Invited plenary talk

"SCIENTOMETRIC INDICATORS: A FEW CHALLENGES data mine-clearing; knowledge flows measurements ; diversity issues"

Michel Zitt*, **

Lereco, INRA, BP 71627, F-44316 Nantes cedex 3, France
***Observatoire des Sciences et des Techniques (OST), 93 rue de Vaugirard, F-75015 Paris, France*
zitt@nantes.inra.fr

Introduction - the new context of bibliometrics

Rather perhaps than a revolution, scientometrics and bibliometrics are encountering a strong development in demand and new forms of supply.
Some signs of this new context are:

- on the supply side, the availability of publication sources and statistics, due to Thomson-ISI and Internet developments among others, and the relatively easy way to draw quick statistics from this background.
- on the demand side, the coming of age of university evaluation practices, which recognize, among other means, the pertinence of bibliometrics. Along with these responses to increasing research management needs, the internalization of bibliometric tools by other disciplines, such as economics, for cognitive purposes, is also on the rise.

This situation results in contrasting perspectives: on the one hand, it can favour spectacular "hit-parades" and some frenzy of numbers; on the other hand, it paves the way for more cautious and sophisticated evaluation systems, rooted in a better understanding of the dynamics of science.

This speech gives me the opportunity to sketch a few of the challenges that bibliometric indicators are facing: some are classic issues, while others are more related to the new context of science. Section I deals with "data mine-clearing"; Section II with measurement of knowledge flows; Section III with diversity.

1st. Challenge: data mining and "data mine-clearing"

Bibliometrics mainly works with empirical models and data, and should not overlook where data come from. To put it bluntly, in data-mines there are mines in the data. This is of course very well known by bibliometricians, but it is still worth recalling for users of indicators and/or policy makers. Data illusion takes on many forms:

- "*in databases we trust*": the delineation of the standard bibliometric database (Web of Science) is largely arbitrary for low-impact and low-internationalized journals: the tail expresses a sub-population, which has devastating effects on indicators for some countries or actors in those countries (ex. FSU). This does lead to envisage alternative measures based on calibrations of data, depending on what is at stake (Zitt, Ramanana, Bassecoulard, 2003). Inter-disciplinary balances are not guaranteed either. Particularly serious issues are met in social and human sciences, reviewed in several occasions (e.g. Hicks, 2004). Each discipline exhibits different issues, due to the mode of production (articles, conferences, books) and the national biases of the databases are quite severe in some disciplines.

- "*in names we trust*": the lack of standardisation of individual data for authors and especially for addresses, as well as content markers (words). Unification/disambiguation remains a challenging issue in bibliometrics, which in this respect could place itself under the auspices of Confucius'.[1] Several of the most

[1] Confucius: "There must be a correction of terminology."

successful studies currently available, for example ISI-ESI and Shanghai study (Liu et al.), exhibit low standards of unification/disambiguation. Let us take the Shanghai ranking: in the record chart, for one of the universities in Southern France, as many as 3/4 of papers are lost due to the lack of unification (primarily due to poor initial mention by authors themselves), and many other examples could be found, though perhaps not in this order of magnitude. This situation may change, but, in the meantime, ex post self-identification by institutions themselves is necessary to yield credible evaluations[2]. Globally however, the lack of unification of items (names, institutions, vocabulary...) hinders analyses of many types: co-activity science-technology, individual evaluation, etc.

- "i*n indicators we trust*": Naturally, data is not the only aspect concerned. Many bibliometricians from the US (ISI, CHI Research) or Europe (ISSRU, SPRU, ISI-FhG, Leeuwen Un.. among many others) have stressed the questions of the sensitivity of classic indicators to methodology, especially since these are typically based on skew distributions (Rousseau, 1990); Egghe, 1991), where the heavy tail strongly affects conventional measures. Theoretical and practical significance of indicators are a concern, not to speak of other issues... Even the most famous historical indicator, ISI's journal impact factor, appeared to be slightly flawed (Moed, 1995).

