Growth of Science and Database Coverage

Carolin Michels[1] and Ulrich Schmoch[2]

[1] *Carolin.Michels@isi.fraunhofer.de*

[2] *Ulrich.Schmoch@isi.fraunhofer.de*

Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Straße 48, 76139 Karlsruhe (Germany)

Abstract

In recent years the Web of Science (WoS) displays a tremendous growth of articles, but is unclear whether this observation is linked to a growth of science or simply an additional coverage of already existing journals by the database producer. An analysis of the category of journals in the period of 2000 to 2008 shows that the number of the basic journals from 2000 covered by WoS steadily decreases, whereas the number of new, recently established journals increases. In addition a rising number of older journals is covered in addition. These developments imply a crescive number of articles, but a relevant effect is the enlargement of the traditional, basic journals in terms of annual articles. All in all it becomes obvious that the data set is rather instable due to a high fluctuation caused by the annual selection criteria, the impact factor. In any case, it is possible to distinguish growth effects induced by the growth of science and those referring to an enlarged coverage linked to a lowering of the entry thresholds. It is important to look at the specific growth structures of a field for differentiating between "real" and "artificial" growth.

Introduction

The process of globalisation goes along with a strong trend towards a knowledge society (Stehr 1994) a major aspect whereof is an increased scientification. This latter development has various aspects such as the growing relevance of scientific expertise in political decisions or the transformation of formerly experience-based professions into science-based ones. A further key element of scientification is the accretive volume of scientific research, primarily research in public research institutes and higher education institutions (HEIs). However it proves to be difficult to measure the growth of scientific research and to provide an objective basis for this felt change. The most useful indicator would be the number of research staff in public research institutes and HEIs, but the available statistics at an international level are rather fragmentary at least at an international level. For various countries such data are available at an aggregate level, but the differentiation by scientific fields is too crude.

Against this background, it is manifest to use broad publication databases such as the Web of Science (WoS) for the analysis of the growth of science. Already de Solla Price referred to the number of journals to provide evidence for the fast growth of science since the beginning of the 17th century (Price 1971). However, the number of journals covered by the Web of Science steadily enlarges, and it is unclear whether this observation is linked to a growth of science or simply an additional coverage of already existing journals by the database producer. In particular in recent years, the number of publications registered in Web of Science increases at high speed as consequence of the competition with the new bibliometric database Scopus. As reaction on Scopus, the producer of Web of Science, formerly the Institute of Scientific Information (ISI), now Thomson Reuters, improved various features, e.g. in a large part of the records the full first name of the authors are indicated instead of the initials, and in particular the number of journals covered is augmented. Therefore it is probable that the increase of publications in the Web of Science primarily reflects the amended coverage of this database and not the growth of science. Nevertheless various authors use a growing number of publications as indication for growing research activities in specific fields (see Gupta and Dhawan 1998, as one of many other examples). To cope with

this problem, Leydesdorff et al. (1994) suggested to use constant journal sets, but to include new journals in addition to capture the dynamics of new scientific fields. The aim of this contribution is to analyse whether this approach conceived at the beginning of the 1990s in a situation of a relatively slow growth of the Web of Science is still appropriate in the present situation of a turbulent change of the database coverage.

Approach

For analysing scientific growth within a database with increasing journal coverage, we decided to analyse the change in Web of Science in the recent period of 2000 to 2008 and to distinguish different types of journals:

(1) Journals covered by WoS in the start year 2000, called basic set
(2) New, recently established journals introduced in WoS in the period 2001 to 2008. This type was operationalised by a low volume number below or equal to 10, called new journals.
(3) Additional old journals introduced in WoS in the period 2001 to 2008. This type was operationalised by a higher volume number of 11 or more, called additional journals.
(4) Journals contained in the database before the year 2000 for one or several years, which were not listed for several years and then were included in the database again after 2000, called recurring journals.

The underlying assumption is that the database covers a set of basic journals representing a stem set which was steadily enlarged in the period since 2000. On the one hand the added journals are new journals that reflect the necessity of new publication media beyond the existing ones, for instance new specific journals on nanotechnology for complementing the existing journals in physics and chemistry. Thus these new journals reflect the dynamics of science in emerging fields, as already suggested by Leydesdorff et al. (1994). The appearance of a new journal can also reflect the case where a traditional one is terminated and continued with a new name or split into two new journals. Then the discontinuation of the old journal is compensated by the introduction of the new one.

The third category of additional journals stands for the policy of the database producer to broaden the coverage.

Thus the basic assumption is that the number of the journals in the basic set is relatively stable and that the growth of publications is primarily due to the inclusion of new and additional journals where in the recent period the share of additional journals is distinctly larger than that of the new ones. If this assumption held, it would be possible to calculate a deflator for the number of the total publications in order to determine the real growth of scientific activities.

Furthermore we guessed that the development differs by field of science and we conceived six broad areas of science defined by category codes of the Web of Science, thus fields defined by specific journal sets. These areas are:

(1) Medicine
(2) Life sciences
(3) Chemistry
(4) Physics
(5) Engineering
(6) Social sciences and humanities
(7) Other fields

Our assumption was that relatively new areas such us life sciences are more dynamic than traditional ones such as physics or chemistry. In addition we supposed that the areas of engineering and social sciences grew faster than other ones linked to the endeavour of the database producer to enlarge the coverage in insufficiently represented fields.

We used the full data set of the Web of Science, but focussed on articles and in particular did not consider proceedings.

