30 November 1998

Analysis of High Energy Physics Records in Databases

Science and Technology Indicators in High Energy Physics

Claus-D. Hillebrand

Vienna, Austria

November 1998

Abstract

An analysis of the literature on high energy physics was performed on the basis of the contents of the bibliographic database INIS (International Nuclear Information System). Quantitative data were obtained on various characteristics of the relevant INIS records such as subject categories, language and country of publication, publication types, etc. It was found that the number of records in high energy physics has increased over the last two decades. The analysis opens up the possibility for further studies, e.g. on international research co-operation and on publication patterns.

Contents

ANNEX

Introduction

High energy physics publications are stored in bibliographic databases such as the International Nuclear Information System (INIS). The objective of this study is to quantify and analyse bibliographic records on the physics of elementary particles and fields (hereafter simply referred to high physics records) and to offer an overview of the developments in this research field. For the first time, a scientometric study (p.15) has been performed to investigate a selected field of science and technology and the INIS database has been used as a source of data. A variety of science and technology indicators are retrieved. The possible applications of this study are outlined. The scope of high energy physics records in INIS is described in the Categories section. (A description of INIS is given in the Annex).

High energy physics lies within the scope of INIS and represents about 7% of the whole INIS database (more than 135 000 high energy physics relevant records were entered in the period from 1970 to mid-1998). In this field, there is an input of 6000--8000 records every year. Fig. 1 shows the time development of these records over the publication year for the last 28 years. In the 1970s, the increase in the number of records was due to the start-up of INIS. In the 1980s, there were 4000--6000 records per year. In the 1990s, the number of records per publication year totalled between 6000--7000. The year 1992 was exceptionally high, with more than 8000 records. Over the last 28 years, the average number of records per year increased steadily. The decline between 1973 and 1976 is not understood. The input for the last 3--4 years is still continuing as the input preparation of each publication represents an extra step. The projection of input for the last four years is indicated by the dashed lines.

Four Member States provide about 76% of the INIS input in high energy physics, Fig. 2. High energy physics records come from 66 different input centres (INIS members which also include international organizations such as CERN, JINR) in the period between 1970 and mid--1998. It is to be noted that the number of publications per country reflects the concentration of scientific publishing houses in those countries rather than research activities (see Author section).

Language

About 88% of all documents related to high energy physics are published in English. This includes translated publications. These are mainly published in the United States of America and these records are therefore provided by the INIS centre in the USA. Translated records represent nearly 10% of the input from the USA. Of all the authors listed in the high energy physics records, roughly 60% are from non-English speaking countries. Altogether there are high energy physics records in 28 different languages (Fig. 3).

Categories

INIS records are categorized according to the INIS subject categories and scope descriptions arranged in conformity with the International Classification System for Physics developed by the International Council for Scientific and Technical Information. The physics category scheme was changed in 1992. For that reason, the following analysis mainly covers the period from 1992 until mid-1998, with more than 41 000 records. The main subject category of interest is G2000 Physics-of-elementary-particles-and-fields, the subcategories are: G2100 General-theory-of-particles-and-fields, G2110 Theory-of-fields-and-strings, G2120 Symmetry,-conservation-laws,-currents-and-their-properties, G2130 S-matrix-theory,-relativistic-scattering-theory, G2200 Specific-Theories-and-Interaction-Models;-Particle-

Systematics, G2210 Unified-theories-and-models, G2220 Quantum-electrodynamics, G2230 Quantum-chromodynamics, G2240 Models-for-strong-interactions, G2300 Specific-Interactions,-Decays-and-Processes, G2310 Weak-and-electromagnetic-interactions-of-leptons, G2320 Neutrino-interactions, G2330 Photon-and-charged-lepton-interactions-with-hadrons, G2340 Hadron-interactions, G2350 Decays-of-mesons, G2360 Decays-of-baryons, G2370 Decays-of-leptons, G2380 Decays-of-intermediate-bosons, G2390 Electromagnetic-processes-and-properties, G2400 Properties-of-Specific-Particles-and-Resonances, G2410 Properties-of-baryons-and-baryon-resonances, G2420 Properties-of-mesons-and-meson-resonances, G2430 Properties-of-leptons, G2440 Properties-of-other-particles-incl.-hypothetical-particles; see Annex: INIS subject scope: high energy physics subject categories in INIS see Table I.

In the period from 1992 to mid-1998, the subject fields with the highest number of records were G2110 Theory-of-fields-and-strings, G2340 Hadron-interactions, G2230 Quantum-chromodynamics, G2120 Symmetry,-conservation-laws,-currents-and-their-properties, (Fig. 4).

The time development of records within the subfields in general shows a small fluctuation in yearly number of records in high energy physics within the considered time period (1991--1996).

The multidisciplinary nature of the INIS database allows the study of the correlation between scientific disciplines. For each record, up to three subject categories can be assigned, if the record covers more than one subject field. The high energy physics records with second and third categories have the following subject fields assigned (listed in order of importance): (Category G) other physics (within this category by order of importance: general physics, nuclear physics, atomic and molecular physics, condensed matter physics and plasma physics followed by (Category E) engineering and technology and (Category F) other aspects of nuclear and nonnuclear energy and (Category B) chemistry, materials and earth sciences (Category B), the other two categories: Life and Environmental Sciences (Category C) and Isotope and Radiation and Isotope Applications (Category D) are in with respect to number of records negligible, Fig. 5.

Publication types

The record type (journal articles, reports, books, miscellaneous and patents) and literary type (e.g. short communications, conferences, numerical data and progress reports) of each record entry are indicated in the database. This allows the publishing format to be characterized (Table II).