At this point, the challenge for bibliometrics is to contribute in the selection and control the data and to resist, in some cases, the whirlpools of data-mining.

2nd. Challenge: measuring knowledge circulation

Positioning indicators, stressing the position of actors in strategic networks, have in the last decades, challenged the traditional input-output rationale (see for example R. Barré, 2006). Bibliometrics does not address all these networks, but those embodying circulation of knowledge are the bull's eye:

- scientific collaboration of many kinds. For example, determining factors of co-publication at the macro-level have been identified, and sometimes ranked. Understanding the collaboration process at the micro-level is perhaps trickier. Co-authorship now enters into promising information circulation models proposed, among others, by economists.

- knowledge transfers made visible by their citation counterpart, following Merton's hypothesis that citations recognize and an intellectual debt. The approach is

---

Tzu Lu: "Are you serious? Why is this so important?"
Confucius: "You are really simple, aren't you? A serious man is cautious about jumping to conclusions about that which he does not know." (*Analects*)

[2] French public research "Cooperative" exercise (Zitt M., Bauin S. and Filliatreau G., coord., 2004)

quite efficient at the macro-level to assess transactions between fields, but, at the micro-level, a qualification of the citation types, according to the various citation theories proposed by sociologists (reviewed by Cronin, see for example 2004), would be helpful.

- content markers allow the uncovering of thematic proximity by linguistic approaches, either limited to lexical aspects or including semantic relations.

- mobility of scientists: this question has many facets, and bibliometrics is only one of them, which face unification/disambiguation issues when addressing this problem on a large scale.

- the web has become a fantastic platform for knowledge circulation, both for disseminating classic media and inventing new ways, formal or informal, of dissemination and interaction. Circulation of scientific information through the web raises the issues of future bibliometric indicators. It is often said that bibliometrics is a mirror reflecting the self-organization of scientific communities and their intricate webs of publications, texts, authors. If the model of publication changes, will webometrics replace bibliometrics? Is webometrics a Chamber of mirrors creating illusions, or the future of bibliometrics? We can think of many fruitful analogies and dangerous mixes, some of them raised in the debate on web "impact factors" (Bjorneborn and Ingwersen, 2001).

In many occasions, a problem implying knowledge circulation may be addressed by competing or complementary measures on each of these networks. This holds for homogenous transfers, within science, as well heterogeneous transfers, for example from science to technology, a central subject both at the theoretical level and for indicators of the "third mission" of universities. What is perhaps lacking is a unified perspective on knowledge circulation.

Tools for network studies have now been used for a long time in bibliometrics applications, for instance actor-networks introduced by the relativist school (Callon, 1986). But the renewal of Milgram's small world theory by Watts and Strogatz (1998) has boosted the research on social networks.. Analyses of scientific networks can now be found in papers by mathematicians ("Erdos project"), computer scientists (Barabasi model), physicists, and recently economists, the latter particularly concerned by designing profit and costs functions. The network modelling of science also draws attention on scale issues, for example scale invariance attached to some categories of models, with important consequences in terms of indicators. The various networks describing scientific activity are not expected to exhibit the same parameters of scaling (word networks, citation networks, co-authorship networks). There may also be strong differences between the smooth scale-invariance of scientific themes on the one hand, whatever the descriptor, words or citations, and the discontinuous embedment of research organizations on the other hand. In spite of overlaps mentioned above, institutions have contours whatever the scale (teams, labs, departments, universities, etc.), while scientific themes are fundamentally fuzzy, and of course overlapping, obeying the laws of informetrics.

At the same time, bibliometricians have to care about the sensitivity to methodology. Approaching knowledge circulation by each of abovementioned networks may lead to quite different estimates, the results are highly sensitive to the weighting and counting options even within a single network, such as citations.