Findings

First of all, a growing number of journals between 2000 and 2008 can be observed with an increase of about 27 percent, equivalent to an average annual growth of about 3 percent. In particular a stronger increase is visible especially in 2005 (3.5 percent) as well as in 2007 and 2008 (4.7 and 8.8 percent respectively), thus obviously confirming the thesis of the attempt of the database producer to achieve a broader coverage (Figure 1).

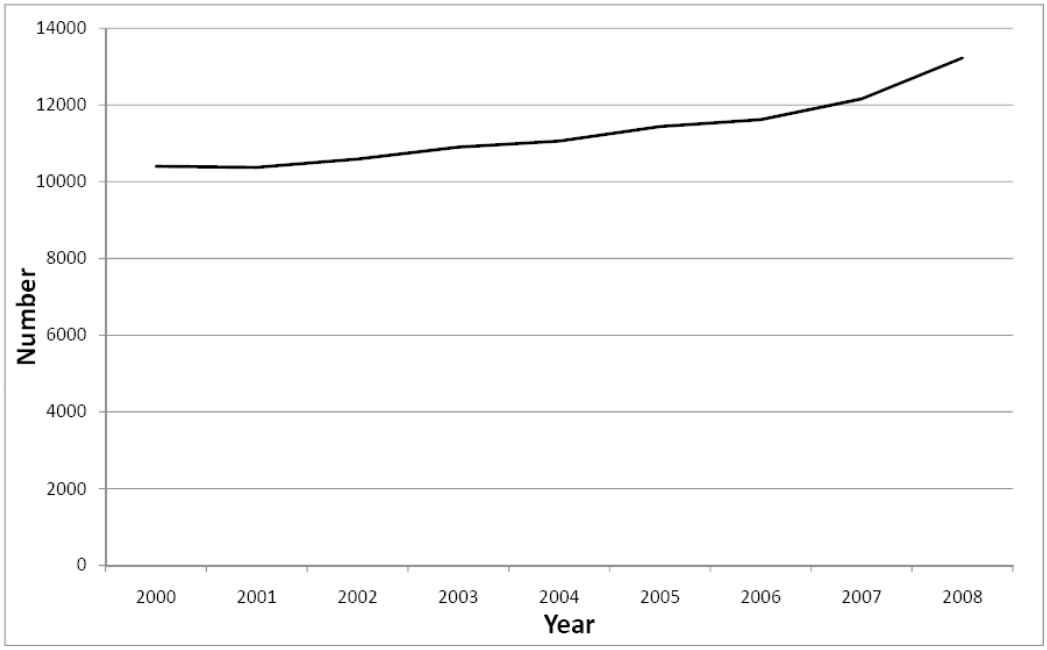

Figure 1. Number of journals in the database Web of Science

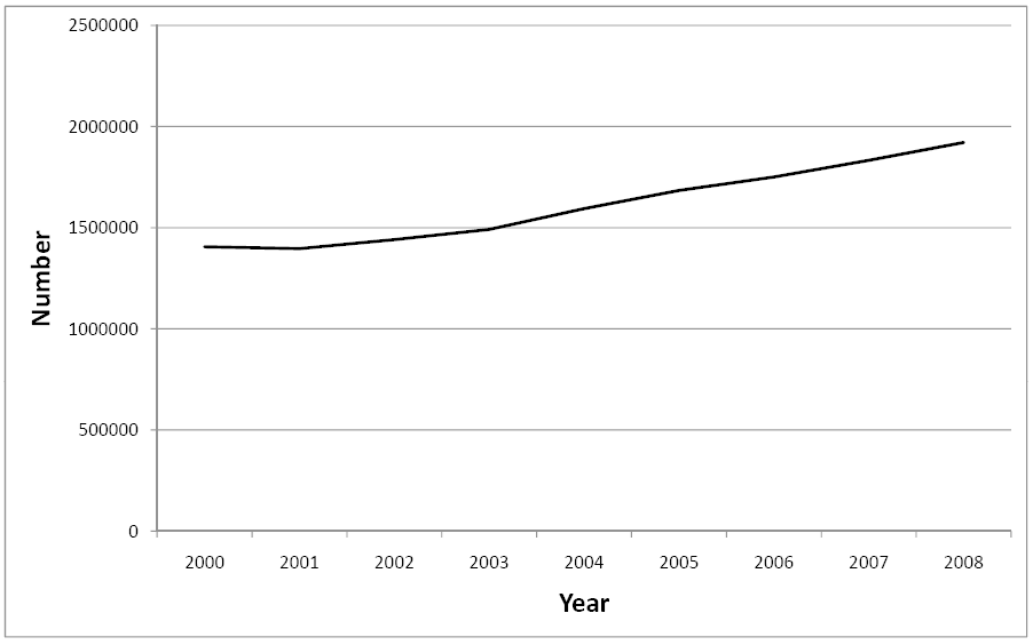

Figure 2. Number of journal articles in the database Web of Science

The growing number of journals covered implies a growing number of articles growing even stronger than the journal number with 37 percent in the period 2000 to 2008 (Figure 2). Here again, the accelerated increase is conspicuous, but primarily in 2004 and 2005 (7.0 and 5.6 percent respectively).
Looking in more detail for the underlying structure of journals, the development in the area of medicine is shown in Figure 3. A first striking result is that the number of basic journals from the year 2000 steadily decreases by about 16.3 percent between 2000 and 2008. In particular in the first year between 2000 and 2001, the drop is substantial with about 7 percent. The share of the recurring journals is low and the number of new journals appears to be moderate, whereas that of additional ones is substantial. Thus the total growth seems to be primarily due to additional journals.
These structures of the development of journal coverage for medicine are similar in the life sciences, chemistry, and the social sciences.
The development in engineering is quite different, as documented in Figure 4. In this area, the number of journals increases until 2005 and then decreases primarily due to a discontinuation of many "additional" journals. In the last years a moderate re-increase can be stated. The drop of the "basic" journals proves to be more substantial than in medicine. But all in all, the number of journals in 2008 is still higher than in 2000.
A similar development, but less pronounced, can be found for physics.

