Analysis document – INRIA Evaluation Committee
What do bibliometric indicators measure?

Prepared by A-M. Kermarrec, E. Faou, J-P. Merlet (rapporteur), P. Robert and L. Segoufin.
Validated by the INRIA Evaluation Committee, 12 September 2007.

Abstract

The INRIA Evaluation Committee (EC) decided to examine bibliometric indicators to better understand what they measure, their relevance and their reliability, particularly in the context of research at INRIA. This document was produced by a study group led by Jean-Pierre Merlet and composed of A-M. Kermarrec, E. Faou, P. Robert and L. Segoufin. The document was discussed during the Evaluation Committee meetings held in January and September 2007 and integrates the remarks of the committee members.

Table of contents

1 Introduction

In September 2006, INRIA's Evaluation Committee (EC) initiated an examination of the elements involved in scientific evaluation. These elements include bibliometric indicators[1], figures which are increasingly used to evaluate the scientific activity and quality of researchers, teams, institutions and even countries.

Mentioning indicators in the global scientific community is sure to provoke emotional reactions based on deeply held beliefs. In the coming years, the reaction is likely to be even stronger given the heightened importance of indicators in research, whether it be in organisations, strategic orientation or funding. This is the case in France, as LOLF [constitutional bylaw on budget acts] is quite specific about which indicators will be used to evaluate organisations (see list in Appendix 1). Their use is spreading to all levels, as illustrated in a very recent report by the French inspectorate-general of finance and administration of education and research (IGAENR)[2] which provides a detailed look at the nature and use of the indicators that will be used to evaluate and manage contractualisation between research institutions and the French government.
However, it would be difficult to deny the sometimes disproportionate influence of certain research indicators, such as the journal impact factor (JIF, Section 5) or the number of citations. These metrics form the basis for staunchly held opinions, even in the national press, without any critical analysis of the figures or methodology used to produce them, nor even a perfect understanding of their definitions.

It can also be observed that these indicators play a role at INRIA, particularly in the work of the Evaluation Committee (EC). For example, certain indicators were used for a thematic evaluation report prepared by a panel whose members were all independent of INRIA. The members of the Committee can use these indicators when required to evaluate activities at the limits of their fields of expertise (e.g. evaluation feedback, entrance exam jury). INRIA's Evaluation Committee thus decided to review these indicators to better understand what they measure (and what they don't) as well as their relevance and their reliability, with a view to facilitating the Committee's internal work. For this purpose, a workgroup was given the task of drafting an analysis document, which was then discussed during two Evaluation Committee meetings in January and September 2007. The present document integrates comments from the Committee, which validated it. The analyses presented below are obviously geared toward the Institute's specific context and thus cannot be transposed to any other situation without further analysis.

2 Document outline

The outline of this document is as follows:

- The basic principles of bibliometrics are presented in Section 3.

[1] There are of course other indicators, such as those used to evaluate transfers or other applications, which are not covered here.
[2] Report 2007-012, La contractualisation des organismes publics de recherche avec l'État, April 2007.

- Most bibliometric indicators are based on citation analysis. We examine the organisations, companies and Web search engines that provide these citations (4), then present certain elements for assessing the reliability of their offerings (4.3). We also discuss the problem of attributing citations to institutions (4.4).
- Indicators based on citations are presented in Section 5. These indicators can be applied to scientific publications (mostly journals), institutions or individuals.
- The consistency of indicators used for scientific quality is examined in Section 5.6 and concluding remarks are presented in Section 5.7.
- Use of indicators in science is not neutral because it opens the door to manipulation and more generally to ethical problems, presented in Section 6.
- The cost of producing indicators is discussed in Section 7.
- Section 8 looks at the complex issue of attributing citations to an institution in order to calculate indicators for it.
- Practical examples involving INRIA researchers are presented in Section 9: we examine four senior researchers at INRIA working in different fields, the results of a citation search using the three most widely used search engines, and how the inconsistency of these results affects the indicators. We also show that the researchers' work is not attributed to INRIA in a uniform manner.

In conclusion, we suggest that while indicators can show trends for limited aspects of scientific activity, they should be used with considerable caution given the possibility of misinterpretation, measurement errors (often substantial) and the different types of bias which affect them. The misuse of indicators is facilitated by the quantitative nature of the result, which introduces the possibility of producing a variety of statistics very quickly, without taking the time to analyse their quality or content, while overlooking other elements of scientific activity, such as innovation or intellectual and industrial transfers.

Our principle observations and recommendations are as follows:

1. Measurement errors in citation searches are substantial, and this is particularly true for research conducted at INRIA. We therefore strongly recommend consultation of several sources, critical analysis of the results (specifically by calling on experts in the relevant fields) and a greater focus on an indicator's order of magnitude.
2. We recommend referring to the literature on bibliometrics to determine the real significance of indicators and their bias.
3. Indicators give only a partial and biased view of certain aspects of science, without covering the whole of scientific activity. Specialists from the scientific field must therefore complete, correct and provide feedback on them. If used, indicators should be interpreted within the context of the evaluation or decision-making process.
4. We recommend that indicators never be used to compare different scientific fields.
5. Indicators are based on citation analysis, which is not very conducive to scientific risk-taking and innovation. Misuse of indicators (or worse yet, automating their use) would be a major obstacle to innovation.
6. Individuals, institutions and other players in the scientific realm (such as journals) can easily manipulate quantitative indicators. Cases of such manipulation are on the rise, and this can be correlated with the growing impact of these indicators.
7. For the scientific fields in which INRIA works, we observed very poor coverage by the usual citation sources (Web of Science, Scopus), bordering on the absurd in certain cases. As a result, the indicators based on these sources misrepresent the Institute. By contacting these sources, INRIA might improve the attribution of citations for its own

research, but this would not change the fact that these sources are insufficient for computer science and applied mathematics in general.

3 Bibliometrics

Bibliometric indicators are based on analysis of citations, i.e. the "References" section of a scientific article mentioning related work conducted by the scientific community (including the article's authors). It should be noted that as they are citation-based, indicators only give a partial vision of scientific output since 90% of the papers published in scientific journals are never cited [9], and it is difficult to believe that only 10% of the science being done is significant.

The first step in building an indicator is thus collecting these citations; given the magnitude of scientific output, this is a colossal task and an impossible one to perform completely. So the professionals who work in this area limit themselves to a select part of the literature produced throughout the world (see Section 4). The second step in building an indicator is to define the mathematical formula that transforms the raw citation data into a quantitative evaluation. Meta-indicators can also be built, which combine the quantitative evaluations of various indicators to obtain other indicators (used, for example, for the Shanghai academic ranking of world universities).

These indicators and the methodology used to calculate them raise the same questions a physicist faces in examining the results of a measurement device:
1. What quantity is actually measured by the device?
2. What is the margin of error for this measurement?
3. What is the measurement's quality/cost ratio?
As M. Zitt of INRA notes [18]: "At the same time, more precautions are needed, as 'desktop' bibliometrics may reinforce a danger already present in less reactive scientometrics: the contrast between the highly sensitive nature of evaluation issues and the eagerness of users or clients to elicit a figure or a display and forget crucial warnings about statistical distributions and methodology artifacts. If scientometrics is a mirror of science in action, then scientometricians' particular responsibility is to both polish the mirror and warn against optical illusions".

As citations are key to producing indicators, we must examine where they come from and the actors in this field (who generally don't limit themselves to collecting citations but also offer indicators). The methodology used to collect the citations must also undergo critical analysis. In France, the main purveyor of indicators is the Observatoire des Sciences et des Techniques (OST)[3]; see Appendix 2.

4 Citation sources

4.1 Sources with fee-based access

Using Web search engines, two types of searches can be performed for scientific publications:

[3] www.obs-ost.fr

1. Article retrieval: Search engines like PubMed Central (PMC) for medical fields or Science Direct can be used to find articles and retrieve the full text, most often for a fee.
2. Citation search: Mainly involves searching a database for articles that cite a given author or institution in their "References" section. These are the searches used to build the citation database, which is then used for the indicator calculations.

We will focus only on the second type of search engines here.

4.1.1 Thomson Institute for Scientific Information (ISI)

The private company Thomson ISI is a key player in the area of indicators, and also the most long-standing. Founded in 1960, the company publishes its *Journal Citation Reports* (*JCR*) annually. *JCR* includes a number of indicators, such as the journal impact factor (JIF), and comes out 6 to 7 months after the end of the year.