3rd. Challenge: coping with diversity

Diversity and its consequences on the stability of indicators

In sociological terms, and even when adopting the Mertonian views, it is clear that within a common system of publication and communication norms, scientific communities exhibit quite different behaviours depending on the type of research, the degree of application and the nature of the field …. They do not publish with the same frequency, they do not behave in the same way neither for co-publications nor for referencing. This has been established long ago by pioneers in citation studies who proposed various typologies and solutions for normalisation (Pinski and Narin, 1976; Murugesan and Moravcsik, 1978; Schubert, Glanzel and Braun, 1998, etc.).

In bibliometric terms, we could state that within the framework of general laws, a large variation of the parameters takes place amongst local sub-networks of science, expressing the specificity of behaviour, organisation and diversity within each field and type of research. Besides theoretical challenges of modelling self-organized scientific activity, which takes the generation of diversity and complementarity as an object, a practical challenge in evaluation-oriented scientometrics is to cope with the consequences of this diversity, namely the heterogeneity of areas, and to find appropriate reference sets for indicators in order to avoid mixing apples and oranges (see Zitt, 2005).

Moreover, due to the particular structure of scientific networks, variety is present whatever the scale of observation. For example in terms of impact measures, a sensible (i.e. area-normalized) measure of performance proves (to be) pretty unstable when successive embedded sets of growing sizes (research front, speciality, subfield, field...) are used as references for normalisation. We observed for example that the content of the top-cited "excellence" class is quite dependent on the level of observation/normalisation (Zitt, Ramanana, Bassecoulard, 2005).

Diversity of science is mirrored by the variety of actors' involvements. The spectrum of actors' activity amongst fields has been one of the first tool of bibliometricians ("activity index", i.e. the economists' Balassa index, and their distribution over fields, Herfindhal and other specialisation indexes) to describe the strategy of actors (for a recent development see Adams, 2003). Variety of research actors' portfolio belongs to this family of indicators where 'ranking' as such makes little sense, since the value of high specialization or high-variety strategies depends on the context and strategies.

Delineation of scientific areas

The question of field delineation is inseparable from that of diversity. Depending on the point of view, many definitions of scientific areas and disciplines can be put forth: institutional setting of academic actors; shared topics and possibly shared journals; shared terminology; close connections of collaboration or citation... Scientometrics can operationalize any definition which can be translated in terms of scientific items and networks. Information may be extracted from published output or from external sources, e.g. lists of researchers or institutional sites; built *ex ante*, in nomenclatures (ISI subject categories are *de facto* a nomenclature), or *ex post* by statistical classifications; rely on usual lexical queries or on sophisticated co-word or co-citation studies; vary in level of aggregation; be general or specific to a player; stem from coarse grain or fine-grain observation.

Fine grain classifications detect dense sub-areas in the scientific networks, and mapping of science is quite a common exercise in bibliometrics, based on various methods of data analysis and graph analysis. Overlaps and fuzziness of scientific areas, when outlined by bibliometric methods, are a crucial point. In other words, one can find practical borders based on some information-retrieval trade-off in terms of recall-precision and costs, and detect some local optima for clustering, but natural borders, generally speaking, are an illusion. We recently addressed the delineation issue in the case of nanosciences, by using standard lexical queries as a starting point. To cope with fundamental limitations of terminology, a complementary algorithm based on citation networks (extension algorithm) has been designed (Zitt and Bassecoulard, 2006). Modelling leads to informetric laws *à la Bradford*, however less concentrated in most cases, expressing that borders can be largely extended in circles of decreasing relevance.