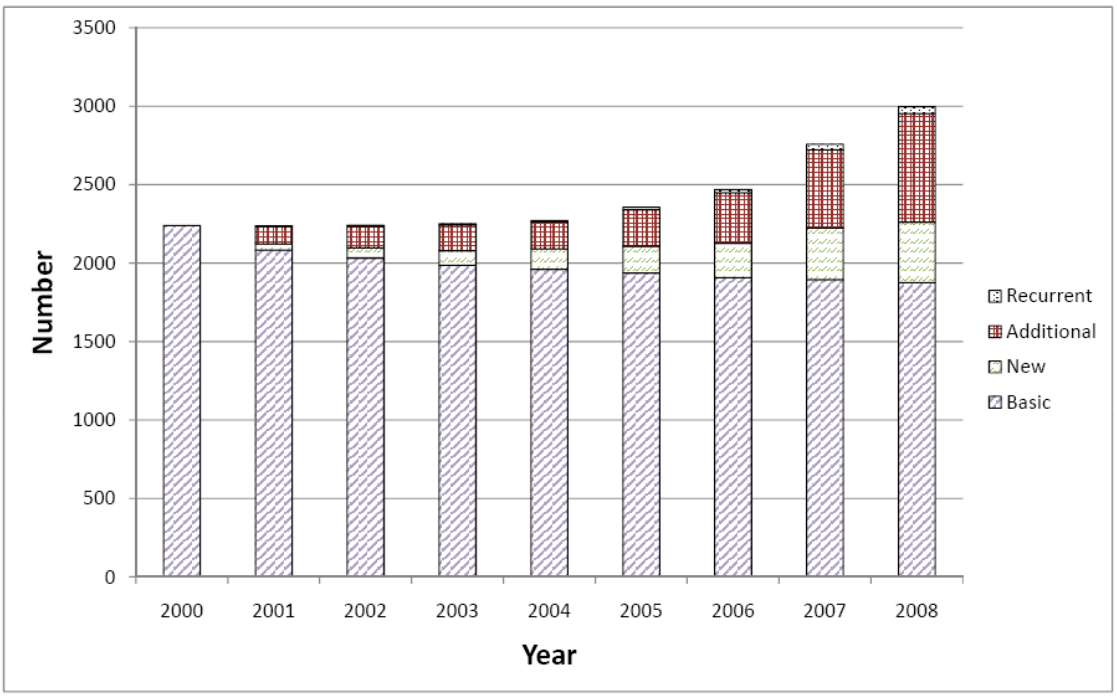

Figure 3. Number of journals in the category Medicine

In total, the number of journals is increasing in all areas between 2000 and 2008. The strongest relative growth can be observed in medicine and social sciences, thus our hypothesis about the higher dynamics of new scientific areas such as life sciences cannot be confirmed by these data (Figure 5). The hypothesis of strong growth in engineering and social sciences linked to the endeavour of the database producer to achieve a broader coverage is only confirmed for the social sciences, whereas for engineering the increase is moderate. Instead of that, the surge of journals in medicine appears to be relevant.

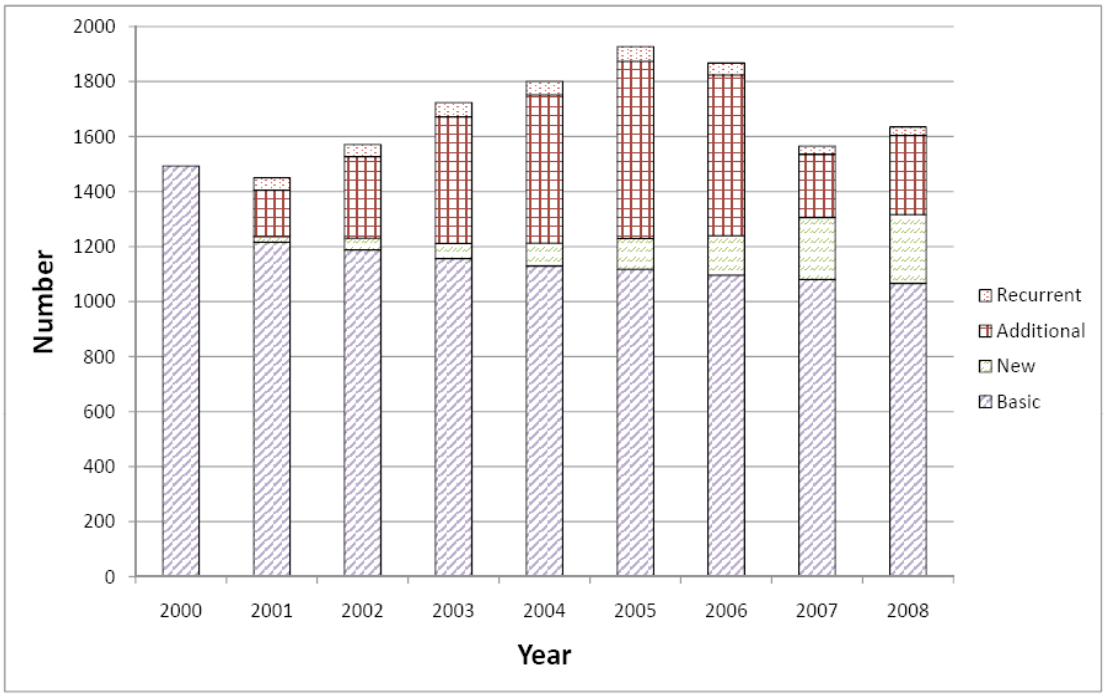

Figure 4. Number of journals in the category Engineering

Looking at the change of structure between 2000 and 2008, the number of journals from the basic set in 2000 is considerably reduced in 2008, but with the inclusion of the other journal categories, in particular the additional journals, the total number is higher than in the basic set in all areas, first of all in the life sciences and the social sciences (Figure 6).