Web of Science (WoS) is the Web-based source for the citations used by ISI for *JCR* (its paper version is the *Science Citation Index* or *SCI*). The expanded version is said to contain a total of 38 million citations dating back to 1900. The WoS includes analysis of 8700 journals covering the "hard" sciences (80%) and the social sciences (20%), a small number of which (190) are freely accessible. The content is updated every week. It should be noted that equal coverage is not given to all fields. For example, the life sciences are not as well represented as in biomedical databases such as Medline. It's also acknowledged that fields closer to societal needs or to applications (engineering sciences in particular) are not always as well covered as fundamental research. WoS coverage only represents a small percentage of the scientific literature worldwide; in 1999, the total number of scientific journals was estimated at 100,000, with 25,000 for the biomedical field[4] alone. Finally, the majority of conference proceedings are not covered by WoS. The same goes for articles published in open access journals, open archives or on personal pages.

WoS is often accused of having an American bias. In 2005, 98.74% of the articles in WoS were in English, 0.234% in French, 0.205% in Chinese and 0.09155% in Japanese [7]. It is obviously easier to cite articles from one's own country[5], a bias which has been confirmed in specific cases (see Section 4.3).

It is important to mention here how ISI defines a citation. All material published by a journal and cited as a reference counts as a citation for this journal. This applies to articles of course, but also to editorials, book reviews and letters to the editor.

ISI indexers assign a document type to each item published in a journal, and sometimes in an inconsistent manner[6]. The journals themselves are classified into categories. Some may be assigned two categories and in rare cases, three categories. For all authors of a given publication, the indexers fill in the following WoS fields:

- Main Organisation: single field
- Sub-organisations: 3 separate fields
- City
- Province/State
- Country
- Post code

[4] G. Roelants, INFI Newsletter, December 1999.

[5] A study showed that on average, the percentage of foreign references cited by American biomedical authors was 12% for articles less than 2 years old and 30% for articles less than 10 years old (figures which greatly underestimate the real impact of non-American literature). N. Pinhas, Inserm Actualités, 154, September 1997.

[6] "It further complicates the situation that the same genre of item is inconsistently classified", Péter's Digital Reference Shelf.

The choice of a single main organisation is not a neutral one; for example, this field serves as the basis for the Shanghai ranking.
In contrast, the number of articles published in a year by a given journal only includes research articles. It is also interesting that ISI does not count references to books as citations[7] (by default, querying the ISI database for a given author will not account for the books this author may have published). ISI tries to avoid counting self-citations (i.e. citations by the author(s) of their own papers) by simply eliminating cited papers where the first author is one of the citing authors (meaning that self-citation between members of a group might be very under-evaluated).
The citations themselves are attributed to all authors. Addresses are standardised, but the process is far from perfect.
A few key remarks by Thomson on *JCR*[8] are worth noting:

"There has been a guide for a long time in the help file about how to use the *JCR*. Many ignore it and use it in a way it was not meant to be used, such as faculty tenure position, or used alone...
There are excellent educators who do not necessarily publish in journals processed by ISI or do not research in a field where citations are profuse...
I just wish a better handling of the citable items, and plausibility checks of the scores by experts who know well the journals of a discipline to spot errors at a glance".

[7] ISI justifies this by the existence of edited books, where each chapter is written by a different author. As only the names of the editors are mentioned in the citations, the authors of the chapters would be penalized: "For books, the bibliographic information presented in an article's reference list often contains the first author only, or a partial list of authors…Linking to the source item allows us to expand the reference to include the complete list of authors. Lacking this expanded information would decrease the accuracy of our analysis by under-representing the contribution of additional authors". It might however be noted that by far not all books are edited books.
[8] Péter's Digital Reference Shelf

The difficulty of attributing citations to authors (due to different practices between communities) but also to institutions should be highlighted as well. It is clear that institutions with multiple partnerships such as INRIA, where authors can be associated with several institutions, are going to be penalised. Indeed, a WoS search shows considerable differences in the name of the author's organisation, which doesn't facilitate matters for the indexers.
Articles which could be attributed to INRIA unquestionably are instead attributed to LORIA, IRISA, INRIA Sophia, etc. This problem of address standardisation in France is discussed in Section 8.
Finally, it should be noted that ISI is not simply a purveyor of citations; it also supplies the vast majority of bibliometric indicators such as JIF (Section 5).

4.1.2 Scopus

Scopus was launched in autumn 2004 by Elsevier and contains 30 million citations, which nonetheless only go back to 1996 (1900 for WoS). They are taken from 15,000 peer-reviewed titles, 12,850 journals including 1000 with open access, 700 proceedings, 275 million Web pages and 125 book series. Scopus is updated daily and offers more exhaustive coverage than WoS for engineering. It also appears to be more user friendly than WoS [1,2]. Citations appear to be distributed as follows [2]:

Health and life sciences: 25.3 million documents
- Engineering: 8.5
- Agriculture and biology: 3.6
- Earth and environmental sciences: 1.9
- Chemistry: 1.3
- Physics: 0.59
- Social sciences: 0.29
- Mathematics: 0.26
- Psychology: 0.23
- Economics: 0.22

The geographical distribution of the titles is more varied than for WoS, with 60% originating outside the US. Until 2003, only the first author's organisation was used, but since then all authors' organisations have been integrated. A special effort has been made to better manage homographs by accounting for name variants and by using data such as affiliation and subject matter to better define the author.
From a bibliometric standpoint, Scopus offers richer statistics than WoS, but it doesn't allow analysis by institution. In any case, it is a serious competitor for WoS.

4.1.3 Common problems for fee-based resources

The problem is essentially related to the corpus, which is not always well characterised. We have seen that this corpus represents only a small proportion of scientific output. Consequently, the choice of journals and conference proceedings is significant and will influence the citation indicators calculated from these sources. This can favour one sub-theme at another's expense, or even completely distort the results. For example, at the time of this writing, it appears that WoS contains all the proceedings published in the LNCS series but nothing published by ACM. In fields where good conferences are part of the ACM library and low-quality conferences are published via LNCS, an indicator calculated from WoS data gives a doubtful result. See Section 4.3.

4.2 Free sources

The predominance of WoS and Scopus and the practices associated with them has provoked reactions in the scientific community, leading to the development of other citation search and evaluation methods.

4.2.1 CiteSeer

CiteSeer is a specialised computer science database providing sophisticated indexing of citations as well as a system for ranking computer science conferences, journals and authors.
The EC compared the ISI 2003 ranking of computer science journals, based on JIF, with the corresponding CiteSeer ranking. Figure 1 shows CiteSeer's ranking of the 20 best journals according to ISI, whereas Figure 2 shows ISI's ranking of the 20 best journals according to CiteSeer.
The figures show very clear and surprising discrepancies between the rankings. We considered 20 journals ranked by ISI (which ranks 200) as the most important in the field by examining how CiteSeer ranked these same journals (CiteSeer ranks 1200 journals and conferences). The top-ranking journal according to ISI, *ACM Computer Surveys*, is in 195th place according to CiteSeer; the second-place ISI Journal, *Bioinformatics*, isn't in the CiteSeer ranking and the 6th ISI title is 958th in CiteSeer... Inversely, the top CiteSeer journal, *ACM Transactions on Computer Systems*, is 26th according to ISI, and the 4th CiteSeer journal, *Computer Networks*, is 122nd.

4.2.2 Google Scholar

The stated goal of Google Scholar is to include "peer-reviewed papers, theses, books, preprints, abstracts and technical reports from... academic publishers, professional societies, preprint repositories and universities... available across the Web". However, neither the sources nor the coverage dates are provided. Google offers different types of indicators based on Scholar sources:

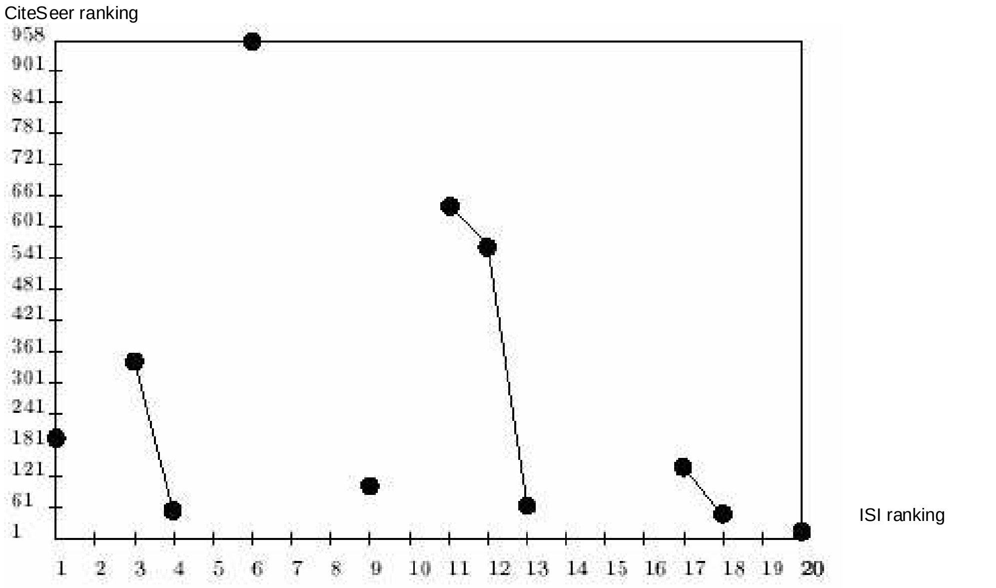

Fig. 1 - The Y-axis shows the CiteSeer ranking of the 20 best computer science journals according to ISI. Note that certain ISI journals aren't included in the CiteSeer ranking (and vice versa).

