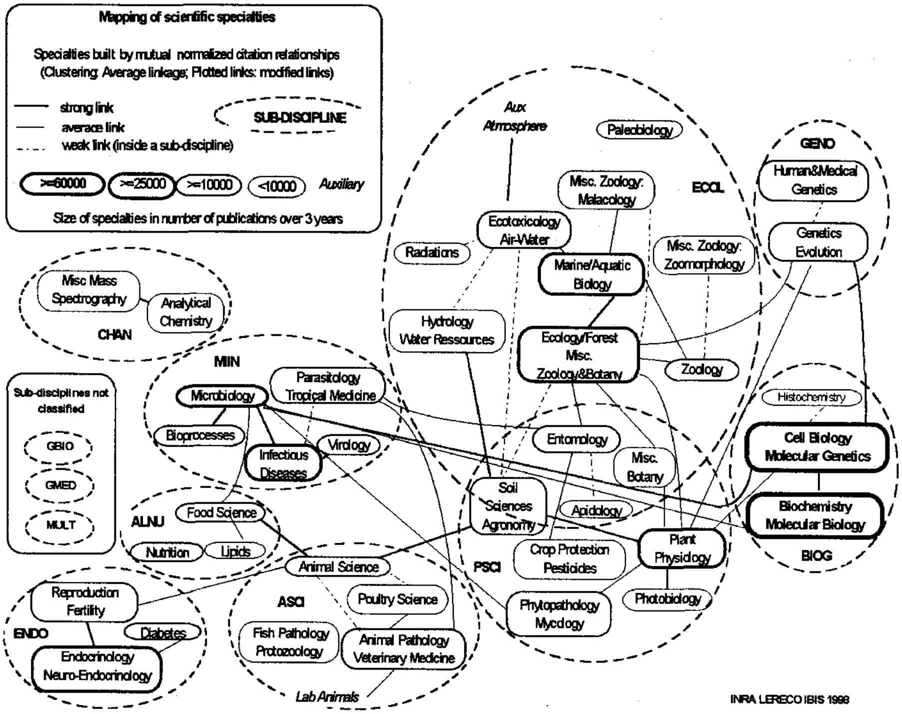

Fig. 2B Modified Links

The basic design of specialties is unchanged but we show a semi-local inter-specialty linkage, calculated after the average distance between the mutual 10 closest neighbour journals for each couple of specialties. Thresholds have been set to obtain the same number of links for the whole SCI/CMCI as on Map 2A. On both maps, for legibility, weak linkages are represented only within subdisciplines borders (subdisciplines defined in the final stage, see below). Other linkages are represented in all cases, except those to subdisciplines not shown on the map.

Both views implicitly carry interpretation of what mapping should reflect. The first one 2A, built on pure average linkage, masks local linkages, especially for big specialties. The second map 2B probably seems more realistic, because local connections appear: big specialties gain some links (e.g. the major subfields in molecular biology, biochemistry, genetics) whereas small specialties are more isolated (e.g. lipids). However, Map 2B implicitly suggests that sharing a common issue is sufficient to get close, which is not necessarily a good argument to build subdisciplines, especially for tranversal subfields like "Microbiology", locally linked to many different specialties.

Let us go back to the global landscape of Fig. 1. Strong linkages and networks are consistent with traditional groupings, for instance:

- neurosciences (NEUR): neurosciences/neuropathology as a global set is retrieved; it appears as a constellation of specialties with fairly strong linkages. Trying to split it e.g. between a more fundamental subfield (centered on neurosciences) and an applied subfield (centered on psychiatry/behavior) is a difficult task if using citation criterion only.
- applied biology. Four large sets emerge: environment/ecology (ECOL), plant science (PSCI), animal science (ASCI) and food science & nutrition (ALNU); experts advice is necessary to precise boundaries/overlaps, especially for the first two.
- other dense expected groups include the couple biochemistry/molecular biology and cellular biology (BIOG), the microbiology network (MIIN), the neurosciences network (NEUR) and the geosciences network (GEOS); less specific of INRA activity: mathematics, statistics-econometrics (MATH), information/ computer sciences (INFO), toxicology and public health (HEAL).