Journal articles represent about 64% of all records, reports 22%, books 10%, miscellaneous 4% and patents 0.001% (Fig. 6). The percentage of journal records seems high when compared with other physics disciplines. This may be due to the high number of records in theoretical high energy physics categories (see subject category section and the discussion in the next paragraph). The number of report records is relatively high. Reports are usually produced in research centres in which many device specific documents and experimental results such as data tables are generated. The readership of these documents is small and they are not suitable for publication in journals or books because of their length, technical content, etc. The `miscellaneous´ type is often used for progress reports, listing of numerical data and dissertations. This publication type is low but this fact has to be seen with respect to the high number of journal records. The high number of book records results from the publication format of some conference proceedings in which each contribution counts as a book record. Also, under journal articles and reports one can find a high number of conference contributions, numerical data and short communications (Fig. 7, records with two publication types). The input of patents covered has been somewhat erratic over the years. This has to do with the change of patent records copyright in some countries and the difficulty of converting records from patent to bibliographic databases. The number of patent records in high energy physics are low when compared with other physics fields. No logical explanation of this has been found.

The number of journal records in percentage terms is higher (about 30%) in the field of high energy physics than in plasma physics and fusion R&T, whereas in the latter the number of report and patent records is higher. Both fields have almost the same total number of records. Regarding the number of records distributed over the subject categories, the high number of journal records in high energy physics is probably due to a high proportion of theoretical physics publications in this field. Research on theoretical physics can be done at many places, many authors of these publications work at universities and not necessarily at facilities (see Authors section). This fact, and the fact that publications on high energy theoretical physics are of interest to a large scientific community, explains why journals are preferably chosen as a medium for publication.

The time development of publication types gives an indication of research activities. The number of journal articles varied by 10--15% between 1973 and 1990. The number of journal records per publication year entered in the INIS database averaged around 3000 in that period. A steep increase in number of records (up to 5000) was observed between 1990 and 1992. The number of journal records remained high. The number of report records has increased since 1988. It has to be noted that the number of reports made available on the Internet increased which are not entered into the INIS database. The frequency of books records over the publication years is very irregular, the reason probably being the irregular choice of formats in publishing conference proceedings.

Authors

The country tag in the author field indicates the actual national research activities better than the country in which the document has been published. the actual national research activities. More than 73% of the authors come from eight countries (with more than 5% of the total number of records per country), (Fig. 8). The distribution of authors according to country is different from the distribution of input countries because in some of these there is a high concentration of science publishing houses. The number of different countries and international organizations (including CERN and JINR) from which authors are publishing totals 130. This is a very high number. About 52% of the publications have at least one author of the document who is affiliated to a university. This indicates that less than 50% of authors working in the field of high energy physics are affiliated to institutions other than universities.

Journal statistics

Journal articles on high energy physics are published in more than 1500 different journals and represent more than 64% of all high energy physics records in INIS. The number of high energy physics relevant records can be found in core high energy physics journals and in general physics journals, as well as in national physics journals and in neighbouring disciplines. The few core journals have a high number of records and comprise more than 50% of all journal records. A detailed analysis can be performed in a specific study on high energy physics journals in which the most frequent and average page number per journal articles can be determined.

A profile of the main journals can also be plotted against the high energy physics subfields. The profiles allow comparison of the scope of each journal. The list of high energy physics journals, a ranking of journals by the number of records (which is a function of publication years, input years, articles published per year and scope) can be compared with the list of the Science Citation Index (SCI) of the Institute of Scientific Information (Philadelphia, USA). The comparison made in fusion R&T shows that, for instance, the scope of the SCI list in fluids and plasmas (not controlled fusion itself) is broad but does not cover certain fields such as material studies, engineering, etc. Furthermore, fusion technology journals are not separated from fusion research journals in the SCI list. That means that the publications of high energy physicists are not totally covered by the SCI Plasma Physics Category.

Keywords and Free Text

A common feature of a bibliographic database is the subject indexing of records by the assigning of keywords. As the subject index is used in books, each database record is complemented by a list of 'controlled terms' (keywords, or in INIS terminology - descriptors) which are chosen to describe better the content, concepts, methods and models. These descriptors are scientific and technical words listed in the INIS thesaurus, which also defines relationships (hierarchical, affinitive, etc.) to other descriptors. The descriptors are used for the retrieval of documents. Descriptors are assigned to each input record by indexers working in each INIS centre. Descriptors in high energy physics records in INIS database indicate that the main emphasis of the records is on high energy physics.

An alternative retrieval tool is the search by `free text´ (that is, natural language words and phrases occurring in all textual fields, including titles and abstracts). The free text can be a scientific term which appears in the title or abstract and is not necessarily a descriptor but nevertheless, can be used for retrieval. In addition to the use of descriptors, `non-standard keywords´ (in INIS terminology - free text terms) are permitted to be input in another indexing field and allow flexibility of indexing and searching. Newly proposed descriptors are usually accepted with a delay of several months.

In the high energy physics records (from 1992 to present), some elements of the periodic table have a high frequency (number of records) in the textual fields (in order of importance) such as hydrogen, lead, deuterium, helium, carbon, gold, silicium etc. (Table III).

Outlook

This analysis of high energy physics records in the INIS database contains many tables and graphs, which form the basis of this summary, and which provide more detailed information. A basic analysis was performed, aimed at different interest groups such as the scientific and technology community, science publishers and editors, librarians and science managers.