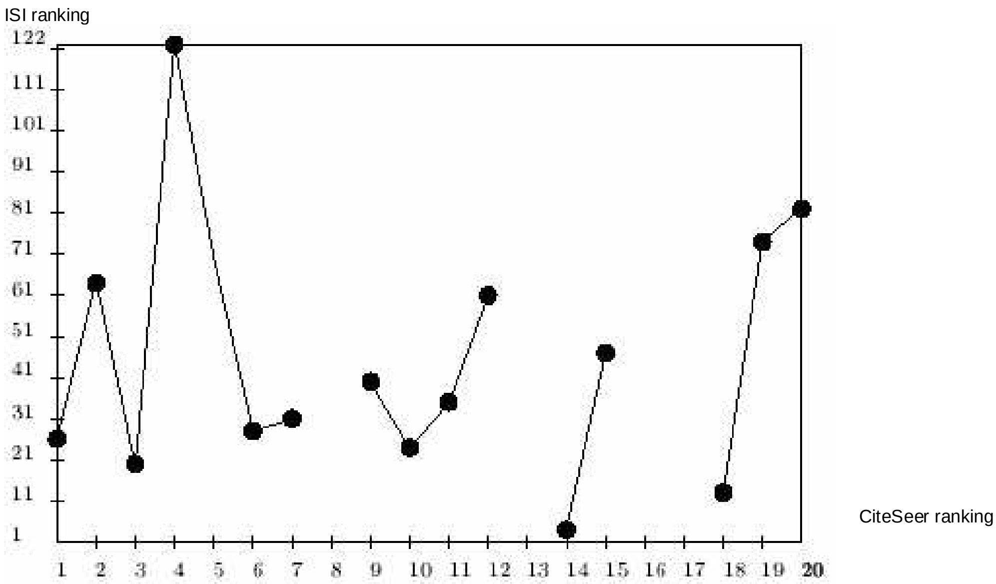

Fig. 2 - The Y-axis shows the ISI ranking of the 20 best computer science journals according to CiteSeer.

- Web documents
- article citations
- book citations

Web documents are links to Web pages which either describe documents or provide access to them. Web document bibliographies are used to count citations.

Different analyses and comparisons with other search engines have shown that:

- Google Scholar's coverage varies considerably from one field to the next. Certain studies indicate that the results are well below what can be obtained using more specialised databases, but this was not the case in our study (see Section 9).
- The documents used as citation sources by Google Scholar are even less well characterised than those used by commercial databases.
- Like the citation count, the indicators obtained are very inaccurate.
- There is only one output format (per citation score) whereas WoS offers a more detailed analysis (number of publications and citations per year, self-citations eliminated, etc.).
- The database is not regularly updated (recent documents will only be found in Google Web).

Google Scholar's poor management of homographs led Anne-Will Harzing to develop a software application called Publish or Perish[9], which retrieves Google Scholar data but performs a more detailed analysis and calculates several indicators:

- Quantitative: Total number of papers or citations, mean number of citations per paper or per author, number of authors per paper.
- Individual: g-index, h-index (two variants), citation rate weighted by publication dates.

4.2.3 Citebase

In 1999, Harnad and Brody began experimenting with a citation algorithm which differed from the ISI algorithm and aimed to produce more objective indicators. This culminated in the creation of the Citebase[10] tool. In addition to citations, this model counts the number of article downloads (but only uses data available in a limited number of open archives like arXiv) as well as the time between downloads and citations to better define an article's scientific quality.

4.2.4 Common problems for free resources

The flaws common to the free resources stem from the fact that the data are retrieved (most often via the Internet) and processed automatically. Below are a few of these flaws.

1. Identification of individuals. It is often impossible to differentiate between people with the same name. Homographs are very poorly managed: a citation search for the author D. Cohen also retrieves citations from authors with the same last name and a first name starting with D. On the other hand, sometimes the papers and citations of a single author are divided into two separate entries[11]. Accented names or those that don't have a first name/last name format may also be problematic since ISI stores author names exactly as they are found (while allowing the storage of variants, particularly for Asian names). This means that the frequent case of missing accents in a citation will penalise the author cited.

---

[9] www.harzing.com

[10] www.citebase.org

[11] For example, if your name contains "fi", try your real name, then your name without the "fi". This works in Google Scholar as well as CiteSeer. It is due to the inability of certain PDF documents to manage these two letters in the form of a ligature, the letter "i" disappearing from the document.

2. Identification of articles. As articles are often automatically retrieved via the Internet, merging the various versions published by all the co-authors on their personal pages is not a simple matter. Consequently, it isn't rare to see two or more entries in Google Scholar or CiteSeer for the same article, which of course distorts the citation count.
3. Identification of journals. It is possible in theory to search by journal name, but you must provide the journal name yourself. Given the considerable number of abbreviations used for journals, it's difficult to actually access all volumes of a given journal.

All these problems occur very frequently, which means that any indicator calculated from these sources is not very credible.

4.3 Comparing the sources

An article published in 2007 [9] studied the citations obtained for 25 well-respected computer science researchers using WoS, Scopus and Google Scholar. Scopus offers 35% more citations than WoS and 160% more than Google Scholar (these differences vary enormously according to discipline). The Google score should be weighted to account for source and citation inaccuracies, but it is also clear that searching Web sources better integrates the new publishing habits of researchers.

Recent studies have focused on the citations of the 25 most cited authors in computer science, crosschecking between CiteSeer, Scopus and Web of Science [10], with the finding that only 44% of the citations were common to all three databases. A similar difference was observed by Burnham [1] in the medical field[12].

Nisonger (2004, see article by Meho) examined his own citations in Web of Science. The database contained 28.8% of the total citations of his articles, 42.2% of his journal article citations, 20.3% of his citations outside the US and 2.3% of his non-English citations.

Meho [10] published an exhaustive comparison of WoS, Scopus and Google Scholar, using all the citations for his 15-member laboratory (School of Library and Information Science) as the search set, hence a pre-established set. The first point of comparison was how easily reliable data could be obtained. WoS clearly came out ahead, with its more reliable data and powerful tools; Scopus required twice as much time, and Google Scholar took 30 times longer (in addition to not including 12% of the laboratory's citations). Meho also observed that having the list of citations beforehand was useful because of the numerous homographs. He was able to correct citations attributed to other authors having the same or similar names. For the fields considered, Scopus provided 13.7% more citations than WoS, and the total citations from WoS and Scopus together represented 35.1% more citations than for WoS alone. Even more troubling is the high variability of this percentage depending on the research field, with a range of 4.9%–98.9%. Only 58.2% of the citations were common to WoS and Scopus, mainly because Scopus provides in that case considerably more conference citations than WoS.

As for Google Scholar, it found 53% more citations than WoS and Scopus combined, bringing the number of citations found to 93.4%. Although some of these citations came from journals with a low JIF, a large percentage of them were from conferences (and were mainly found on the authors' personal pages). But once again, the increase in the number of citations found depends very strongly on the field (from 143.6% for human-computer interaction to 23.4% for bibliometrics).

The overlap between Google Scholar and the WoS/Scopus combination is very low (30.8%), and Scholar does not find 40.4% of the WoS and Scopus citations. However, introducing

[12] Number of citations found for Arthritis and environment, 1992-date: 405 (WoS), 395 (Scopus); for Tubercular meningitis, 1992-date: 21 (WoS), 76 (Scopus); for Barik S*, 1992-date: 93 (WoS), 126 (Scopus) and for American Journal of Cardiology: 1992-date: 15,619 (WoS), 21,993 (Scopus).