More generally, from the three "new disciplines" often evoked (neurosciences, ecology/environment, material sciences), the last one does not really emerge as a dense area, at least for 1993 JCR. This frontier area between physics, chemistry and engineering sciences still appears splitted between somewhat disconnected sets (materials/ceramics, polymers, composites).

Some subfields are fairly isolated from the citational point of view, especially medical specialties; any grouping of these is more or less arbitrary. But some different situations appear. As mentioned above, citations often stress basic-applied relationships, e.g. immunology and rheumatology; or the relation between an organic system and its dominant pathology, e.g. oncology attracting urology (rather than conventional intra-system association urology/nephrology) and, at a lesser degree, hematology; shared issues also link well-identified specialties (say osteoporose for osteology and endocrinology or histocompatibility for nephronology and immunology). Such bridges should not be considered as constraints to build subdisciplines: it is up to the experts/users to consider whether partaking in a common problem is sufficient to aggregate specialties.

To conclude this point:

- in most cases citation linkages are efficient to build consistent specialties, but, as far as our protocol is concerned, they fail to split (or split in a questionable way) the hard cores in chemistry, physics, and pure biology. While many other fields would easily be broken down, these large scale-resistant specialties determine the skewness in field size distribution, and this in a wide range of cutting levels corresponding to "specialties" or "sub-specialties" design.
- citation linkages are very helpful but not sufficient to delineate subdisciplines. In many cases they show specialties groupings which are good candidates, but other criteria, possibly conveyed by experts, should be mobilised as well as classifications based on factor analysis.[22]

Final structure

We have carried out the study starting from citation linkages and expert-advice. The final list of subdisciplines is given in Table 3.

For the sake of longitudinal studies, we have assigned in a second stage ISI subfields to subdisciplines they were closer to. Journals of both IBIS and ISI categories build the final subdiscipline. An example of assignment is given in Table 4 for some subdisciplines. In the first three cases, Science & Nutrition (ALNU), Animal Science & Pathology (ASCI)and Chemical Engineering, Polymers & Colloids (GCPO), IBIS classification is more detailed than ISI nomenclature. The opposite is true for some other subdisciplines such as Optics, Signal & Electronics (OPEL), an example of high aggregation in IBIS classification and of course for Physics and Chemistry.

Table 3
List of subdisciplines

Subdisciplines with significant INRA activity		Other subdisciplines	
Short name	Content	Short name	Content
ALNU	Food Science & Nutrition	ASTR	Astronomy-Astrophysics
ASCI	Animal Science/ Animal Pathology	BING	Biomedical Engineering
BIOG	Biochemistry/Molecular&Cellular Biology&Genetics	CANC	Oncology
CHAN	Analytical Chemistry	CHME	Medical Chemistry/ Pharmacy
CHIM	Chemistry	GCPO	Chemical Engineering/ Polymers/ Colloids
ECOL	Ecology/Ecosystems/Ecotoxicology/Environment	GMED	General Medicine
ENDO	Endocrinology/ Reproduction	GMFE	Mechanical Engineering/ Fluids/ Energy
GBIO	General Biology	HEAL	Public Health/Epidemiology/Life cycle/Toxicology
GENO	Genetics/ Evolution	INFO	Computer & Information
GEOS	Earth&Atmosphere	MATH	Mathematics/ Statistics
IMMU	Immunology	MATL	Materials /Metals/ Cristallography
INTE	Gastroenterology/Cardiology/Pneumology/Surgery	MSPE	Medical Specialties, Miscellaneous
MIIN	Microbiology/ Virology/ Infections/ Bioprocesses	OPEL	Optics/ Photon/ Signal/ Electronics
MULT	Multidisciplinary	PHCH	Physicochemistry/ Spectroscopy/ Spectrography
NEUR	Neurosciences/ Neuropathology	PHY1	General, Mathematical, Nuclear Physics
PSCI	Plant Physiology/ Plant Protection/ Agronomy	PHY2	Applied Physics/ Solid State/ Cristallography