In the study, additional information on science and technology indicators and trends is also shown, as well as information on high energy physics related publications and their formats. Furthermore, more advanced and focused analyses and evaluation of the data for some of these interest groups are also possible. The analysis opens the possibility of further studies, e.g. the co-operation between different institutions and countries, mapping publication patterns, highlighting scientific co-operation, development of human resources and journal structure.

Scientometric studies can assist in analyses and formulation of science and technology policy by mapping changes in research activities, providing thematic and strategic analysis of relative position of research communities and sketching profiles of activities and performance of countries and institutions.

Conclusion

The number of publications in high energy physics increased steadily over the last two decades. The number of countries from which authors are publishing totals 130. High energy publications can be found in 28 different languages. More than 73% of the authors of high energy publication come from eight countries.

References

- Low and Medium Energy Physics Records, Claus D. Hillebrand, NDS Collection, IAEA Vienna, 1998

- Fusion Research and Technology Records in INIS Database, C.D. Hillebrand; Div. of Scientific and Technical Information; IAEA, Vienna, Austria. ITER-Newsletter, April 1998; p.4-6

Annex

Table I

High Energy Physics Subject Categories in INIS

-- G2000 PHYSICS-OF-ELEMENTARY-PARTICLES-AND-FIELDS
-- G2100 General-theory-of-particles-and-fields
-- G2110 Theory-of-fields-and-strings
-- G2120 Symmetry,-conservation-laws,-currents-and-their-properties
-- G2130 S-matrix-theory,-relativistic-scattering-theory
-- G2200 Specific-Theories-and-Interaction-Models;-Particle-Systematics
-- G2210 Unified-theories-and-models
-- G2220 Quantum-electrodynamics
-- G2230 Quantum-chromodynamics
-- G2240 Models-for-strong-interactions
-- G2300 Specific-Interactions,-Decays-and-Processes
-- G2310 Weak-and-electromagnetic-interactions-of-leptons
-- G2320 Neutrino-interactions
-- G2330 Photon-and-charged-lepton-interactions-with-hadrons
-- G2340 Hadron-interactions
-- G2350 Decays-of-mesons
-- G2360 Decays-of-baryons
-- G2370 Decays-of-leptons
-- G2380 Decays-of-intermediate-bosons
-- G2390 Electromagnetic-processes-and-properties
-- G2400 Properties-of-Specific-Particles-and-Resonances
-- G2410 Properties-of-baryons-and-baryon-resonances
-- G2420 Properties-of-mesons-and-meson-resonances
-- G2430 Properties-of-leptons
-- G2440 Properties-of-other-particles-incl.-hypothetical-particles

ANNEX

About INIS Database

The decentralized multidisciplinary bibliographic database of the IAEA is a part of **INIS** which was created in 1970 and is administered by the INIS Section of the IAEA with the purpose of collecting and disseminating information on science and technology through its Member States.

INIS has 120 Members including 18 International Organizations which provide records on science and technology documents published in the states where the 120 INIS members are located. Records of documents are provided to INIS in English, along with the titles in the language of origin.

The main INIS fields of scope are: (i) chemistry, materials and earth sciences; (ii) life and environmental sciences; (iii) isotopes, isotope and radiation applications; (iv) engineering and technology; (v) other aspects of nuclear and non-nuclear energy; (vi) physics.

The largest subject category is physics with about one third of all records, followed by engineering and technology with one fourth. Chemistry, material and earth sciences as well as life and environmental sciences represent about one fifth each.

Annex

Definition of Scientometrics and Bibliometrics

The terms bibliometrics and scientometrics were introduced almost simultaneously by Pritchard and by Nalimov and Mulchenko in 1969. While Pritchard explained the term bibliometrics as “the application of mathematical and statistical methods to books and other media of communication, Nalimov and Mulchenko defined scientometrics as “the application of those quantitative methods which deal with the analysis of science viewed as an information process. According to these interpretations, scientometrics is restricted to the measurement of science communication, whereas bibliometrics is designed to deal with more general information processes. The at best fuzzy distinction between the two has virtually disappeared over the course of the last three decades and, today, the terms are more or less synonymous. Meanwhile, the term infometrics has come to replace the originally broader specialty of bibliometrics.

(Source: 2nd European Report on Science and Technology Indicators, Dec. 1997, page 111, EC-Luxembourg, EUR17639)

Table II Publication Types

Records Types

J	Journals
R	Reports
B	Books
I	Miscellaneous
P	Patents

Literary Types

E	Short Communication
N	Numerical Data
V	Computer Program Description
X	Nonconventional Literature
Y	Progress Reports
U	Dissertations
K	Conference
Z	Bibliography

Annex

Table III

Frequency of Periodic Elements in High Energy Physics Records (since 1992; total 41462)