Scholar data does not noticeably modify the researcher ranking obtained using the WoS/Scopus combination.
Finally, Meho notes that Google Scholar provides much better coverage of non-English literature than WoS or Scopus, finding 6.95% of the relevant citations (compared to 1.14% and 0.7% for WoS and Scopus respectively).
The Center for Science and Technology Studies (CWTS) at Leiden University has observed that conference proceedings are major means of dissemination in computer science. They thus undertook to complete WoS by adding Springer's Lecture Notes in Computer Science (LNCS), ACM conferences and those of the IEEE Computer Society, thereby increasing the size of the database by 66% [12]. However, this addition required considerable work, particularly to extract data from PDF files. To measure coverage quality for this extended database, CWTS examined its articles' citations of other articles in the database, a measure of internal coverage. A 100% finding would indicate that overall, the database articles refer only to other articles in the database. The internal coverage turned out to be 51% (a substantial increase given that internal coverage for computer science items in WoS is 38%, much lower than for physics or chemistry where it exceeds 80%), but this is still "moderate" according to the authors, who consider it unlikely that adding conferences could increase this coverage to more than 80%.
Dutch laboratory directors responded to this initiative by citing publications they considered important which weren't in the extended database. CWTS therefore recommended in its report that a bibliometric study in computer science should be based on a list of publications provided by researchers themselves or at least checked by them (this is in line with the EC's current effort to identify journals and conferences considered important by the Institute's researchers). The second recommendation stems from the diversity of citation sources (e.g. beyond the standard journals and conferences). It is very difficult to automate the tally of citations, thereby necessitating a manual analysis. This represents a very difficult task as demonstrated in Section 9, which presents the results of a preliminary study that, while it only involved four INRIA researchers, confirmed the conclusions of Meho. In fact the results were even worse.
The inconsistency of these sources clearly casts doubt on the value of indicators based on them. The question can be raised as to the validity of publishing a JIF with 3 decimal places, given that over 20% of the citations in the journals processed may have been omitted (a conservative figure, according to Meho's article). This uncertainty comes into play right from the first decimal place in JIF.

4.4 Problems of attribution

Indexers are confronted with two types of problems when they examine articles:

- To which organisation should the article be attributed: The only identifying elements are author addresses, which are so complex they must be dealt with manually (limiting compatibility with the automated data processing offered by citation sources). This problem is addressed in Section 8.
- For articles with several authors, attributing the work between them: The simplest solution is to divide a publication equally between the authors (this appears to be the method used by WoS[13]). It is nonetheless very difficult to obtain the rules used to attribute citations. An effort can also be made to account for publication practices, but there is extreme variation from one field to another, making automatic processing very difficult.

5 Indicators

There is a multitude of bibliometric indicators and their number is growing steadily (likely a corollary of their increased use)[14]. Hence our aim is not to make an exhaustive inventory, but rather to examine the most utilised indicators, those which have undergone in-depth analysis. We have also attempted to describe trends for new indicators.
All these indicators are based on one citation source. They obviously inherit their source's problems, mentioned in the preceding section.

5.1 Indicators for journals

Evaluating an article's scientific quality is a delicate problem. A simple approach is to link its quality to that of the medium in which it was published. In this way, evaluating the medium (a scientific journal in most cases) takes the place of an individual evaluation, which of course simplifies the task enormously because there are far fewer media than articles. This was ISI's initial model, which has been forced to evolve to some extent under pressure from customer organisations.

5.1.1 Journal impact factor

The journal impact factor (JIF) is an indicator made available by ISI in its *JCR*. It is calculated from WoS. The concept was initially invented in the early 1960s by Gene Garfield, founder of ISI. A journal's JIF for year n is defined as the ratio between the number of citations during year n of the journal's articles published during years $n–1$ and $n–2$ (and only then), and the total number of articles published during these two years:

$$JIF_n = \frac{C_{n-1} + C_{n-2}}{P_{n-1} + P_{n-2}}$$

[13] "Citations to each article from any other article in the dataset are counted, and each indexed author is credited with the full tally of citations to that article. Thus, an article with ten authors which has been cited 100 times, will attribute 100 citations to each of the ten indexed author names".

[14] Ample literature on this subject is available at indicasciences.veille.inist.fr.

Limiting the articles to these two years is somewhat mysterious[15]. JIF is often considered an indicator of journal quality and plays a considerable role in the scientific community[16].

However, ISI indicates that JIF should not be used for different fields. In particular, the JIF for a journal in a field requiring long studies will automatically be lower than for journals covering rapidly evolving fields. This has been shown to explain why journals in molecular biology (a field where articles rapidly become obsolete) have much higher JIFs on average than mathematics journals [15]. In 1999, the best JIF in mathematics was that of the 51st cell biology journal, and only 4 references out of the 84 in Andrew Wiles' paper on Fermat's Last Theorem were to items published in the previous two years [16]. Recent analyses show that this trend has not been reversed. For example, a JIF study of 181 mathematics journals and 124 genetics journals based on *JCR* 2005 data showed that while the distribution was comparable for the two disciplines, the mean JIF value varied by a factor of 10, with genetics coming out ahead [7][17].

Another logical argument against using JIF for comparisons between fields is that journals in a field with little scientific activity may nonetheless have similar JIFs to journals in a field with a high level of research, since JIF depends only on citations. Likewise, a very active field may have journals with a low JIF due to different citation practices[18] or a limited community. Indeed, community size (measured as the number of articles published) appears in many cases to considerably influence JIF. This is not true for all fields, but a number of other factors may arbitrarily influence JIF: e.g. an increased number of articles published in a journal where research is very active will usually lead to a lower JIF because the denominator increases while in most cases, the papers will be cited outside the timeframe used to calculate the JIF numerator [16]. Even the type of material published in a journal can introduce bias in JIF, estimated at 5%–40% (see Section 6.1).

Moreover, JIF measures a mean for the journal's citations, but it should be noted that even for journals with high JIFs the citations come from at most 15% of the articles published [15]; in other words, JIF does not really measure quality for a specific article or author.

As it is based on WoS, JIF is often accused of having an American bias, even in the English-speaking community[19]. This is due to the fact that authors more readily cite articles from their own countries[20], a bias which has been confirmed in specific cases (see Section 4.3).

And while life sciences are not as well covered by ISI (see Section 4), they are in a class of their own in terms of JIF. Twelve of the 15 journals with a JIF above 10 are in these disciplines. The relatively general nature of life science journals gives them a clear advantage over the high-quality specialised journals that are rarely cited outside their own communities.

5.1.2 Immediacy index

[15] ISI has recently released a guide for calculating JIF using a 5-year window.

[16] "Editors get hired and fired soon after the annual JIF data are published", Péter's Digital Reference Shelf, August. Published by Thomson Gale, which is an affiliate of Thomson Institute for Scientific Information (ISI).

[17] E. Garfield has drawn attention to the absurdity of comparisons based on JIF: "[I]t is absurd to make invidious comparisons between specialist journals and multi-disciplinary general journals like *Nature* and *NEJM*. To compare journals you should stick to a particular category as is explained very carefully in the Guide to JCR". Der Unfallchirurg, 48(2) p.413, June 1998.

[18] For example, the mean citation rate in 2000 for pharmacology was 11 whereas in genetics it was 28 [13].

[19] Guide entitled *Journal Impact Factors*, June 2004, University of Auckland, New Zealand, www.boulder.nist.gov/div853/Publication%20files/journalimpactfactors.pdf.

[20] A study showed that on average, the percentage of foreign references cited by American biomedical authors was 12% for articles less than 2 years old and 30% for articles less than 10 years old (figures which are far below the real impact of non-American literature). N. Pinhas, Inserm Actualités, 154, September 1997.

ISI also makes this indicator available in its *JCR*. It is defined as the ratio between the number of citations to articles published during year *n* (and only then), and the number of articles published in the journal the same year:

$$II = \frac{C_n}{P_n}$$

This index is often considered to measure a journal's immediate impact. However, in many cases journals with a high immediacy index obtain this figure because of the high number of references to editorials, which are not included in the indicator's denominator.

5.1.3 Cited half-life

ISI also makes this indicator available in its *JCR*. For year *n*, the cited half-life is the number of years *j* such that 50% of the journal's citations for year *n* are citations predating year *n–j*, with the other 50% published thereafter. For example, *Nature Genetics* had a cited half-life of 4.7 in 2003 because 46.38% of the citations for this year predated 1999. In other words, this indicator provides information about the long-term value of research in a given field.
Indicators such as JIF only account for relatively recent citations, which means that journals with shorter cited half-lives will automatically have higher JIFs than those with longer cited half-lives.

5.2 Downloads

The trend amongst publishers toward online access has provided another indicator: the number of downloads. The advantage of this indicator is that it can be obtained in real time, and according to Meho, a correlation has been established between the number of downloads and the number of citations, although the degree of correlation varies significantly depending on the discipline. Under these conditions, the number of downloads would provide an initial estimate of the number of citations to expect from articles.

However, certain restrictions apply to this indicator:

- It is difficult to establish for a given author, who can publish in a wide range of journals, because it would be very costly to examine all journals in a database for each author.
- It does not account for the new means of dissemination used by researchers (personal pages, open archives).
- The practice of indicating the most downloaded articles, used by certain publishers, automatically gives these articles an advantage.
- Reliability is relative. These indicators are calculated by publishers themselves, resulting in a clear conflict of interests.

5.3 Quantitative indicators

These indicators are the easiest to determine using citation databases.
Here are some examples:

- Number of publications and citations for a group of researchers
- Number of publications and citations per researcher for a group of researchers

- Percentage of worldwide output
- Number of publications used for ISI indicators
- Number of publications in high-JIF journals

It goes without saying that the first three indicators don't give any information about the quality of the scientific work; at most, they allow evaluating whether the group has a "normal" level of publication, which must be considered relative to the average level of other groups working in the same field. The validity of the following two examples is correlated to the ISI indicators for the field considered.