Challenges for data analysis and data mining methods aiming at classification and delineation of scientific areas are of course the efficiency of algorithms but also the capability to reflect skew distributions, embedments and overlaps/fuzziness of areas at various scales. Another challenge is to take advantage of advances in natural language processing, adapting them to the particularity of the scientific jargon

Measuring multidisciplinarity

Measures of multidisciplinarity can be operationalized in a variety of ways by bibliometrics, depending on the particular definitions used to delineate disciplines or fields and measure flows and overlaps. The literature on the subject is quite abundant and is based on various points of view. Measures of multidisciplinarity in static terms often play with the distribution of an actor's activity over a grid of "disciplines" with some degree of permanence, defined by institutional or thematic rationales. Flows and/or overlaps between categories created by bibliometrics delineation or some hybrid process (ISI subject categories) are common direct measures. Many authors have used the citation flows approach, e.g. Rinia et al, 2001. In a more radical way, multi-disciplinarity can be observed through weak ties amongst contexts in a thematic network, or in dynamic terms with the way and speed of their rearrangement.

Size issues

Many bibliometric studies deal with collective actors such as countries and institutions. Comparisons between actors may involve size-depending measures such as total publications or citations of the actor, or true "performance" indicators, such as productivity measures of various kinds, or impact measures. It can be shown, for example, that the Shanghai ranking is biased towards size (Zitt and Filliatreau, 2006). In contrast, most bibliometric indicators deal with both "power indicators" (embodying size effects) and "performance indicators" not -- directly -- size-dependent. A central question of science policy is whether performance correlates with power or size, and we have some indications that the questions of increasing returns and critical masses are scale-dependent, with different answers at the national systems level (Katz, 1999) and at the laboratory level. Productivity indicators are confronted with serious data issues (on allocation of human as well as financial resources), despite international norms. There is a sharp contrast between this fragility of data and the sophistication of methods such as DEA, which allow relative positioning on a collection of inputs and outputs.

conclusion

I will conclude with a few global challenges, very conscious of the somewhat arbitrary character of this list.

Growth-diversity models

Amongst the founding fathers, de Solla Price (1963) established scientometrics on models of growth. With the arising of new methodologies of networks analysis, a new horizon can perhaps be found in understanding how the creation of local variety on the one hand, and the fabric of weak ties and multidisciplinary connections on the other hand, shape growth regimes in science. Some attempts of new models (Van Raan, 2000), or reformulation of Price's questions (Bonaccorsi, 2002), may pave the way for a better understanding of science dynamics, the real reservoir of renewed indicators.

Science and outside

Although still functioning on particular norms or habits, laboratories do not only produce science but relations with many partners, as metaphors such as the "research compass card" (Laredo, Callon et al., 1992), the "mode II" (Gibbons et al., 1994) and the "triple helix" (Etzkowitz and Leydesdorff, 1997) illustrate. The role of bibliometrics is all the more easy than these transfers involve codified knowledge, and much work still needs to be done to measure, for example, science-technology relations. Perhaps less easy to handle, but also promising, are some attempts to measure particular science-society relations (Lewison, 2004).

Scientometrics and science: Feedbacks, backlashes, paradoxes

As shown by sociologists (Wouters, 1997), scientometrics is part of the system. Scientific communities are quite reactive to changes in evaluation systems, especially when funding is at stake. The star of bibliometric tools was undoubtedly Garfield's impact factor, which durably shaped the behaviour of competitive science. Recently, Butler (2003) showed the perverse effects of naive bibliometric-based funding in Australia. In the data-mining section, I mentioned the low standard of the Shanghai ranking as far as actors identification is concerned. The paradox is that the lower the standard, the stronger the expected reaction by actors themselves to correct the names for the next round - just because of the international visibility of this particular study. This virtuous circle was probably unexpected by most bibliometric offices. Mastering the feedback loops implying scientometrics is far from easy.

Ranking, positioning, benchmarking

A great threat for scientometrics is to succumb to the frenzy of numbers, rankings and data-mining techniques, and forget about the necessity of criticizing the sources, controlling the data, and making sensible comparisons. Only a few indicators can be “ranked”: impact measures, certainly, with many warnings issued by sociologists; and others by bibliometricians (normalization, aggregation properties); gross publications and citations, with respect to size-dependence and field-normalisation. Number-based excellence indicators, H-index for example (Hirsch, 2005), could participate in the worship of figures if they were applied without paying attention to the diversity of contexts.