Table 4
Examples of final subdisciplines

Subdiscipline	IBIS specialties	ISI subfields
ALNU	Food Science	Food Science & Technology
	Lipids	Nutrition & Dietetics
	Nutrition	
ASCI	Animal Pathology/Veterinary Medicine	Agriculture, Dairy & Animal Science
	Animal Science	Veterinary Medicine
	Fish Pathology/Protozoology	
	Poultry Science	
	Laboratory Animals	
GCPO	Chemical Engineering	Engineering, chemical
	Colloids	Materials Science, Paper & Wood
	Polymers/ Rheology	Materials Science, Textiles
	Textiles	Polymer Science
	Wood/ Paper	

Conclusion

To position a research institute in ISI databases, we tried a seed-free hierarchical structuration of SCI/CMCI journals based on direct journal citation-transaction. At the subfield level, results show a large convergence with classifications based partly or solely on citations. In some cases the seed-free construction gives more accurate pictures than usual nomenclatures. The problem of generalist journals remains the principal limitation of the process, especially in disciplines where they create non-splittable aggregates. The alternative are the document-level process mentioned below, or the extra treatment of "problematic journals".[7] For subdiscipline delineations, journal transactions are helpful but, in this range of weaker linkages, the citation links cannot be compelling for subdiscipline delineation. Let us notice that citationnist logics, even moderated by symmetrical indexes, tend to emphasize vertical connexions between background and applications in some cases. Expert advices are required. Another limit of the process is the changing structure of science networks: some temporal adaptations are possible, but the classification process should be repeated on recent data.

However, this classification on SCI/CMCI journals proved to be operational for our preoccupations. It is not a mere catalogue: the underlying structure is easily understandable and zooms are possible. Borders and close or large neighbourhoods can be visualised and studied. Besides, it provides an external but appropriable view of science fields, particularly useful when investigating new developments.

Of course, classification at the journal level cannot be more than a surrogate of direct document assignment.[6] Whether the quality of this surrogate may decline over time depends on the proportion of mutlidisciplinarity in journals among other factors. On the other hand, document-level classification at very large scale is possible, either through federation of nomenclatures or bibliometric classifications. More powerful computers promise that sophisticated means, very efficient for small-size studies, may be generalised for new-generation classification based on co-words, or improved co-citation analyses.[19,23] An alternate way is to exploit complementarities between databases, often used in monitoring science studies after a proper unification.[2] High-quality assignments of some specialised databases at the document level, may be combined to achieve a fine-grain classification. Such processes, far more ambitious but costly, are explored by indicators producers. But journal classification remains a handy and efficient solution in many applications at macro or meso level.

*

The authors thank A. *Chartier* and H. *Pampouille* (INRA-Direction of Information and Communication) for their collaboration in validation processes and anonymous referees for helpful comments.