In their report, the CWTS of Leiden University make an interesting observation for computer science [12]: peer-evaluation of 42 Dutch computer science laboratories, as carried out by the Review Committee for Computer Science of the Quality Assurance Netherlands Universities (QANU), does have a positive correlation with the indicators mentioned above, but it is nonetheless a weak correlation. This observation raises the following question: Is peer-evaluation a poor measurement of a laboratory's impact, or are the indicators poorly suited to measuring this impact?

5.4 Individual indicators

The organisations which use indicators are looking for measurements to individually evaluate researchers, which is not the objective of journal indicators[21]. In this context, journal indicators only give them an indirect view, which doesn't lead to the quantitative values they seek. Researchers object as well, noting that journal indicator analyses based on averages cannot reflect the quality of a particular element. There are many examples of articles published in a journal with a low JIF that nonetheless constitute a major contribution to modern science and inversely, articles of poor quality, or those only aimed at stirring up controversy, published in journals with high JIFs. Consequently, there is a current trend (often criticised by professional bibliometricians) toward offering indicators which supposedly evaluate the scientific quality of an individual's work.

[21] E. Garfield is very clear on this point:"The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the Habilitation process. In many countries in Europe, I have found that in order to shortcut the work of looking up actual (real) citation counts for investigators the journal impact factor is used as a surrogate to estimate the count. I have always warned against this use". Der Unfallchirurg, 48(2) p.413, June 1998.

5.4.1 The h-index

The h-index for authors was defined by J.E. Hirsch [6] as the number of articles *h* by a given author that have been cited at least *h* times each. It was proposed as an alternative to other indicators (for which the advantages and drawbacks have been summarised from Hirsch's paper):

- Total number of papers: Measures productivity but not impact.
- Total number of citations: Measures a form of impact but can be very strongly influenced by the number of co-authors and by review articles.
- Citations per paper: Enables comparisons between researchers of different ages; difficult to estimate and tends to reward low productivity while penalising high productivity.
- Number of significant publications: Number of papers cited more than *y* times; lacks the drawbacks of the above indicators but the arbitrary choice of *y* is a disadvantage (measurement difficulties could also be mentioned).
- Number of citations in the *q* most cited papers: Lacks the drawbacks of the indicators above, but difficult to obtain and difficult to determine *q.*

Hirsch indicates that the h-index can be easily obtained from Web of Science by using the "times cited" ranking made available by ISI. But this assumes good coverage of all scientific fields (or at least uniform coverage for comparing individuals) by Web of Science, which is far from the case for INRIA's disciplines. Hirsch goes even further by proposing h-index values for obtaining tenure, becoming a full professor or a member of the US National Academy of Science. It's only at the end of his paper that he briefly addresses the problem of h-index values according to discipline ("h-indices in the biological sciences tend to be higher than in physics... more research in understanding similarities and differences... in different field of science would be of interest").
Among the numerous problems raised by this indicator, there is the reliability with which it can be measured. ISI initially refused to make this indicator available, leading to "makeshift" tools based on Google Scholar, which in addition to its poorly identified sources, does not manage homographs and self-citations very well. For authors with relatively common names, the results obtained via Google Scholar are often outlandish (without doing an exhaustive search, it can be shown that authors initially qualified with a very high h-index of 35 had this value fall to 5 (average) as soon as the citations found in Google Scholar were examined more carefully).
Meho claims that the h-index can now be easily calculated from Web of Science (in October 2006, ISI changed policies and made the h-index available in its *JCR*), Scopus and Google Scholar. We do not share this opinion: the diverse results from these three databases require very detailed crosschecking and a manual analysis of the data's validity before a more or less correct value relative to the indicator's definition can be obtained (without ensuring its validity). Moreover, the sources used do a very poor job of managing references to books or book chapters, strongly penalising book authors. For example, certain authors who have written very few articles but whose books are acknowledged have very low h-indices. In such cases, only the maximum citation indicator provided with the h-index gives some sense of the discrepancy between the index value and the author's real influence. As a result, the citations themselves must be examined with great care.
A critical analysis of the h-index was published by H.L. Roediger [14], revealing the following characteristics:

- The h-index is correlated with age.

- Lag: The h-index can increase substantially even if the researcher is inactive for a long period.
- The h-index is underestimated for researchers who have published books.
- The h-index doesn't highlight an author's important achievements.
- Because citations are attributed to authors, there is no way to account for practices in different fields, which are very diverse. The resulting author rankings often reflect the significance of their contributions. But Roediger, a psychologist, indicates that in his field the achievements of the first and last authors are the most significant.

In other fields, alphabetical order is the rule, or the first author is the one who has made the most important contribution.
There are other major disadvantages of the h-index: it does not account for negative citations (i.e. very critical of another paper); nor are highly cited works or the total number of citations taken into account.

5.4.2 Variants of the h-index

To make up for the shortfalls of the h-index, variants have been proposed:

- a-index: The mean number of citations for the articles used to calculate the h-index.
- g-index: Proposed by Egghe [3], it's the number of articles g for which the sum of the citations is at least g^2 (a g-index of 10 indicates that the author has written 10 papers for which the sum of the citations is at least 100).

Incidentally, Hirsch encouraged the development of these alternative indicators. The multiplication of indicators without a minimum of critical analysis can nonetheless be seen as casting substantial doubt on the existence of a relationship between the indicators and scientific quality.

5.5 Other indicators

More sophisticated indicators, also based on citations, have been proposed to account for three possible biases:

- Year of publication: Older publications are more often cited.
- Type of document: The number of citations varies considerably according to document type, e.g. review articles are generally more cited than research articles.
- Field: Publication practices differ considerably according to the scientific field.

The indicators proposed are based on a standardisation process which attempts to correct these biases. The most well-known among them include the field-normalised citation score, one form of which is the crown indicator, which compares the mean number of citations attributed to a "unit" (researcher, laboratory) to the mean number of citations in international publications of the same year, in the same field and for the same type of document. A crown indicator of 0.9 means that the publications analysed are cited 10% less than the world mean.
Another of these indicators calculates, for a group of authors, the proportion of their publications found in the top 5% of the most cited publications in the world for the same year, in the same field and for the same type of document. A value above 1 indicates that the group has more publications in the 5% most cited category than the world mean.
The commendable effort to better account for the specific characteristics of different scientific fields with these indicators runs up against a few difficulties:

- Defining the field is highly subjective (what level of granularity is necessary to accurately reflect a field?)
- Data processing is necessarily manual (for example, a journal can cover several fields, so its citations must be sorted).
- The considerable inaccuracy of citation sources is seen at two levels: for the calculation of the world mean and the citations of the group considered.

The use of citations alone to evaluate an article's scientific quality and impact has raised many questions in the scientific community, which is obviously most keenly aware of the biases this methodology can create. These questions have led to initiatives proposing other models. As an example in biology and medicine, BioMed Central's Faculty of 1000[22] offers article analysis based on cooperative reading by a group of specialists selected for a specific field.

5.6 Complementarity and consistency of the indices

A study of *JCR* indicators for a 19-year period [8] (the study is relatively old) showed that journal indicators could be split into four groups: low, central, high and extreme (journals with very different indicators than other journals). The high and extreme groups only represent 10% of the total, whereas 50% are in the central group. Thus, in 1994, 6.7% of the journals had a JIF higher than 3, 27% had a JIF between 1 and 3, and the rest had a JIF below 1. It should be noted that the journals INRIA researchers publish in are nearly all found in the 2nd and 3rd groups, therefore having JIFs between 1 and 3. Very little attention should be paid to the decimal places in indicators such as JIF; from a strictly methodological point of view, only the first number makes any sense in terms of measurement error.

Concerning consistency, there are discrepancies between expert opinion and the rankings one can establish based on ISI indicators. For example, in 2005 the robotics teams at INRIA ranked the journals in their field[23] and submitted the result to the EC. This ranking differs significantly from what might be established using the *JCR* indicators. A 1995 study [5] compared the ranking established by 50 NIH researchers with ISI's JIF. The top 10 journals according to the experts had a JIF-based ranking that ranged from 3 to 160, effectively demonstrating that JIF characterises the visibility of research rather than the quality of scientific output. A similar discrepancy between the ISI and the CiteSeer rankings in computer science was presented in Section 4.2.1.

Well-respected researchers have even proposed a moratorium be placed on using indicators for evaluation until they have been completely re-assessed and revised [4].

[22] www.facultyof1000.com

[23] This ranking obviously accounts for the specific characteristics of research conducted at INRIA, and it is very conceivable that another ranking would be established in a different context.