Alphabetic Order		Number of Records	
1	Actinium	2373	Hydrogen
79	(Aluminium or Aluminum)	998	Lead
2	Americium	489	Deuterium
1	Antimony	254	Helium
42	Argon	227	Carbon
5	Arsenic	206	Gold
1	Astatine	185	Silicon
7	Barium	156	Beryllium
0	Berkelium	119	Oxygen
156	Beryllium	119	(Sulphur or Sulfur)
13	Bismuth	98	Copper
0	Bohrium	96	Silver
53	Boron	90	Iron
21	Bromine	90	Uranium
22	Cadmium	90	Tritium
32	(Caesium or Cesium)	82	Gallium
38	Calcium	79	(Aluminium or Aluminum)
3	Californium	63	Germanium
227	Carbon	58	Lithium
3	Cerium	56	(Tungsten or Wolfram)
53	Chlorine	53	Boron
13	Chromium	53	Chlorine
10	Cobalt	49	Nitrogen
98	Copper	47	Neon
0	Curium	42	Argon
0	Dubnium	40	Xenon
4	Dysprosium	38	Calcium
0	Einsteinium	32	(Caesium or Cesium)
0	Erbium	30	Nickel
2	Europium	25	Fluorine
0	Fermium	22	Cadmium
25	Fluorine	21	Bromine
0	Francium	20	Iodine
8	Gadolinium	20	Titanium
82	Gallium	18	Tin
63	Germanium	16	Tantalum
206	Gold	16	Thallium
0	Hafnium	15	Molybdenum
0	Hassium	14	Mercury
254	Helium	13	Bismuth
7	Holmium	13	Chromium
2373	Hydrogen	13	Lanthanum
5	Indium	13	Platinum
20	Iodine	12	Magnesium
3	Iridium	12	Thorium
90	Iron	10	Cobalt
10	Krypton	10	Krypton
13	Lanthanum	8	Gadolinium
0	Lawrencium	8	Phosphorus
998	Lead	8	Sodium
58	Lithium	7	Barium
5	Lutetium	7	Holmium
12	Magnesium	7	Potassium
5	Manganese	6	Tellurium
0	Meitnerium	6	Zinc
0	Mendelevium	5	Arsenic
14	Mercury	5	Indium
15	Molybdenum	5	Lutetium
1	Neodymium	5	Manganese

Table III continued

47	Neon
1	Neptunium
30	Nickel
3	(Niobium or Columbium)
49	Nitrogen
0	Nobelium
2	Osmium
119	Oxygen
5	Palladium
8	Phosphorus
13	Platinum
3	Plutonium
0	Polonium
7	Potassium
0	Praseodymium
0	Promethium
1	Protactinium
2	Radium
3	Radon
4	Rhenium
0	Rhodium
1	Rubidium
2	Ruthenium
0	Rutherfordium
2	Samarium
3	Scandium
0	Seaborgium
4	Selenium
185	Silicon
96	Silver
8	Sodium
2	Strontium
119	(Sulphur or Sulfur)
16	Tantalum
2	Technetium
6	Tellurium
4	Terbium
16	Thallium
12	Thorium
0	Thulium
18	Tin
20	Titanium
56	(Tungsten or Wolfram)
0	Ununbium
0	Ununnilium
0	Unununium
90	Uranium
4	Vanadium
40	Xenon
4	Ytterbium
3	Yttrium
6	Zinc
4	Zirconium
489	Deuterium
90	Tritium

5	Palladium
4	Dysprosium
4	Rhenium
4	Selenium
4	Terbium
4	Vanadium
4	Ytterbium
4	Zirconium
3	Californium
3	Cerium
3	Iridium
3	(Niobium or Columbium)
3	Plutonium
3	Radon
3	Scandium
3	Yttrium
2	Americium
2	Europium
2	Osmium
2	Radium
2	Ruthenium
2	Samarium
2	Strontium
2	Technetium
1	Actinium
1	Antimony
1	Neodymium
1	Neptunium
1	Protactinium
1	Rubidium
1	Astatine
0	Berkelium
0	Bohrium
0	Curium
0	Dubnium
0	Einsteinium
0	Erbium
0	Fermium
0	Francium
0	Hafnium
0	Hassium
0	Lawrencium
0	Meitnerium
0	Mendelevium
0	Nobelium
0	Polonium
0	Praseodymium
0	Promethium
0	Rhodium
0	Rutherfordium
0	Seaborgium
0	Thulium
0	Ununbium
0	Ununnilium
0	Unununium

Annex

Figure Caption:

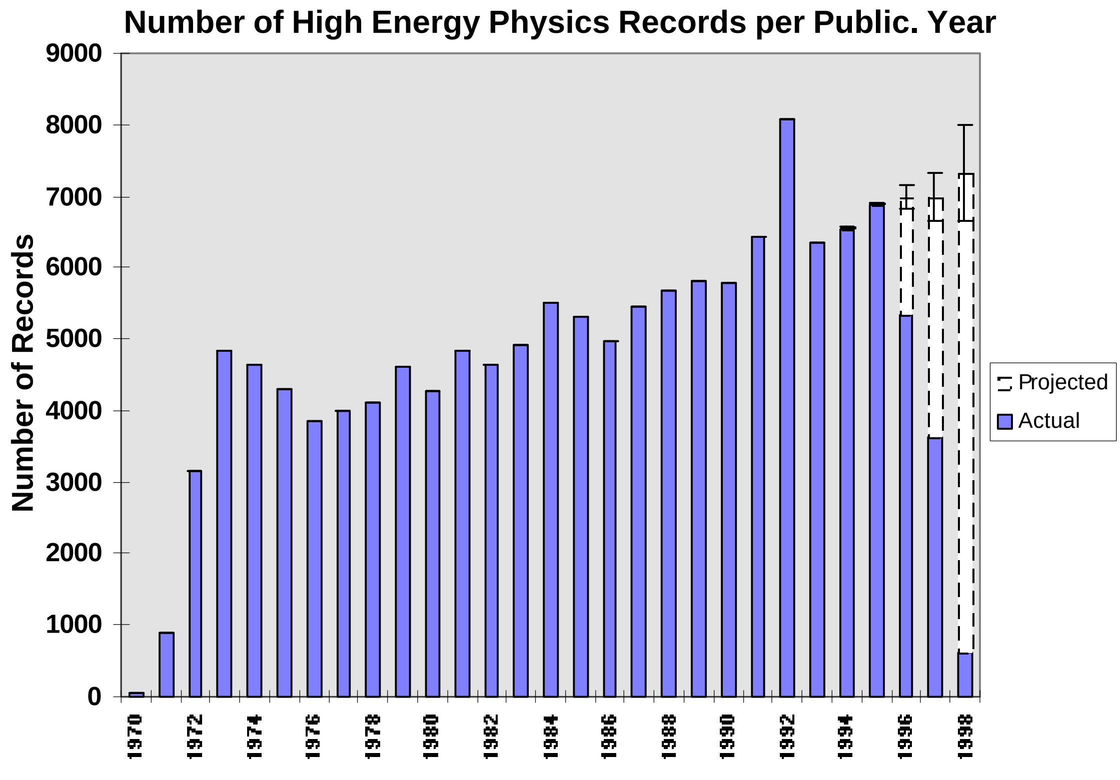

Number of High Energy Physics Records per Public. Year
Number of Records
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
Projected
Actual
Publication Year

Fig. 1

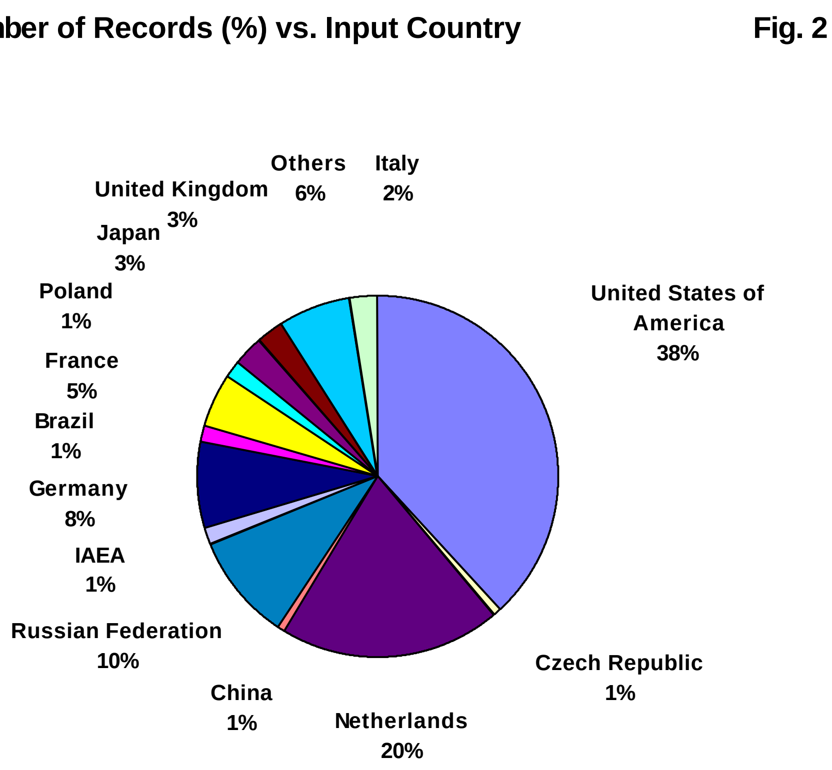

Number of Records (%) vs. Input Country
Others 6%
Italy 2%
United Kingdom 3%
Japan 3%
Poland 1%
France 5%
Brazil 1%
Germany 8%
IAEA 1%
Russian Federation 10%
China 1%
Netherlands 20%
Czech Republic 1%
United States of America 38%

