

Impact and public acceptance of INIS' bibliographic studies

W.Mandl, INIS, April 2002

Introduction

For more than 30 years, from 1970 to the present, INIS¹ has assembled a bibliographic database in the nuclear field and collected the associated full text of non-conventional literature. 123 member states and international organizations participate in the collaborative system where data are collected and prepared locally and sent to Vienna for quality control and assembly into a common database.

The INIS Bibliographic Database maintains a quality standard and an international coverage of world rank. With well over 2 million data records, it is the most comprehensive bibliographic database in the nuclear field.

In 1998, INIS started to utilize the contents of its database for analysing trends in nuclear related areas. Such bibliometric studies exploit the fine granularity of the database records and are based on the high quality of indexing performed by INIS and its participating members.

INIS' bibliometric studies are available to the public via the Internet². The present evaluation tries to assess the impact and public acceptance of INIS' bibliographic studies, based on their usage as recorded by the Agency's³ Internet server log files.

Scope of INIS' bibliometric studies

The bibliometric studies present a statistical analysis of thousands of database records concerning publications in a particular field covered by INIS' subject scope. The main objectives are to provide statistics and scientific indicators for INIS' users, namely science managers, researchers, engineers, operators, scientific editors and publishers, decision-makers in a particular field.

The quantitative data in these reports are obtained for various properties of relevant INIS records, such as year of publication, secondary subject categories, countries of publication, language, publication types, literary types, etc.

Presently, the following studies are available⁴:

1. [Research Reactor Records in the INIS Database](#)⁵
N. Marinkovic, April 2002
2. [Records related to Nuclear Power Infrastructure in the INIS Database](#)
N. Marinkovic, February 2002
3. [Fast Breeder Power Reactor Records in the INIS Database](#)⁶
N. Marinkovic, August 2000
4. [Fission Reactor Records in the INIS Database](#)
N. Marinkovic, July 2000
5. [Analysis of Condensed Matter Physics Records](#)
C.-D. Hillebrand, March 1999
6. [Science and Technology Indicators in High Energy Physics](#)
C.-D. Hillebrand, November 1998
7. [Science and Technology Indicators in Low and Medium Energy Physics](#)
C.-D. Hillebrand, November 1998
8. [A Survey on Publications in Fusion Research and Technology](#)
C.-D. Hillebrand, October 1998

¹ *INIS*, the International Nuclear Information System is a section within the International Atomic Energy Agency (IAEA), located in Vienna, Austria. (<http://www.iaea.org/inis/>)

² <http://www.iaea.org/inis/ws/articles/index.html>

³ *Agency*, in this context, refers to the International Atomic Energy Agency (IAEA) in Vienna, Austria

⁴ <http://www.iaea.org/inis/ws/articles/index.html>

⁵ This study has been made available only in April 2002 and is therefore not covered by the present report

⁶ This study has been translated to Serbian and is published in the *Journal of the Yugoslav Nuclear Society: 'Nuklearna Tehnologija'* 16(2001), 1, pp. 42-48

One may distinguish two groups of studies: The works by N. Marinkovic (1-4) deal with one of the core nuclear fields (nuclear reactors) whereas the older publications by Hillebrand (5-8) concentrate on fringe areas with a more academic emphasis.

Data preparation

The log files of the Agency's web servers record millions of transactions every month. For the present report only records which document the delivery of one of the 8 bibliometric studies are retained. The 15 month time period from January 2001 to March 2002 is considered.

Records which indicate the occurrence of error conditions are discarded. The only records retained are marked with either the error flag:

- 200 – download without error
- or
- 304 – file not sent because it has not changed since a given date

The latter records (62) are included into the statistical analysis, because they indicate, in most cases, that an intermediate Internet cache server has been able to fulfil the download request of the end-user.