Regarding the number of journal articles, there are again different phenomena to discover in the different areas; in medicine, the number of journal articles in total increases continuously after a short drop in 2001 (see Figure 7). Similar developments can be found for life sciences, chemistry and physics.

Figure 8 shows the number of journal articles for engineering. Here, the instable data set of journals (Figure 4) is reflected in the journal articles. The number of articles deriving from additional journals increases and develops in a rather volatile way, but less pronounced than on the journal level.

While the number of journals in social sciences shows a tendency to increase similar to medicine (see Figure 5), the number of journal articles in this area first drops due to the decreasing number of basic journals that is not compensated with a corresponding number of new, additional or recurrent journals (Figure 9). After 2003 the number of journal articles increases because of a growing number of additional journals. The decrease period in the social sciences is described in Larsen and von Ins (2010), but proves to be intermediate based on more topical data.

The development total number of journal articles by area between 2000 and 2008 is shown in Figure 10. In comparison with the development of journals numbers (Figure 5), medicine and life sciences are the only areas, where a high journal growth rate is reflected in a high article growth rate. Nevertheless, chemistry, physics and engineering even have a relatively high growth rate in terms of articles despite the relatively small increase in journals.

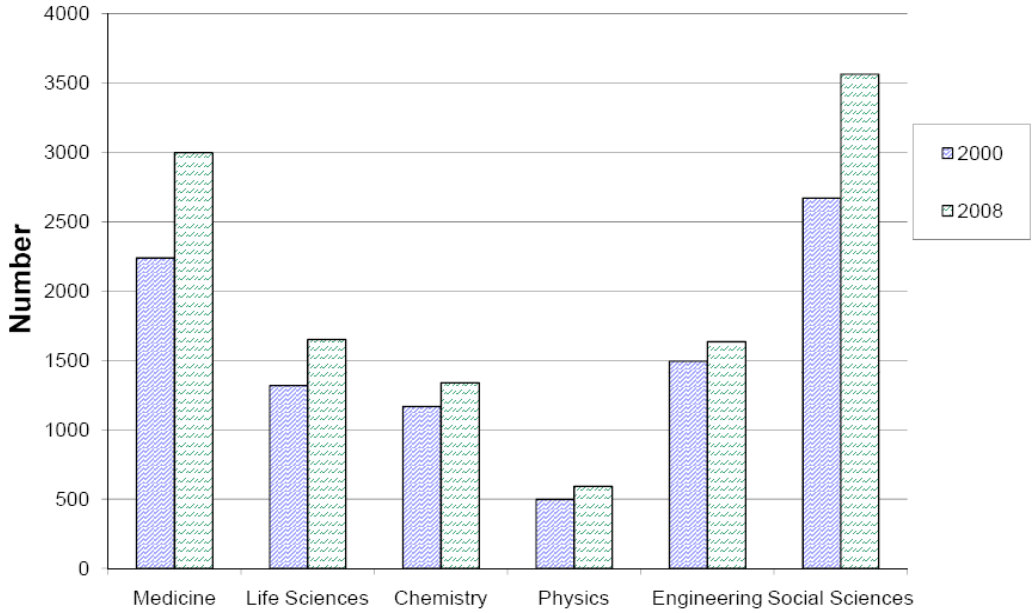

Figure 5. Number of journals in the years 2000 and 2008 in comparison

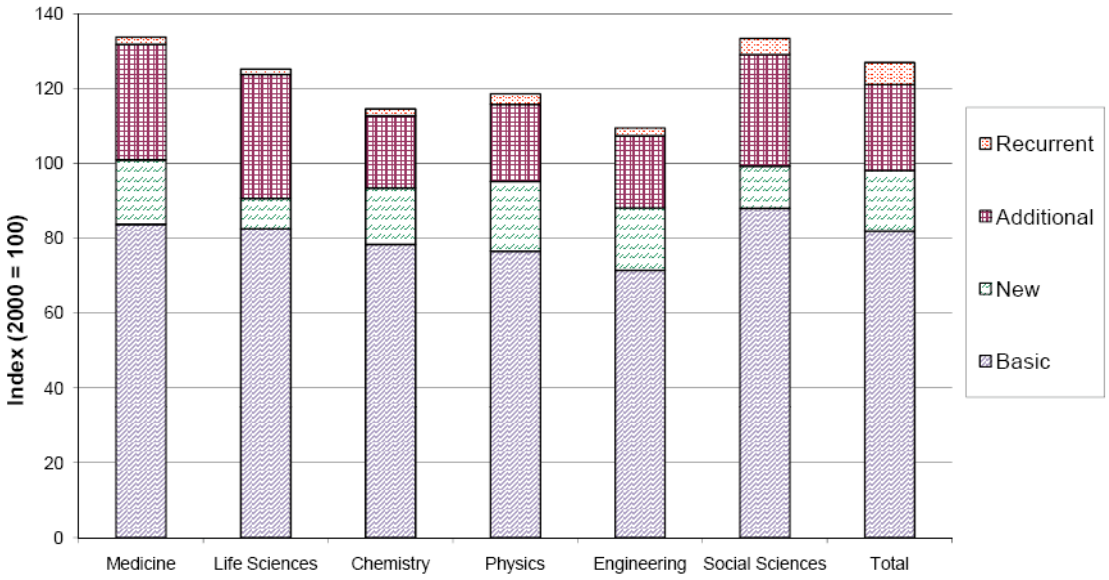

Figure 6. Number of journals by categories in 2008 with reference to 2000 (Index 100 = total number of journals in 2000)