Fig. 2

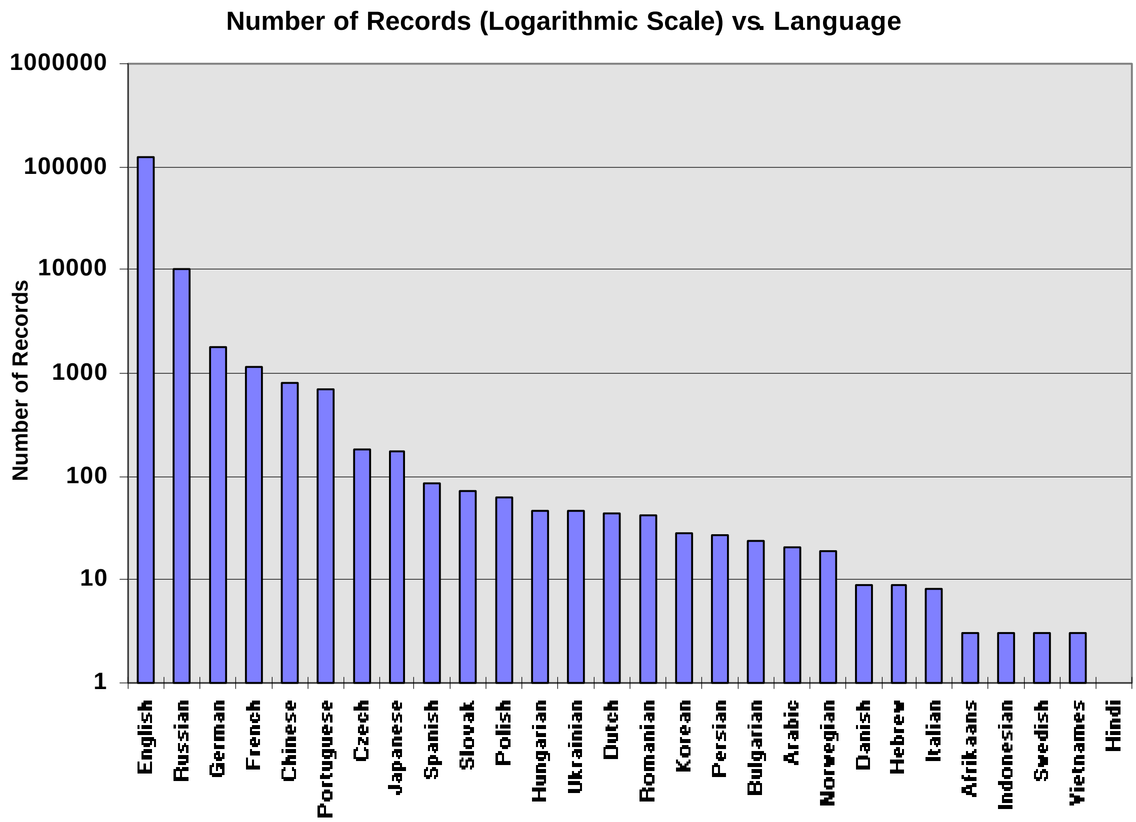

Number of Records (Logarithmic Scale) vs. Language
Number of Records
1000000
100000
10000
1000
100
10
1
English
Russian
German
French
Chinese
Portuguese
Czech
Japanese
Spanish
Slovak
Polish
Hungarian
Ukrainian