References

1. GOMEZ I., BORDONS M., FERNANDEZ M.T., MENDEZ A., Coping with the problem of subject classification diversity, *Scientometrics*, 35, 2 (1996) 223-237.
2. TIJSSEN R., VAN LEEUWEN T.N., On generalising scientometric journal mapping beyond ISI's journal and citation databases, *Scientometrics*, 33, 1 (1995) 93-116.
3. GRUPP H., HINZE S., International orientation, efficiency of and regard for research in East and West Germany: a bibliometric investigation of aspects of technology genesis in the united Germany, *Scientometrics,* 29, 1 (1994) 83-113.
4. COWARD R.H., Science and technology indicators in a policy context, *Proceedings of the Joint EC-Leiden Conference on Science and Technology Indicators*, Leiden, 23-25 oct 1991, A.F.J. VAN RAAN, R.E. DE BRUIN, H.F. MOED, A.J. NEDERHOF, R.W.J. TIJSSEN (Eds), DSWO Press, Leiden University (1992) 234-241.
5. KATZ J.S., HICKS D., The classification of journals: a new approach, *Proceedings of the Fifth Biennal Conference of the International Society for Scientometrics and Informetrics*, River Forest (Chicago), June 7-10 1995, Learned Information Inc., Medford, New Jersey, 245-254.
6. BOURKE P., BUTLER L., Institutions and the map of science: matching university departments and fields of research, *Research Policy,* 26,6 (1998) 711-718.
7. GLÄNZEL W., SCHUBERT A., CZERWON H.J., Improving ISI's subject classification system. A paper-by-paper assignment of papers in multidisciplinary and general journals based on reference analysis, *Book of Abstracts of the Fifth international Conference on Science and Technology Indicators*, Hinxton, Cambridge June 4-6 1998, SPRU-CWTS, 16.
8. NARIN F., CARPENTER M., BERLT N.C., Interrelationships of scientific journals, *Journal of the American Society for Information Science*, 23 (1972) 323-331.
9. CARPENTER M.P., NARIN F., Clustering of scientific journals, *Journal of the American Society for Information Science*, 24 (1973) 425-436.
10. DOREIAN, P., FARARO T.J., Structural equivalence in a journal network, *Journal of the American Society for Information Science*, 36, 1 (1985) 28-37.
11. TODOROV R., GLÄNZEL W., Journal citation measures: a concise review, *Journal of Information Science,* 4 (1988) 47-56.
12. TIJSSEN R., DE LEEUW, J., VAN RAAN, A.F.J., Quasi-correspondence analysis on square biblometric transaction matrices, *Scientometrics*, 11 (1987) 347-361.
13. BURTON M.P., The use of citations matrices to group journals, *Scientometrics,* 33, 2 (1995) 257-262.
14. LEYDESDORFF L., COZZENS S.E., The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index, *Scientometrics*, 26, 1, (1993) 135-156.
15. LUUKONEN T., Why has Latour's theory of citation been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis, *Scientometrics*, 38, 1, (1997) 27-37.
16. VAN RAAN A.F.J., Fractal dimension of co-citations, *Nature*, 347 (1990) 626.
17. HARTIGAN, J.A., *Clustering Algorithms*, John Wiley, New York, 1975.
18. LEYDESDORFF L., The development of frames of references, *Scientometrics*, 9, 3-4, (1987) 103-125.

19. ZITT M., BASSECOULARD E., Reassessment of co-citation methods for science indicators: effect of methods improving recall rates, *Scientometrics*, 37, 1 (1996) 333-351.
20. MURUGESAN P., MORAVCSIK M.J., Variation of the nature of citation measures with journal and scientific specialties, *Journal of the American Society for Information Science*, 29 (1978) 141-155.
21. PUDOVKIN A.I., FUSELER E.A., Indices of journal citation relatedness and citation relationships among aquatic biology journals, *Scientometrics*, 32, 3 (1995) 227-236.
22. REINERT M., Classification descendante hiérarchique: un algorithme pour le traitement des tableaux logiques de grande dimension, data analysis and Informatics, 4, Diday et al. Eds., North-Holland (1986) 23-28.
23. SMALL H., Update on science mapping: Creating large document spaces, *Scientometrics*, 38, 2, (1997) 275-293.
24. ZITT M., TEXEIRA N., Science macro-indicators: some aspects of OST experience, *Scientometrics*, 35, 2 (1996) 209-222.
25. IRVINE J., MARTIN B.R., ISARD P.A., *Investing the Future: An International Comparison of Government Support for Academic and Related Research*, Edward Elgar Publishing, Aldershot (1990) 251-265.
26. DE SOLLA PRICE D.J., The analysis of scientometric matrices for policy implications, *Scientometrics*, 3, 1 (1981) 47-54.
27. NOMA E., An improved method for analyzing square scientometrics transaction matrices, *Scientometrics*, 4, 4 (1982) 297-316.

Appendix

Acronyms of institutions

CHI	CHI Research Inc, USA
FHG-ISI	Fraunhofer Gesellschaft-ISI, Germany
INRA	Institut National de la Recherche Agronomique, France
ISI	Institute for Scientific Information, USA
ISSRU	Information Science & Scientometrics Research Unit, Hungary
NSF	National Science Foundation, USA
OST	Observatoire des Sciences et des Techniques, France
SPRU	Science Policy Research Unit, Great Britain
SRI	Formerly Stanford Research Institute, USA