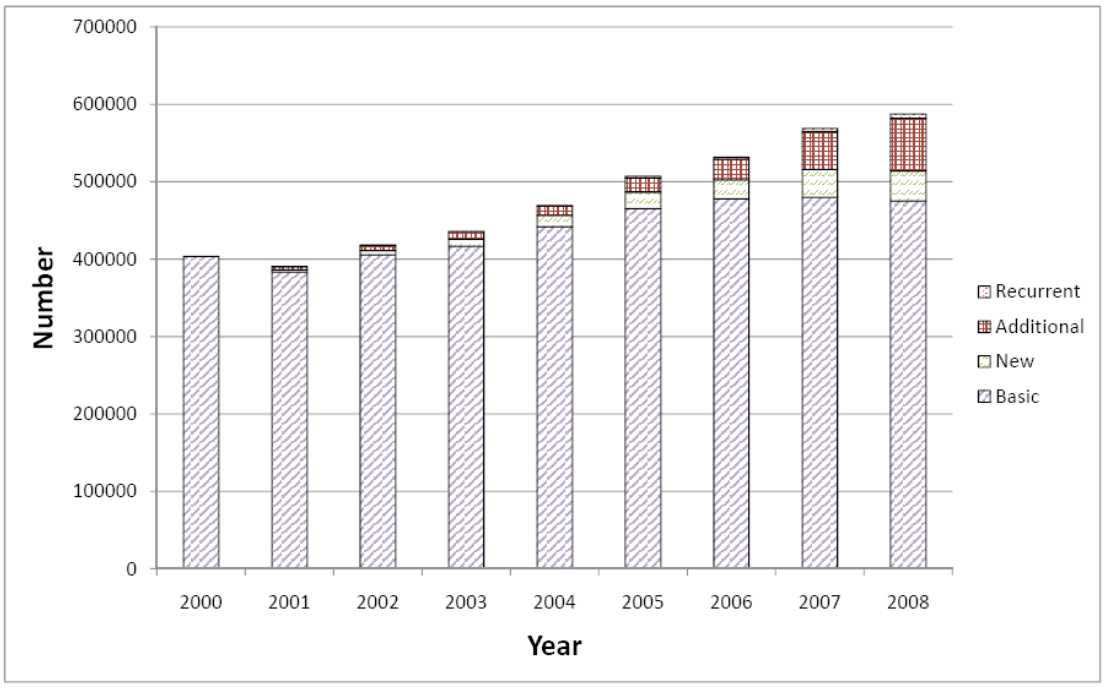

Figure 7. Number of articles in the category Medicine

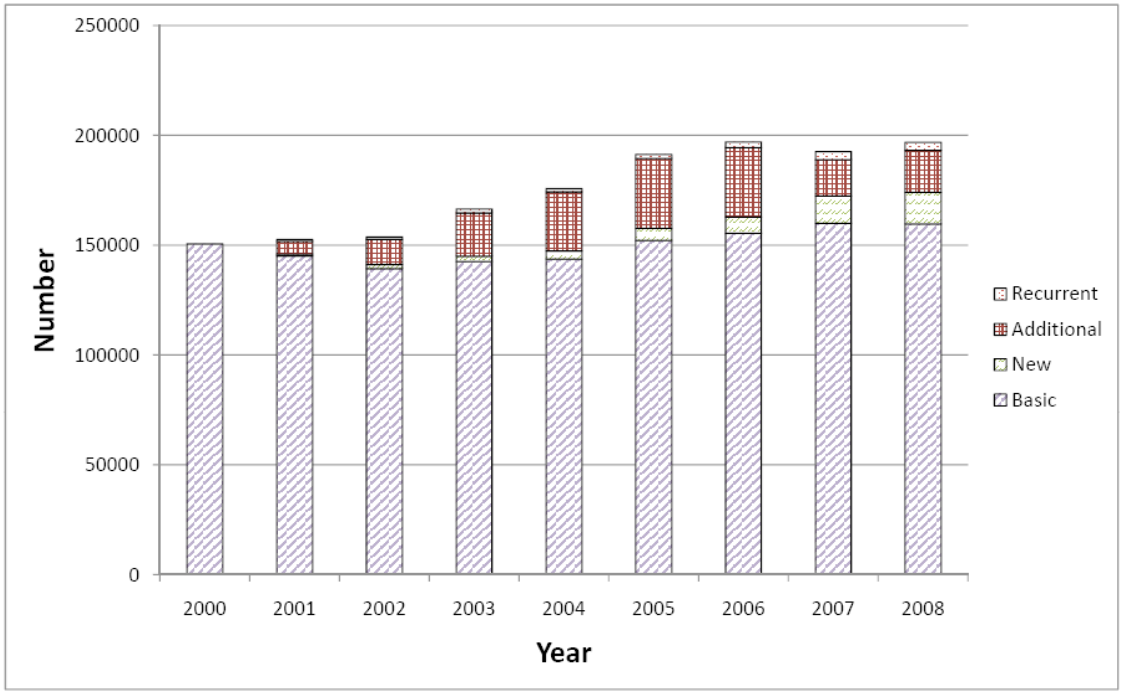

Figure 8. Number of articles in the category Engineering

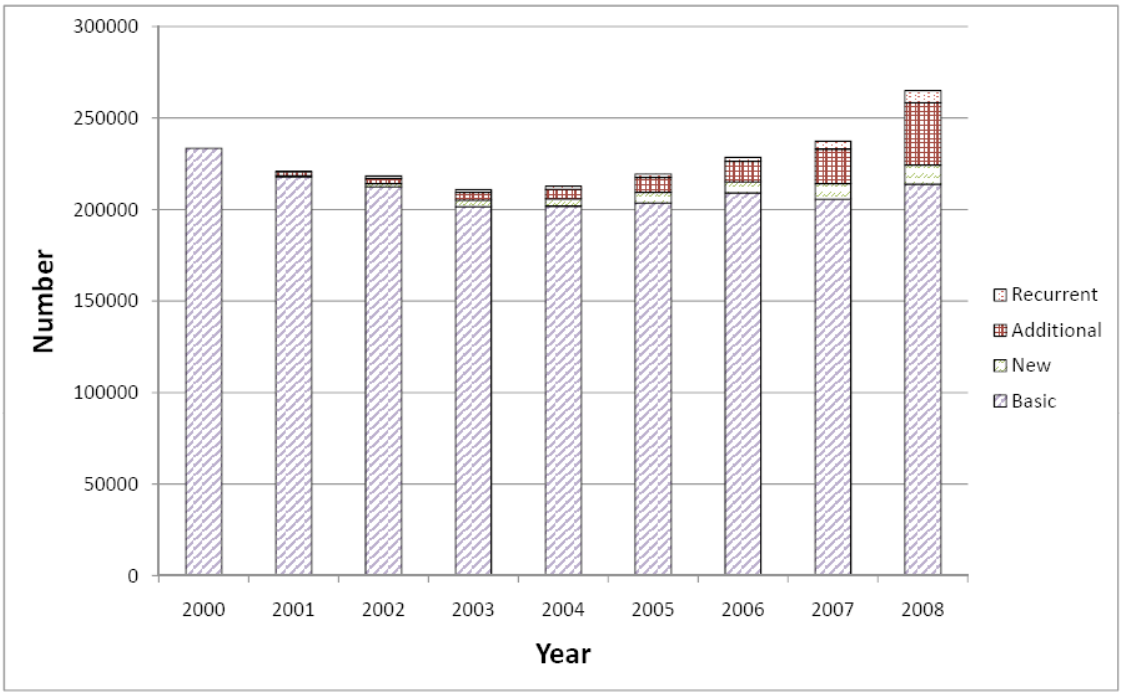

Figure 9. Number of articles in the category Social Sciences

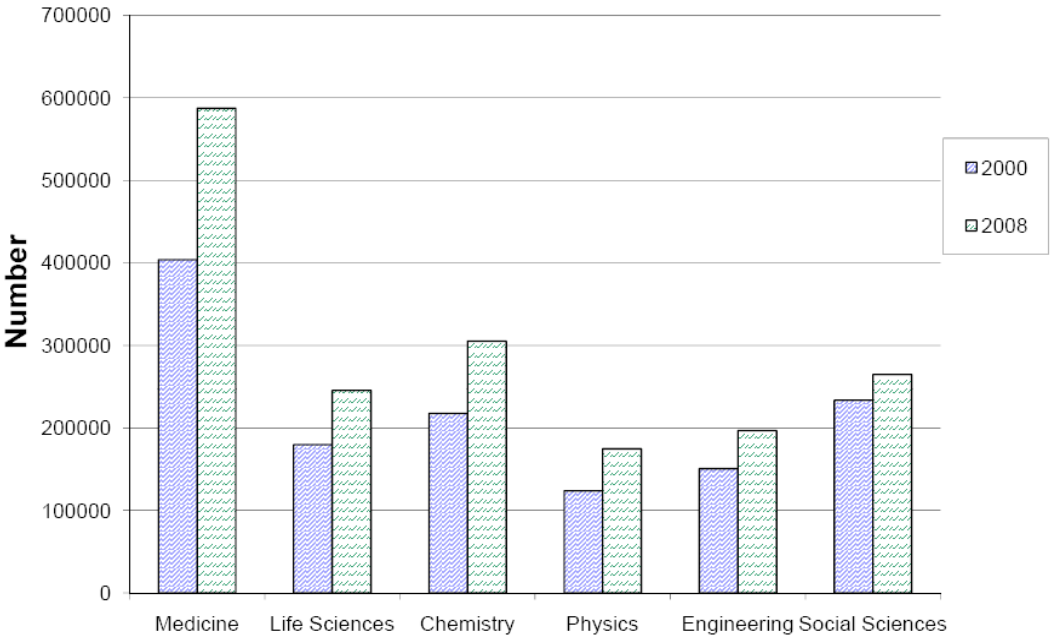

Figure 10. All articles in the years 2000 and 2008 in comparison

The obvious differences between the pattern in terms of journals and in terms of articles have several reasons. One factor is the different size of journals in the different categories, as illustrated by the life sciences in Figure 11. There the size of the basic journals appears to be much larger than for the categories new, additional and recurrent. Therefore, the relation between articles and journals is largely dependent on the relevance of basic journals. Comparing the number of journals and articles by area and by journal categories in 2008 (Figures 6 and 12), the higher weight of journal numbers, in particular in the category additional, compared to article numbers is obvious.