Ranking of indicators such as the gross rate of co-publication, spectrum of partnership or diversity indexes do not make sense out of actor's strategies and context (see for instance Glänzel, Danell and Persson, 2003); they can be used for positioning, not for direct ranking unless some additional rules are introduced. The same is true for the repartition of author's activity in "mainstream" and "transfer", or more generally among the branches of the research compass or multiple helix of research. Aggregate ranking implies that all criteria are performance measures and that combination and some weighting scheme are possible, in accordance with the missions of an institution. Quantitative analysis of research institutions is not only performance ranking such as in the Shanghai exercise or in input/output approaches whatever the sophistication. Whether they are compared to global references, or against each other (benchmarking), research institutions can be studied in multi-criteria positioning, where performance criteria and ranks are present but where other measures, not amenable to ranking, are also documented -- not speaking about qualitative indicators[3].

ACKNOWLEDGEMENTS: The author thanks E. Bassecoulard, Inra-Lereco, Suzy Ramanana-Rahary and G. Filliatreau, OST, for helpful comments.

Adams J. and Smith D. ***Funding research Diversity***. London (UK): Universities UK, Evidence Ltd, 2003.

Barré R. (2006). **The emergence of 'Positioning Indicators'Towards a European STI Indicators Platform (ESTIP)?** *Second PRIME annual conference, 7th February 2006*

Bjorneborn L. and Ingwersen P. (2001). **Perspectives of webometrics**, *Scientometrics*, *vol 50, n°1, pp. 65-82*.

Bonaccorsi A. (2002) **Matching properties. Research regimes and institutional systems in science** in: "*Science as an institution, the institution of science*" *Conference* Siena, Italy, jan. 25-26, p.

Butler L. (2003). **Explaining Australia's increased share of ISI publications - the effects of a funding formula based on publication counts**, *Research Policy*, *vol 32, n°1, pp. 143-155*.

Callon M. (1986). **The sociology of an actor-network : the case of the electric vehicle**. In: Callon M., Law J. and Rip A. (eds). *Mapping the dynamics of science and technology*, London: Macmillan Press, pp. 19-34.

Cronin B. (2004). **Normative shaping of scientific practice: The magic of Merton**, *Scientometrics*, *vol 60, n°1, pp. 41-46*.

Egghe L. (1991). **The exact place of Zipf's law and Pareto's law amongst the classical informetric laws**, *Scientometrics*, *vol 20, n°1, pp. 93-106*.

Etzkowitz H. and Leydesdorff L. (1997). ***Universities and the global knowledge economy: a triple helix of university-industry-government relations***. London: Pinter, 1997.

Gibbons M., Limoges C., Nowotny H., Schartzman S., Scott P. and Trow M. ***The New production of knowledge***. London: Sage, 1994.

Glanzel W., Danell R. and Persson O. (2003). **The decline of Swedish neuroscience: Decomposing a bibliometric national science indicator**, *Scientometrics*, *vol 57, n°2, pp. 197-213*.

Hicks D. (2004). **The four literatures of Social Science**. In: Glaenzel W., Moed H. and Schmoch U. (eds). *Handbook of Quantitative Science and Technology Research*: Kluwer Academic Publishers, pp. 473-496.

Hirsch J.E. (2005). **An index to quantify an individual's scientific research output**, *Proceedings of the National Academy of Sciences of the United States of America*, *vol 102, n°46, pp. 16569-16572*.

Katz S.J. (1999). **The self-similar science system**, *Research Policy*, *vol 28, n°5, pp. 501-517*.