5.7 Conclusions on indicators

There are two problems for each indicator. The first involves the meaning of the indicator itself and how it should be interpreted. The second involves how it is calculated and whether the number obtained is in keeping with the number expected.
Answering the first question would require an investigation surpassing the scope of this report. We can however ask why JIF only accounts for citations in the two years following the publication of an article (ISI describes now a procedure allowing calculations for a window of 5 years). Why is 2 more relevant than 1, 3 or any other number? The h-index is a function which may seem ad hoc, but why would this function be better than some other function? For example, a paper's significance can be calculated from the significance of the papers citing it, thereby setting a point of reference; a person's significance can then be obtained from the papers he/she writes. In this way, thousands of functions can be imagined, and only through an in-depth study could the most relevant possibly be distinguished.
Finally, we would like to stress that all the major indicators, especially those presented here, are based on the number of times a paper is cited. Hence they all measure an impact factor, which is worth whatever it's worth and must be used in view of the reservations presented in this report. In any case, it is a factor of impact and under no circumstances a factor of scientific excellence.
The most troubling aspect concerns the calculation of these indicators. We have seen that the databases used for these calculations have significant limitations both in terms of the corpus and the relevance of the results. This is particularly true for Google Scholar and any other software using automatic processing of documents retrieved on the Internet. The figures obtained are thus approximations, often very gross approximations, of the expected figures. The indicators are thus figures with questionable reliability.

6 Misuse and unintended effects

One of the unintended effects of indicators' importance is the temptation to manipulate them to increase their number, without improving research quality in any way. We will mention a few easy manipulations which have been observed, without attempting in any way to make an exhaustive list.

6.1 JIF manipulations

A standard practice in certain fields, and one which is beginning to spread to journals that publish articles by INRIA researchers, involves asking the authors of accepted articles to include references to articles in the same journal (which constitutes a form of self-citation), even though the relationship between the references and the article's subject is not really clear[24]. An exhaustive analysis of WoS references showed a rate of erroneous references (those having no link with the paper's subject) of around 7%. This rate was much higher in certain specific situations [11]. ISI is aware of this problem; it offers a self-citing rate for journals, which is the ratio between the citations published in the journal and its total citations. This creates a paradoxical situation whereby a journal should prohibit its authors from citing the articles it publishes...

[24] "Some authors need no arm-twisting, they volunteer to include reference to articles in the target journal even though they may not be relevant to the submitted papers". Péter's Digital Reference Shelf, August. Published by Thomson Gale.

Another classic tactic plays on the definition of JIF itself and involves publishing material that can be included in the numerator of JIF but not in the denominator[25], such as editorials[26].

6.2 Salami slicing

Faced with how standard indicators are defined and their growing importance, researchers might be tempted to divide their publications for a given project into several articles containing a particular element of their research, whose significance can only be appreciated by reading all the articles. The aim of course is to ensure a high number of citations[27].

6.3 Self-citations

Self-citation refers to the fact that an author may cite his/her own articles.
Of course this practice isn't reprehensible in and of itself, but it could influence the indicators were it to become malicious. Various studies have shown that self-citation doesn't change the value of indicators in a major way if the field considered is sufficiently vast and the time period sufficiently long, except in specific fields[28]. However, there may be reason to fear that the pressure of using indicators for management would push people to try to optimise them. For example, it would be easy to set up a citation strategy within a group that would allow significantly increasing the h-index of each group member. Such a ploy would be difficult to detect.

6.4 An obstacle to risk-taking

An exaggerated consideration of indicators to evaluate individual research may push young researchers into fields where results can be obtained quickly, to the detriment of more long-term research[29]. Likewise, indicators can act as an obstacle to innovation, given that major innovation often occurs within limited communities. Moreover, indicators are poorly suited to rapidly detecting innovative research[30] (e.g. a new journal won't have a JIF until three years after its creation at best).

7 Cost of indicators and implications

INRIA acquired access to WoS as part of a national ordering group including around 50 institutions (public science and technology institutions, universities, ANR, OST, etc.) for a

---

[25] "[T]ake for example the paper about the accessibility of information on the Web (labeled as editorial material by ISI (workgroup's note: thus not counted in the JIF denominator)) from *Nature*. It was already cited by 325 articles", Péter's Digital Reference Shelf, August. Published by Thomson Gale.

[26] Which can themselves refer to articles in the journal: (regarding an editorial in *Diagnostica*) "I must warn you that the text is about a half page and the rest is the listing of all articles published in *Diagnostica* in the previous two years and usually nothing else". Péter's Digital Reference Shelf, August. Published by Thomson Gale.

[27] "While the practice of 'salami slicing' is frowned upon, it is widely practised by those who think that the length of their curriculum vitae measures its strength, moreover it is inevitable in reporting major studies...". R. G. Newcombe and J. Stebbing, Postgraduate Medical Journal 2006;82:355-356.

[28] "Self-citations that appear in prestigious high-impact economics journals have a statistically positive, but numerically small, effect on a subsequent article's total citation count and on the quality of the citing journal". M. H. Medoff, Scientometrics, 2006, vol. 69, no1, pp. 69-84.

[29] *Resources are often accorded to majority groups, which favours a conservative approach.... It is difficult to give young talent a chance when the trend is more toward immersion in what already exists.* [Translated from the original French] M. Vert, Lettres des Sciences Chimiques, 64, January 1998.

[30] *A very regrettable consequence ... would be the temptation for scientists to neglect or abandon less popular research themes in favour of more profitable ones, in order to improve their status and/or funding. This would result in the impoverishment of scientific knowledge in many fields which are "virtually" interesting.* [Translated from the original French] G. Roelants, INFI Newsletter, December 1999.

period of 3 years. Initially the scope of our access was essentially limited to Science Citation Index Expanded, with our archive access only going back to 1991; in 2007 this scope was expanded to include the *Journal Citation Reports*.
Moreover, a second agreement was reached with the CNRS (ordering group operator) and INRIA whereby the CNRS agreed to cover half our bill, whereas the buying group allows us to obtain these resources at 30% of their institutional rate.
The dual effect of this savings and the participation of the CNRS brings INRIA's WoS bill to around 6000 euros per year (instead of 40,200 USD). Since 2007 an additional 1000 euros per year must be added for *JCR*, for an annual total of 7000 euros.
INRIA is not currently subscribed to Scopus and it is difficult to estimate the cost of such a subscription. However it is generally estimated that this cost is around 85% to 95% of that of WoS [2].
Access to fee-based sources is thus relatively expensive, whereas we have seen that the imperfect coverage rate of these tools should lead users to multiply their sources to establish the necessary indicators in a reasonable manner and/or to control for the methodological biases in the calculation of indicators. Furthermore, in the specific case of INRIA, the preliminary study presented in Section 9 shows that the results obtained with fee-based sources are much more disappointing than those obtained using free tools.
Moving into another area, the use of these indicators to guide subscription choices in a reference library would be extremely pernicious.

8 Standardising addresses

One way to make the indicators provided by ISI more accurate would be to inform the company of the Institute's specific nature and to standardise the addresses of INRIA researchers so that papers published by INRIA project-teams at LORIA or IRISA could be counted as INRIA publications. The CNRS proceeded in this manner by having the term "CNRS" included in the abbreviations selected by ISI. It would be a good idea to see if the same thing could be done for "INRIA".
Faced with the same problems, the CNRS defined for its own needs the concept of a *publication of the CNRS laboratories* as being a publication for which one of the authors indicates his/her address as being a unit supported or evaluated by the CNRS, whether the author is an employee of the CNRS or not[31]. However, this concept in no way resolves the problem of indicators provided by outside entities such as ISI.
In a related move, OST has launched the programme *Normaddresses*[32] whose goal is to provide ISI with a nomenclature for French laboratories. We quote from the site on this initiative [translated from the original French]:
The OST initiative is aimed at improving the quality of processing applied to French addresses in Web of Science by Thomson Scientific. This initiative must be accompanied by two types of action: upstream, efforts to standardise the addresses in manuscripts sent by researchers to journal publishers; downstream, efforts to familiarise those who use indicators with the various calculations used by the producers. First of all, researchers should be made aware of the importance of providing complete addresses in the manuscripts they submit to publishers. This step is primordial, because this is the information, necessary to proper identification of the article, its authors and their home institutions, which will be transcribed in the published article. However, researchers should be made aware that publishers often limit

[31] Les publications des laboratoires du CNRS et leur impact, Unité d'indicateurs de politique scientifique, March 1999.
[32] See www.obs-ost.fr/e-docs/00/00/00/A0/documentflactu.phtml

the number of characters allowed for address description. The next step involves creating the bibliographic record. When articles are put into the Web of Science database, as a general rule Thomson Scientific doesn't eliminate any institutions from the addresses, but it sometimes changes the wording for standardisation purposes. Similarly, publishers sometimes change the order of the data: for institutions, and for structures within the same institution (institution, department, laboratory, etc.). Processing "mixed" addresses (joint units) raises specific problems not always well resolved by the publisher. Downstream, rigorous methods are required to produce bibliometric indicators. For example, OST works line by line with each institution to identify its articles, accounting for the multi-affiliations of mixed units, which are differentiated from co-authors. However, numerous studies in the world are based on the choices made by automatic processing, which only identifies certain institutions or certain occurrences of names, or which limits their identification to the first institution mentioned in each address. Depending on the objective, these choices may or may not be suitable, but before using these indicators it is important to have a good grasp of the processing used to calculate them, to allow good interpretation.
Regarding the sources' complementarity, examinations of this type shouldn't be limited to WoS; the same procedure should be applied to Scopus and Google Scholar.