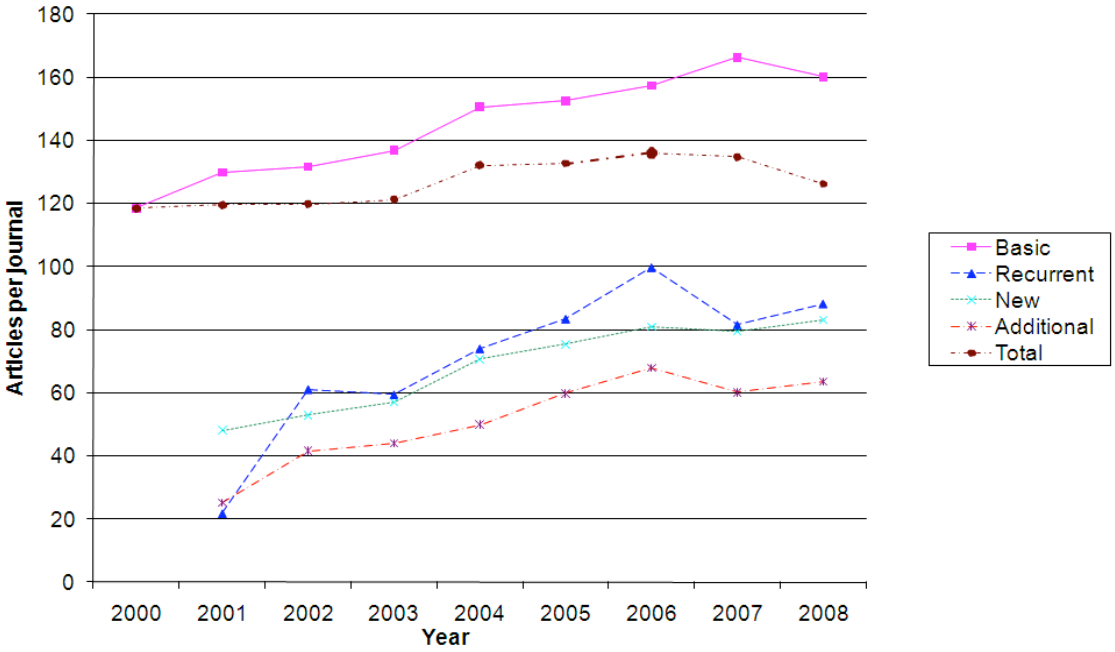

Figure 11. Annual number of articles per journal in the life sciences by journal categories

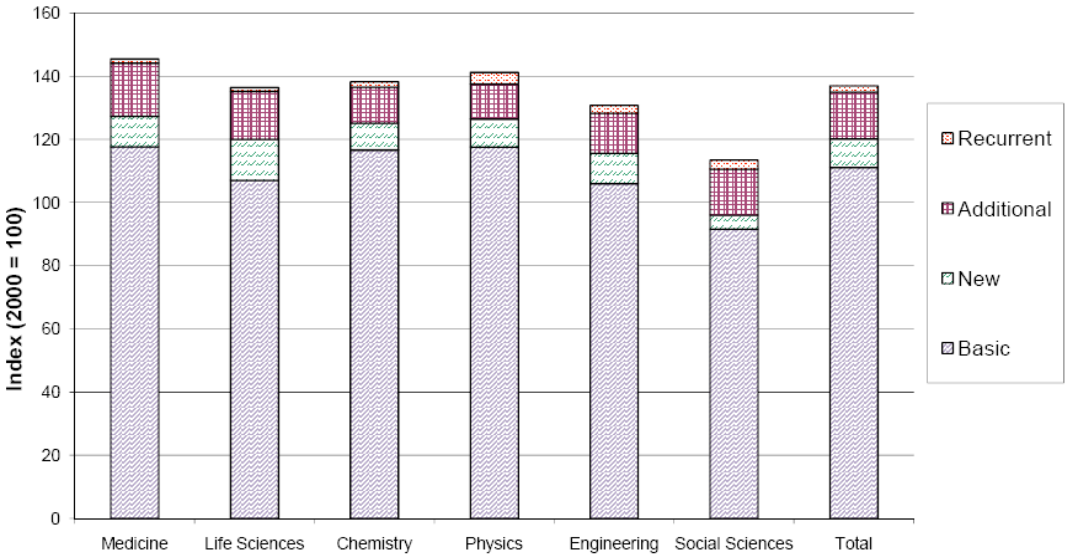

Figure 12. Number of articles by journal categories in 2008 with reference to 2000 (Index 100 = total number of articles in 2000)

A second factor is the increase of the average number of articles per journal. In the relatively brief observation period from 2000 to 2008, this relation augmented by 7.4 percent for the all categories combined (Table 1). Thus there is an obvious pressure on the journals to enlarge their coverage to cope with the growing submission of papers.

A further element that has to be taken into account is the slowly rising number of pages per article (Table 2), for all articles in WoS by about 5.7 percent between 2000 and 2008. So the often deplored slicing of research results which should imply a shortening of the papers cannot be supported by our data.