Larédo P., Callon M., Mustar P., Birac A.M. and Fourest B. (1992). **Defining the Strategic Profile of Research Labs: the Research Compass Card Method.** In: Van Raan A., de Bruin R., Moed H., Nederhof A. and Tijssen R. (eds). *Science and Technology in a policy context*, Leiden: DSWO Press, pp. 184-199.

[3] see for example the project Observatory of European Universities within the Prime project.

Lewison G., Rippon I., de Francisco A. and Lipworth S. (2004). **Outputs and expenditures on health research in eight disease areas using a bibliometric approach, 1996-2001**, *Research Evaluation*, *vol 13, n°3, pp. 181-188.*
Liu N.C. and Cheng Y. (2005). **Academic ranking of world universities - methodologies and problems** in: *1st Intl Conf. on World-Class Universities (WCU-1)* Shanghai (CHN), June 16-18, 2005

Moed H.F. and Van Leeuwen T.N. (1995). **Improving the accuracy of institute for scientific information's journal impact factors**, *Journal of the American Society for Information Science*, *vol 46, n°6, pp. 461-467.*

Murugesan P. and Moravcsik M.J. (1978). **Variation of the nature of citation measures with journal and scientific specialties**, *Journal of the American Society for Information Science*, *vol 29, pp. 141-155.*

Pinski G. and Narin F. (1976). **Citation influence for journal aggregates of scientific publications : theory, with application to the literature of physics**, *Information processing and management*, *vol 12, pp. 297-312.*

Price D.J.d.S. ***Little science, big science.*** New York: Columbia University Press, 1963.

Rinia E.J., van Leeuwen T.N., Bruins E.E.W., vanVuren H.G. and van Raan A.F.J. (2001). **Citation delay in interdisciplinary knowledge exchange**, *Scientometrics, vol 51, n°1, pp. 293-309.*

Rousseau R. (1990). **Relations between continuous versions of bibliometric laws**, *Journal of the American Society for Information Science*, *vol 41, pp. 197-203.*

Schubert A., Glaenzel W. and Braun T. (1988). **Against absolute methods : relative scientometric indicators and relational charts as evaluation tools**. In: Van Raan A.F.J. (ed). *Handbook of Quantitative Studies of Science and Technology*, Amsterdam: Elsevier, pp. 137-169.

van Raan A.F.J. (2000). **On growth, ageing, and fractal differentiation of science**, *Scientometrics*, *vol 47, n°2, pp. 347-362.*

Watts D.J. and Strogatz S.H. (1998). **Collective dynamics of "small-world" networks**, *Nature*, *vol 393, pp. 440-442.*

Wouters P. (1997). **Citation cycles and peer review cycles**, *Scientometrics*, *vol 38, n°1, pp. 39-55.*

Zitt M., Ramanana-Rahary S. and Bassecoulard E. (2003). **Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation**, *Scientometrics*, *vol 56, n°2, pp. 259-282.*

Zitt M., Bauin S. and Filliatreau G. (2004).**Indicateurs bibliométriques des institutions publiques de recherche françaises (2000 -hors sciences humaines et sociales)**, Production coopérative d'indicateurs de politique scientifique des Institutions de Recherche françaises, *41*.p.

Zitt M., Ramanana-Rahary S. and Bassecoulard E. (2005). **Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation**, *Scientometrics*, *vol 63, n°2, pp. 373-401.*

Zitt M. (2005). **Facing diversity of science: a challenge for bibliometric indicators - comments on A. Van Raan's focus article**, *Measurement: Interdisciplinary Research and Perspectives*, *vol 3, n°1, pp. 38-49.*

Zitt M. and Filliatreau G. (2006). **Being a world-class university: Big is (made) Beautiful**. In: Sadlack J. and Liu N.C. (eds). *The World-Class University and Ranking: Aiming beyond Status* Shanghai (CHN),6p + figures

Zitt M. and Bassecoulard E. (2006). **Delineating complex scientific fields by hybrid lexical-citation method: an application to nanoscience**, *Information Processing and Management, in press***.**