9 An example analysis for INRIA authors

9.1 Comparing the sources

We conducted a preliminary comparison of citation searches using WoS, Scopus and Google Scholar (GS) for four INRIA researchers (1 = computer vision, 2 = networks, 3 = financial mathematics and 4 = pure computer science). The raw data were corrected manually to eliminate duplicates and correct the titles (a given article often has different title spellings depending on the database).

The citation data were retrieved and saved to files, before undergoing preliminary manual processing (to ensure consistency of the article titles) so as to allow automated processing in a subsequent step. Such manual processing is time-consuming (around 1/2 day for 2 researchers) but necessary since raw data vary considerably from one database to another (not only for a given article but also for the results, as we will see below). "Total number of articles found" indicates the total number of different papers found in all three databases, "articles found in" indicates the number of articles found in one of the databases, and "number of citations" indicates the total number of citations listed in a given database. We obtained the results presented in Table 1[33].

Researcher	1	2	3	4
Total number of articles found	152	91	86	128
Total number of articles found (with at least one citation in GS)	137	69	75	114
Articles found in WoS	31	13	16	13
Articles found in Scopus	52	23	27	4
Articles found in Google	140	89	77	126
Articles ∈ WoS; ∉ Scopus	6	3	2	12
Articles ∉ WoS; ∈ Scopus	27	12	13	3
Articles ∈ WoS; ∩ Scopus	25	10	14	1
Articles ∈ WoS; U Scopus	58	25	29	16
Articles ∈ WoS; ∉ Google	2	0	3	2
Articles ∉ WoS; ∈ Google	111	76	64	115
Articles ∈ WoS; ∩ Google	29	13	13	11
Articles ∈ Scopus; ∉ Google	11	2	9	0
Articles ∉ Scopus; ∈ Google	99	69	59	122
Articles ∈ Scopus; ∩ Google	41	20	18	4
Number of WoS citations	237	163	104	78
Number of Scopus citations	652	334	122	52
Number of Google citations	1859	981	565	2324

Tab. 1 - Articles and citations found in WoS, Scopus and Google Scholar and in the various combinations of these databases, compared to the total number of articles found in the 3 bases together (1st and 2nd lines).

For informational purposes, we also consulted databases which are more specialised in computer science (free access): DBLP[34] (which doesn't allow citation searches) and CiteSeer.

Researcher	1	2	3	4
Articles found in DBLP	54	19	1	39
Articles found in CiteSeer	58	49	1	42
Number of citations found in CiteSeer	134	103	0	818

Tab. 2 - Articles and citations found in the databases DBLP and CiteSeer for the 4 researchers.

Obviously, the researcher in computer science has the best article numbers for both DBLP and CiteSeer (which are also far higher than the WoS or Scopus results). However, these numbers are far below those for GS. In contrast, for the other researchers, the results range from average (Researchers 1 and 2) to very bad (Researcher 3).

The percentage of articles indexed by the various search engines relative to all indexable papers is given in Table 3.

For these four examples, it would appear that the combined use of WoS and Scopus only locates around 1/3 of the potentially citable articles, with the lowest number found for pure computer science (12.5%). We also observed that for papers found through combined use of WoS and Scopus, there is relatively little overlap

[33] In the tables, the symbol U = union, ∩ = intersection, ∈ = in set and ∉ = not in set.

[34] http://www.informatik.uni-trier.de/»ley/db/

	WoS	Scopus	Google	WoS U Scopus	WoS ∩ Scopus
Researcher 1	20.4%	34.2%	92.1%	38.16%	16.45%
Researcher 2	14.3%	25.3%	98.9%	27.47%	11%
Researcher 3	18.6%	31.4%	89.5%	33.72%	16.28%
Researcher 4	10.15%	3.12%	98.44%	12.5%	0.78%

Tab. 3 - Number of papers found by each of the search engines relative to the total number found by combining the results of WoS, Scopus and GS

between the two search engines (typically 50% or less). Finally, papers which are extensively cited may not be included in a given database.

It can also be noted that the number of citations found varies considerably depending on the search engine. Clearly, Google Scholar's systematic scanning of the Web makes it possible to find a much broader set of relevant citations than can be found by the other engines. However, this kind of search has many flaws:

- The problem of self-citations is not addressed.
- All documents have the same value: a citation in a Master's research report is evaluated at the same level as a reviewed article.
- The documents counted as citations vary greatly in nature: in a random examination of the citations provided by GS, we found unpublished articles, industrial contract reports and even CVs.

Even if we exclude Google Scholar, there are significant differences between WoS and Scopus. To illustrate this, we took the 10 most cited papers for Researcher 4 using a given database (the top 4 in Scopus) and we examined the number of citations found in the other databases for these papers, obtaining the following tables. The results are clearly inconsistent, whether it be for the papers indexed (e.g. a paper often cited in one database but absent in another), or in the ranking by number of citations. For example, while Scopus and GS are consistent and have the same top-ranking paper (with respectively 44 and 208 citations, Tables 5 and 6), this paper is not included in WoS and only has 2 citations in CiteSeer (placing it far below 10th place for this database).

Inversely, the best paper according to WoS (Table 4) doesn't even figure in the GS top 10 and is completely absent from Scopus.

Paper ranking according to citations in WoS	1	2	3	4	5	6	7	8	9	10
Number of WoS citations	22	13	12	11	7	5	4	2	1	1
Number of Scopus citations	-	-	2	-	-	-	-	-	-	-
Number of GS citations	67	62	67	67	85	38	47	8	14	47
Number of CiteSeer citations	22	10	23	6	26	31	-	0	2	-

Tab. 4 - Number of citations for the 10 most cited papers in WoS for Researcher 4, according to WoS (2nd line) and according to the other databases (following lines). "-" indicates the paper doesn't figure in a particular database.

Paper ranking according to citations in GS	1	2	3	4	5	6	7	8	9	10
Number of GS citations	208	160	111	101	95	85	83	74	73	70
Number of WoS citations	-	-	-	-	-	7	-	-	-	22
Number of Scopus citations	44	-	-	-	-	-	-	-	-	-
Number of CiteSeer citations	2	152	68	74	32	26	60	-	43	22

Tab. 5 - Number of citations for the 10 most cited papers in GS for Researcher 4, according to GS (2nd line). "-" indicates the paper doesn't figure in a particular database.

Paper ranking according to citations in Scopus	1	2	3	4
Number of Scopus citations	44	5	2	1
Number of WoS citations	-	-	12	-
Number of GS citations	208	37	67	6
Number of CiteSeer citations	2	-	23	-

Tab. 6 - Number of citations for the 4 most cited papers in Scopus for Researcher 4, according to Scopus (2nd line). "-" indicates the paper doesn't figure in a particular database.

For this researcher's four most cited papers in the four databases respectively, we obtained the number of citations indicated in Table 7.

Paper ranking	1	2	3	4
Number of WoS citations	2	13	12	11
Number of Scopus citations	44	5	2	1
Number of GS citations	208	160	111	101
Number of CiteSeer citations	152	74	68	60

Tab. 7 - Number of citations for the 4 most cited papers of Researcher 4 according to each of the 4 databases. In any given column, the paper considered is not necessarily the same, depending on the database.

9.2 Institutional searches

Institutional searches shouldn't give good results either; for the 4 researchers selected, we found the following 9 affiliations in WoS:

INFIA ROCQUENCOURT (sic!)
INRIA
INRIA RENNES
INRIA RHONE ALPES
INRIA ROCQUENCOURT
INRIA SOPHIA ANTIPOLIS
INST NATL RECH INFORMAT & AUTOMAT
INST NATL RECH INFORMAT & AUTOMAT ROCQUEN COURT (sic!)
NAT RES INST COMP SCI & CONTROL

Scopus does slightly better with "only" 6 affiliations:

INRIA
Inst Natl de Recherche en (sic!)
I.N.R.I.A.
Inst. Natl. Rech. Info. et Automat.
LORIA
LIFIA-IMAG and Iuria Rhone-Alpes (sic!)

It is impossible to search by affiliation with GS for these authors.