The possible influence of so called web crawlers⁷ on the analysis has also been addressed. However, since all full texts of the bibliometric studies are stored in PDF⁸ rather than HTML⁹, the statistical impact of automated indexers was relatively moderate. Out of 2214 successful file downloads within the 15 month period covered by the present evaluation only 561 (≈25%) were associated to automated indexers¹⁰. Amongst the automated indexers, Google's¹¹ crawler was, with 352 downloads, the most active process. The other active web indexers¹² were associated with companies in the nuclear field or the energy business, their activities are therefore more likely the result of specific web mining activities rather than indiscriminate web indexing. Nevertheless, all log-file records which could be associated with processes which seem to download files automatically were excluded from further analysis. Most of the well known Internet search engines showed download activities with HTML files only¹³ and had therefore no impact on the present analysis. 1653 records, all referring to successful downloads of full texts, were retained for further analysis.

Results

During the 15 month time period from January 2001 to March 2002 the full texts of the bibliographic studies were successfully down-loaded 1653 times (see above). This number includes 70 downloads from within the IAEA, the rest were by outside users.

⁷ Crawlers (also known as robots or spiders) are automated processes which download a large number of files regularly for web indexing purposes.

⁸ Adobe Portable Data Format

⁹ HyperText Markup Language

¹⁰ Automatic indexers were identified by their characteristic signatures in the log files: indiscriminate downloading of many files in regular time intervals.

¹¹ <http://www.google.com/>

¹² Fortum Group (<http://www.fortum.com/>), Energy E-Comm (<http://www.energyecomm.com/>), Nuclear Waste Management System (<http://198.22.87.207>)

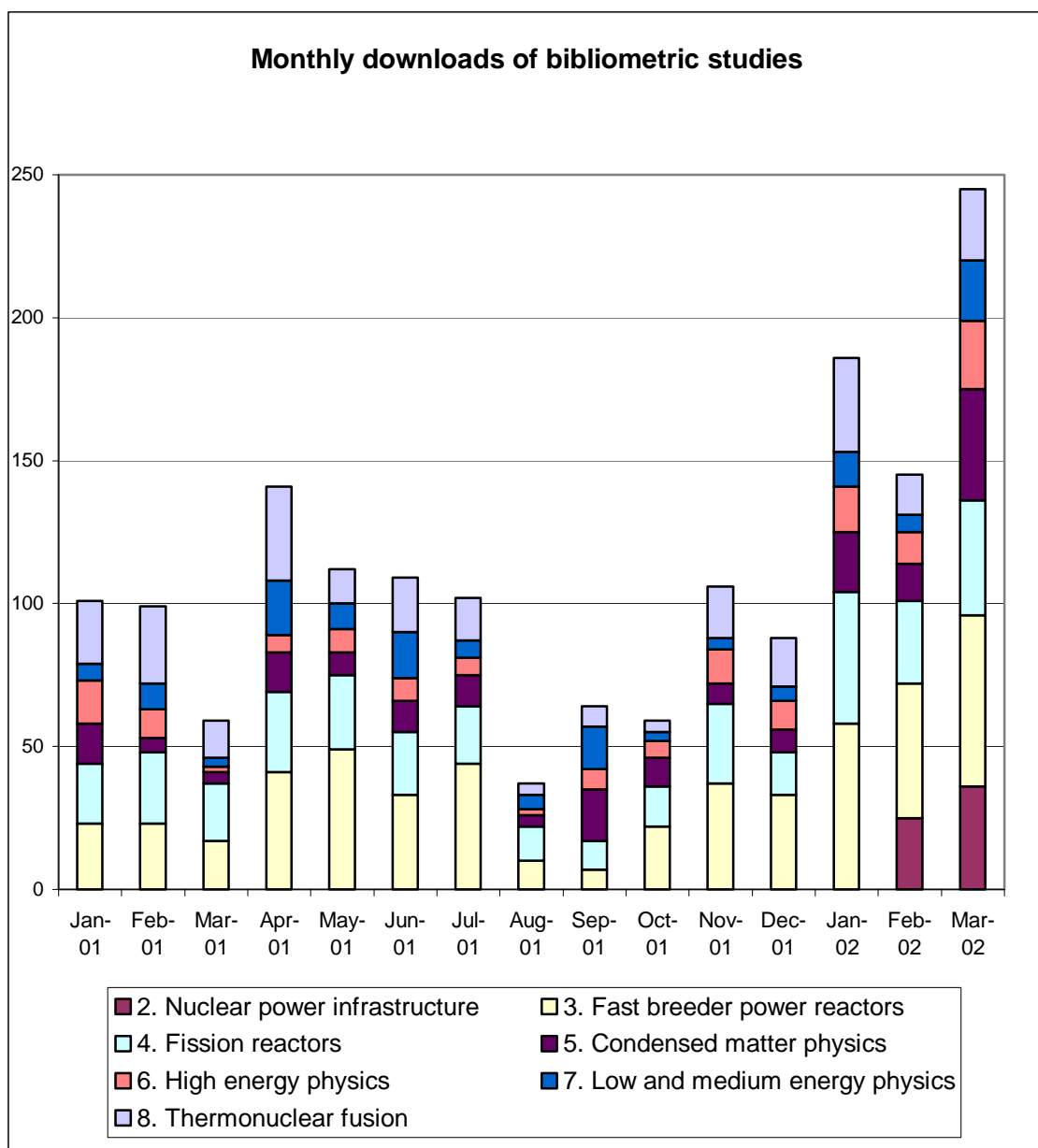
¹³ Inktomi, Fastsearch, Northernlight, Openfind and Altavista were the most prominent general purpose Internet search services found in the log files.