Dutch
Romanian
Korean
Persian
Bulgarian
Arabic
Norwegian
Danish
Hebrew
Italian
Afrikaans
Indonesian
Swedish
Vietnames
Hindi
Language

Fig. 3

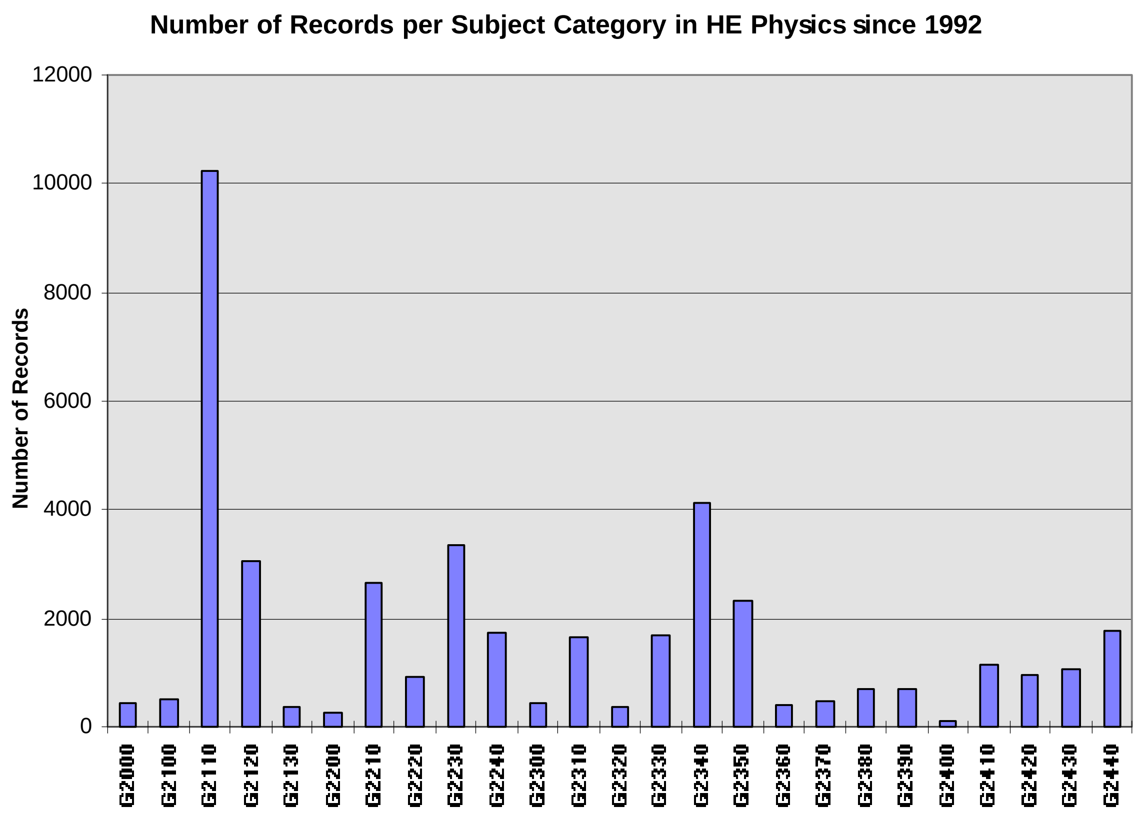
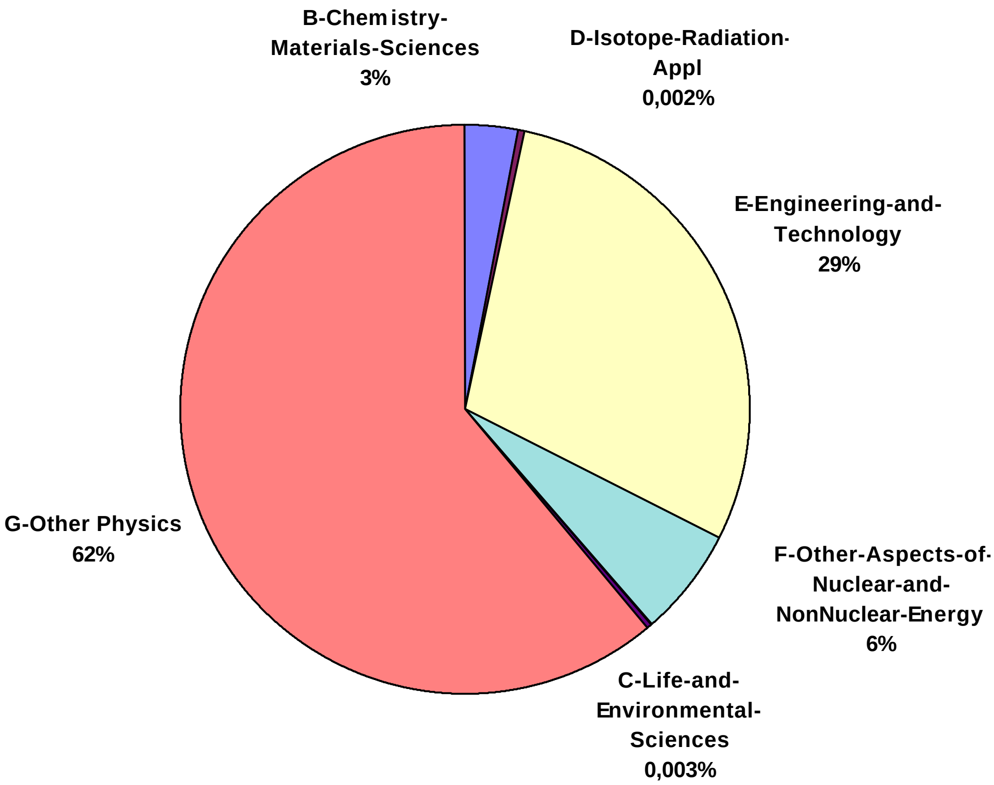
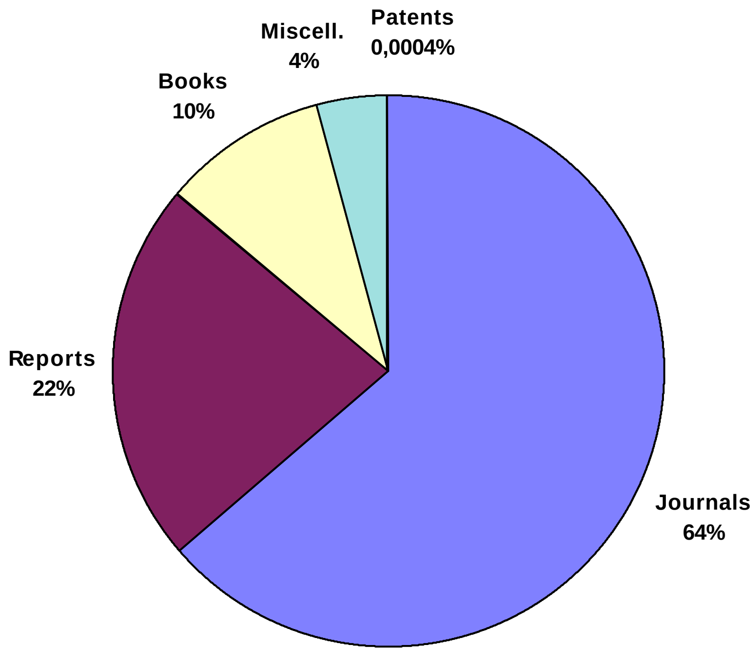