Table 1. Annual number of articles per journal in the period 2000 to 2008 by area

Year	*Medicine*	*Life Sciences*	*Chemis-try*	*Physics*	*Engi-neering*	*Social Sciences*	*Other*	*Total*
2000	180	136	186	248	101	87	97	135
2001	174	139	193	247	105	83	97	134
2002	187	140	195	252	98	82	97	136
2003	193	143	200	251	97	78	98	137
2004	207	158	215	259	98	78	106	144
2005	215	157	218	276	99	79	107	147
2006	215	159	220	287	105	80	112	151
2007	206	160	221	295	123	76	111	151
2008	196	149	228	295	120	74	106	145

Table 2. Number of pages per article in the period 2000 to 2008 by area

Year	*Medicine*	*Life Sciences*	*Chemis-try*	*Physics*	*Engi-neering*	*Social Sciences*	*Other*	*Total*
2000	5.2	7.3	6.7	8.2	8.1	8.0	10.6	7.0
2001	5.4	7.5	6.6	8.3	8.0	8.8	10.8	7.2
2002	5.1	7.5	6.5	8.3	8.4	8.6	10.9	7.1
2003	5.2	7.3	6.6	8.2	8.7	8.7	10.4	7.2
2004	5.0	7.3	6.4	8.1	8.8	8.5	10.6	7.1
2005	5.0	7.5	6.5	8.2	8.9	8.5	10.7	7.1
2006	5.1	7.5	6.7	8.1	8.9	8.6	10.5	7.2
2007	5.1	7.8	7.0	8.1	8.8	9.0	10.7	7.3
2008	5.3	7.8	6.8	8.2	9.2	9.0	10.7	7.4

Assessment of the results

The analysis displays a substantial growth of the journals and articles in the database Web of Science in the recent period of 2000 to 2008. By subdividing the journals into different categories it is possible to distinguish the parts of increase referring to a growth of science on the one hand and to the policy of the database producer to achieve a broader coverage of journals. The total growth of articles in the observation period is at a level of 37 percent, thus tremendous in this short time. If the number of articles in the categories “basic” and “new” are added up, the growth rate in this period is 20 percent representing the growth of science. For also in the “old” journals of the category “basic” new topics can appear and imply a growth of articles. Thus about 17 percent of the increase, represented by the categories “additional” and “recurrent”, can be linked to the database policy to enlarge the coverage.

At first sight, it seems to be misleading to associate all newly covered journals with a volume up to 10 to the category “new”. We selected this limit, as lower thresholds such as 5 implied much lower numbers. It has to be taken into account that a newly established journal needs some time for getting broadly known and for attracting a sufficient number of citations. However, a relevant number of citations or more precisely a high “impact factor” is the criterion for being included in WoS (Glänzel and Moed 2002).

The policy of WoS to use the impact factor for “deciding” on the inclusion or exclusion of journals explains the observations referring to the database development to a large extent.

Thus the sharp drop of "basic" journals in the first year reflects that some journals stay in WoS only for a short time due to substantial annual fluctuation of their impact factor. One of the reasons for this fluctuation is the skewness of the distribution of citation rates (Seglen 1992/1997) and extremely highly cited articles can be decisive for the average citation rates in particular of smaller journals. Thus the "basic" journals remaining in WoS for several years represent a kind of stable stem set.
In the differentiation between science and policy induced growth it is important to look at specific areas or fields, as there are substantial disparities.
It is characteristic for all areas that the "stem" journals are substantially larger in terms of annual articles than the newly includes ones in the categories "new", "additional, and "recurring". This is again a consequence of the decision criterion "impact factor". Small specialised journals have a limited readership and therefore attract fewer citations. In consequence the impact factor is lower than for larger journals.
In the context of a policy to enlarge the coverage of the database the main instrument is to lower the threshold of the impact factor for specific areas and fields. In this way, primarily smaller journals are included and the increasing number of articles in the "basic" journals still proves to be a driving force.

Conclusions

By introducing the journal categories "basic", "new", "additional", and "recurrent" it is possible to differentiate between the growth implied by the growth of science and that due to an enlarged coverage of the database for political reasons. The main instrument for enlarging the database coverage is the lowering of the threshold of the impact factor primarily leading to the inclusion of journals of lower size in terms of the annual number of articles.
The assumption of a sharply decreasing number of pages per article cannot be confirmed.
A consequence of the neutral, technical criterion "impact factor" is the similarity of WoS to a living organism where the cells are steadily renewed. In any case, the idea of a stem set which is continuously enlarged proves to be too simplistic.
All in all, the growth of the number of articles in a specific field should not be directly interpreted as growth of scientific activities. Rather the specific growth structures in this field have to be taken into account.

Acknowledgments

The research underlying this paper was supported by the German federal Ministry for Education and Research (BMBF, project number 01PQ08004D). Certain data included in this paper are derived from the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts and Humanities Citation Index (AHCI), and the Index to Social Sciences & Humanities Proceedings (ISSHP) (all update June 2010) prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, Pennsylvania, USA, USA:

References

Glänzel, W.; Moed, H.F. (2002): Journal impact measures in bibliometric research. Scientometrics, 53 (2), pp. 171–193.
Gupta, B.M., Dhawan, S.M. (2008): Condensed matter physics: An analysis of India's research output, 1993-2001. Scientometrics, 75 (1), pp. 123-144.
Larsen, P.O.a , von Ins, M. (2010): The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics, 84 (3), pp. 575-603.
Leydesdorff, L., Cozzens, S., Van den Besselaar, P. (1994): Tracking areas of strategic importance using scientometric journal mappings, Research Policy, 23 (2), pp. 217-229.
Price, d. de Solla (1971): Little Science, Big Science Frankfurt/m.: Suhrkamp.

Seglen, P.O. (1992): The Skewness of Science, American Society for Information Science, Journal, 43 (9), pp. 628-638.
Seglen, P.O. (1997): Why the impact factor of journals should not be used for evaluating research. BMJ, 314, pp. 498–502
Stehr, N. (1994): Knowledge Societies. London/ Thousand Oaks/ New Delhi: SAGE Publications.