The total number of downloads during the 15 months and the average number of downloads per month are respectively:

	total	per month
• Records related to Nuclear Power Infrastructure in the INIS Database	61	30.5
• Fast Breeder Power Reactor Records in the INIS Database	504	33.6
• Fission Reactor Records in the INIS Database	356	23.7
• Analysis of Condensed Matter Physics Records	187	12.5
• Science and Technology Indicators in High Energy Physics	143	9.5
• Science and Technology Indicators in Low and Medium Energy Physics	139	9.3
• A Survey on Publications in Fusion Research and Technology	263	17.5

A difference in popularity between the studies is apparent. The more recent works on nuclear reactors are downloaded most often. Second comes the study on nuclear fusion and least popular are the works on high, medium and low energy physics.

As a function of time, we observed fairly stable web traffic during the year 2001 with an increase in 2002.



It is also interesting to study the download pattern of individual visitors. A total of 674 different visitor's domain names has been identified. By far the most common scenario is that only one single document is downloaded per domain name, as shown in the table below:

number of documents downloaded	from each of this many domain names
1	441
2	86
3	78
4	14
5	4
6	10
7	4
8	4
9	9
10	3
11	1
12	5
15	1
17	1
18	1
21	1
23	1
24	2
25	1
26	1
28	1
33	1
35	1
36	1
39	1
64	1

It is easy to understand that there are also a sizable number of domains with multiple downloads. The main reasons for multiple downloads will be either a situation where the end user is interested in more than one bibliographic study or where an individual person is reading a given document on the computer screen while online. In the latter case multiple downloads of the same document would be registered if the user closes the web browser between study sessions.

In order to understand the occurrence of the occasional cases where many downloads are associated with one single domain name, we have to recall that in some cases a large number of individuals have access to the Internet via one single domain name. In the case of the last entry in the above table, for example, 64 downloads are associated with the single domain name of the 'Fachhochschule Eisenstadt' an Austrian institution of higher education which might have a large number of students.

Conclusions

INIS' Bibliographic studies are appreciated by a large audience¹⁴. This is well documented by the consistently high volume of downloads of the full texts in PDF. Recent trends show even a noticeable increase of demand during the first months of this year. There is a higher demand for studies which concern core areas of the nuclear field than for fringe areas.

¹⁴ In addition to access via the Internet, bibliographic studies were also distributed on CD-ROMs during conferences.