Number of Records per Subject Category in HE Physics since 1992
Number of Records
12000
10000
8000
6000
4000
2000
0
G2000
G2100
G2110
G2120
G2130
G2200
G2210
G2220
G2230
G2240
G2300
G2310
G2320
G2330
G2340
G2350
G2360
G2370
G2380
G2390
G2400
G2410
G2420
G2430
G2440
Subject Categories (see Annex)

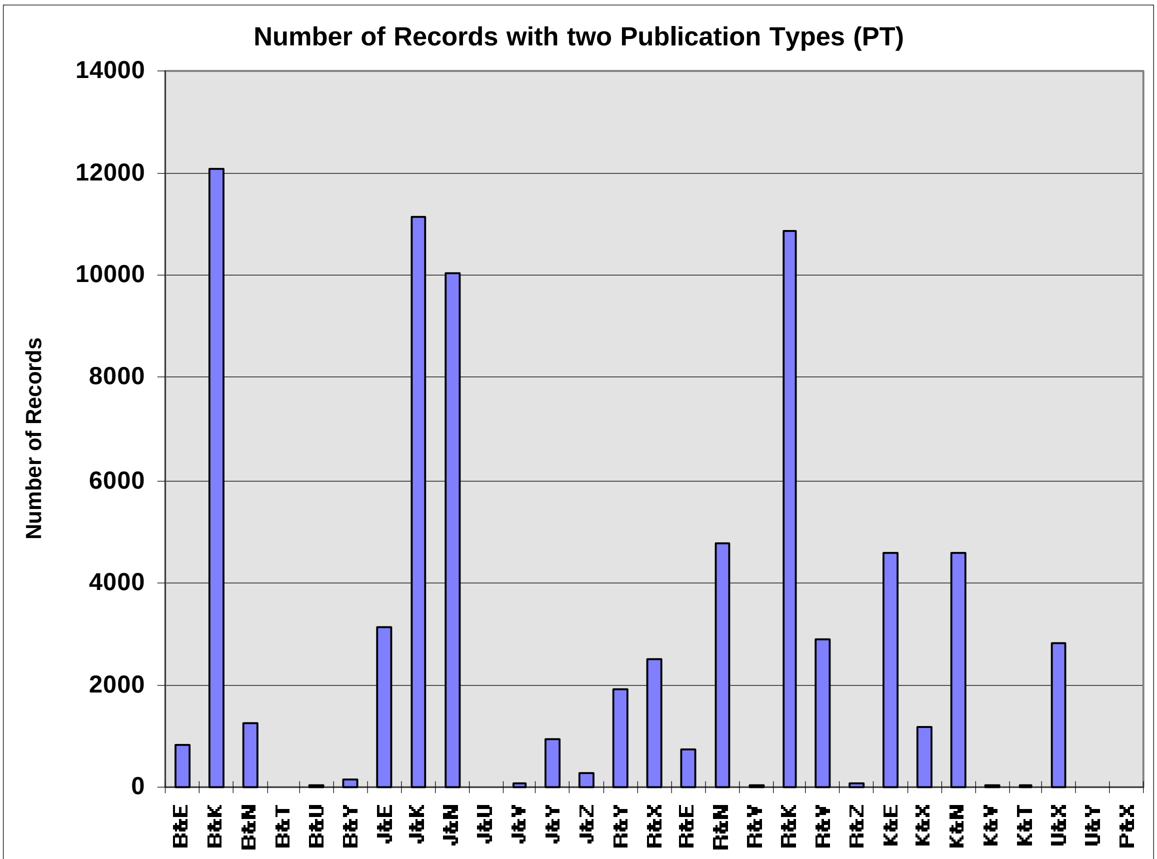
Fig. 4

Number of Records (%) with Secondary Category **Fig. 5**



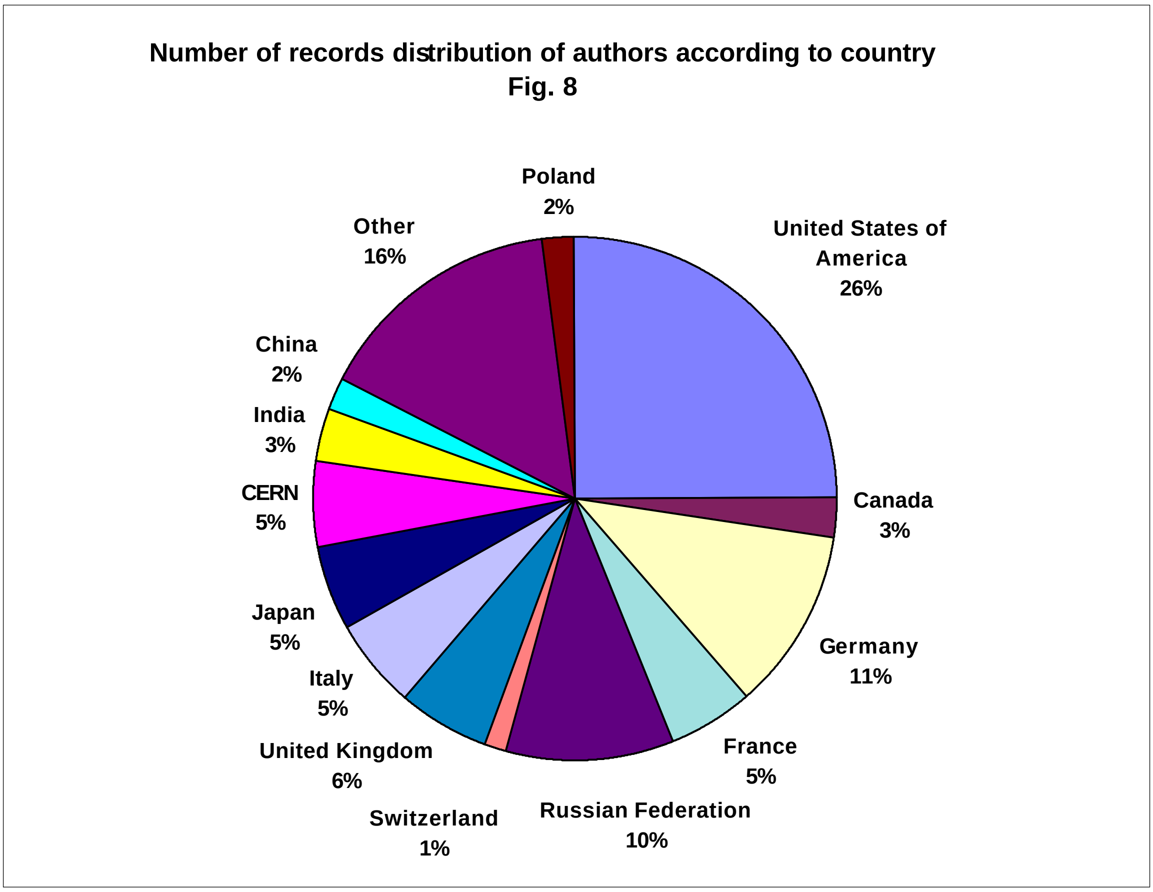
Number of Records (%) per Publication Type **Fig. 6**





Number of Records with two Publication Types (PT)
Number of Records
14000
12000
10000
8000
6000
4000
2000
0
B&E
B&K
B&N
B&T
B&U
B&Y
J&E
J&K
J&N
J&U
J&V
J&Y
J&Z
R&Y
R&X
R&E
R&N
R&V
R&K
R&V
R&Z
K&E
K&X
K&N
K&V
K&T
U&X
U&Y
P&X
Two Publication Types

Fig. 7



Number of records distribution of authors according to country
Poland 2%
United States of America 26%
Canada 3%
Germany 11%
France 5%
Russian Federation 10%
Switzerland 1%
United Kingdom 6%
Italy 5%
Japan 5%
CERN 5%
India 3%
China 2%
Other 16%

Fig. 8