9.3 The h-index

The h-index results for these 4 researchers are also of interest and are presented below. The figures were calculated (after manual sorting) from each of the three search engines (Table 8). The discrepancy in the values of this indicator obviously reflect the pronounced flaws of citation searching, and any indicator based solely on citations will be subject to the same type of uncertainties on a large scale.

Researcher	1	2	3	4
WoS h-index	8	5	6	5
Scopus h-index	14	6	6	2
GS h-index	21	12	12	25

Tab. 8 - The h-index for the 4 INRIA researchers according to the various databases

10 Conclusion

Indicators are important metrics, but they must be used in an enlightened manner and according to some basic rules:

- Current indicators are essentially factors of impact, which is not synonymous with quality.
- One must simply look at orders of magnitude for indicators because even using various sources doesn't allow a high degree of accuracy.
- Expert opinion must be used to correct indicators.
- Several indicators must be used[35]. For example, JIF cannot be dissociated from cited half-life for a given field.
- Crosschecking various sources is necessary to obtain relevant information.
- Rather than measuring the quality of journals with quantitative indicators, it may be better to categorise them into groups based on the qualitative criteria suggested by researchers themselves. A project with this focus is currently underway: INRIA's Evaluation Committee is looking at developing a guide to leading publications, drawing on the expertise of the Institute's project-teams.
- Comparisons between fields should never be made.

These recommendations dovetail to some extent with those proposed by M. Zitt and G. Filliatreau (director of OST) regarding the biases and limitations of indicators, which [19]:

- Only account for part of the spectrum of activity.
- Must be calibrated to "compare like with like".
- Are poorly suited to observation of emerging trends.
- Require a broad range of approaches and levels of observation.
- For evaluation purposes, must be completed by other elements, particularly peer-assessment.

[35] According to Meho, "relying exclusively on Web of Science and a single citation measure will, in many cases, no longer be an option for making accurate impact assessments".

Concerning INRIA, we strongly advise better use of references in the activity report. The Institute is fortunate to have nearly all its publications brought together in one relatively well-structured document each year. P. Robert has shown that this could easily be used as a base to establish accurate statistics on INRIA publications. By extending this effort, we could also correct certain indicators.

In conclusion, we cite the healthy stance adopted by Vinay and Baverel [17] [translated from the original French]: *We must critically assess [indicators] as we use them; above all, we must reject the temptation they propagate to automate evaluation...*

11 Appendix 1: LOLF indicators

See www.obs-ost.fr/projetflu.phtml for a description and counting methods.
There are very few bibliometric indicators. Those which exist were developed by MIRES, the French interministerial mission on research and higher education.
Apparently the French research and innovation directorate (DGRI) hoped for more indicators[36], but the fear of being unable to provide the necessary data has impeded this.

Publications

- Percentage of internationally cited publications, relative to scientific output in France, in EU25 or in the world (overall and by discipline)
- Two-year citation index (relative impact index), overall and by discipline

Patents

- Number of patents filed respectively with the INPI [French national institute of industrial property] and the EPO [European Patent Office].

Framework Programmes for Research and Technological Development

- Rate of participation in projects financed by EU framework programmes (overall and by objective).
- Rate of coordination of projects financed by EU framework programmes (overall and by objective).

[36] A report by IGAENR notes the following: *Observation by the DGRI: An operator's situation and its efficiency are assessed by a set of indicators which, taken together, enable a qualitative evaluation of the institution. For example, no one indicator by itself enables "measurement" of an operator's level of excellence in fundamental research. A series of indicators is necessary: publications, citations, external and internal evaluations, affiliated researchers (including international researchers), number of researchers accredited as research supervisors, number of invitations to international conferences, awards, etc. Moreover, these indicators should be measured quantitatively, but also qualitatively.*

12 Appendix 2: Observatoire des Sciences et Techniques (OST)

OST is a public interest group created in 1990 and renewed in 2002 whose mission is to design and produce indicators for scientific, technological and innovation activities and to facilitate their interpretation in terms of France's position in Europe and in the world.
According to the OST website, its initiatives are organised around three main themes:

- Producing standard indicators, trend charts and a biennial report on science and technology indicators
- Performing studies on research fields, technological sectors, R&D activities of institutions, countries, regions, etc.
- Research and development on indicators, strategic evaluation methodologies, and management tools.

OST was specifically tasked by the French government to produce indicators for the scientific and technological output of the operators of Programme 150, on higher education and university research[37]. The bibliometric indicators OST selected for this programme and the methodology it used are interesting to examine. First of all, a single database (ISI) was used, and the first phase of the programme involved identifying the operators in the institutional lists provided by ISI, apparently by hand. Two types of attribution were used:

- Counting based on presence: One attribution for each operator mentioned in an article.
- Proportional counting: Operator contributions credited on a pro rata basis, depending on the number of author addresses and the number of discipline categories the article was assigned.

The indicators used are as follows:

- Two-year immediacy index, but only after the articles have been categorised into 8 fields (mathematics, physics, medical, engineering sciences, chemistry, fundamental biology, sciences of the Universe, applied biology). The index is then established as the ratio between the operator's percentage of citations and its percentage of publications.
- Degree of visibility: Once the articles are standardised by discipline, specialisation and journal, visibility categories are defined (first 5% of most cited articles, following 5%, following 10%, etc.). The activity index per visibility category is then defined as the ratio between the operator's percentage of articles in a given category and the percentage defining this category (for example, if the operator has 7.5% of its articles in the 5% most cited category, its activity index for this category would be 7.5/5 = 1.5).

OST has expressed reservations concerning the use of these indicators for inter-institution comparisons, however: data stabilisation, analysis of publication and citation "usage" at the specialisation level, data analysis within the context of the institution.

[37] www.obs-ost.fr/projetflu.phtml

13 Appendix 3: Glossary

- Article: Publication of a scientific text in a journal or in conference proceedings. The concept of an article used for citation searching may vary according to the citation source. For example, certain journal components such as editorials are integrated into the citation count but not into the number of articles published by the journal (included in the JIF denominator).
- Self-citation: For journals, this refers to the citation of papers published in the same journal. For researchers, this refers to an author's citation of his/her own work.
- Citation: Reference in an article to work conducted on a particular subject by the scientific community. Traditionally, citations are listed at the end of an article with references inserted in the body of the text. A citation can be positive (the authors note the quality of the research) or negative (the authors mention an article to draw attention to its poor quality, or worse).
- Indicator: Element (usually quantitative) intended to provide information for evaluating a scientific activity.

References

[1] Burnham J.F. Scopus database: a review, March 2006. www.bio-diglib.com/content/3/1/1.
[2] Dess H.M. Database reviews and reports: Scopus. Issues in Science and Technology Librarianship, Winter 2006. www.istl.org/06-winter/databases4.html.
[3] Egghe L. Theory and practice of the g-index. Scientometrics, 69(1):131, 152, 2006.
[4] Ergma E. et al. Science, morality and quality, July 2002. ALLEA Biennal Yearbook.
[5] Foster W.R. Impact factor as the best operational measure of medical journals. Lancet, 346(8985), 11 November 1995.
[6] Hirsch J.E. An index to quantify an individual's scientific research output. 29 September 2005. arXiv: Physics.
[7] Leydesdorff L. Caveats for the use of citation indicators in research and journal evaluations. April 2007. Preprint, SIGMETRICS.
[8] Magri M-H. and Solari A. SCI JCR: a tool for studying journals. Scientometrics, 35: 93,117, 1996.
[9] Meho L.I. The rise and rise of citation analysis. Physics World, January 2007.
[10] Meho L.I. and Yang K. A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. Journal of the American Society for Information Science and Technology, 2006. To be published. http://arxiv.org/abs/cs/0612132v1.
[11] Moed H.F. The impact-factor debate: The ISI's uses and limits. Nature, 415(6873): 731¸ 732, 2006.
[12] Moed H.F., Visser M.S. Developing bibliometric indicators of research performance in computer science: an exploratory study. CWTS report 2007-1, Leiden University, February 2007.
[13] Rehn C. et al. Bibliometric handbook for Karolinska Institutet. 27 November 2006.
[14] Roediger H.L. The h index in science: a new measure of scholarly contribution. The Academic Observer, 19(4), April 2006.
[15] Seglen P.O. Why the impact factor of journals should not be used for evaluating research. British Medical Journal, 314(7079):458¸502, 1997.

[16] Sutherland W.J. What do impact factors tell us. TREE, 14(10):382¸384, 1999.
[17] Vinay P. and Baverel G. Au fur et à mesure: à propos de l'évaluation numérisée de la production scientifique. Médecine/Science, 10:701¸703, 1994.
[18] Zitt M. Facing diversity of science: A challenge for bibliometric indicators. Measurement, 3(1):38¸49, 2005.
[19] Zitt M. and Filliatreau G. Bibliométrie et indicateurs : rôle de l'OST. In Rencontres 2005 des professionnels de l'IST, Nancy, 20-22/06/2005.