

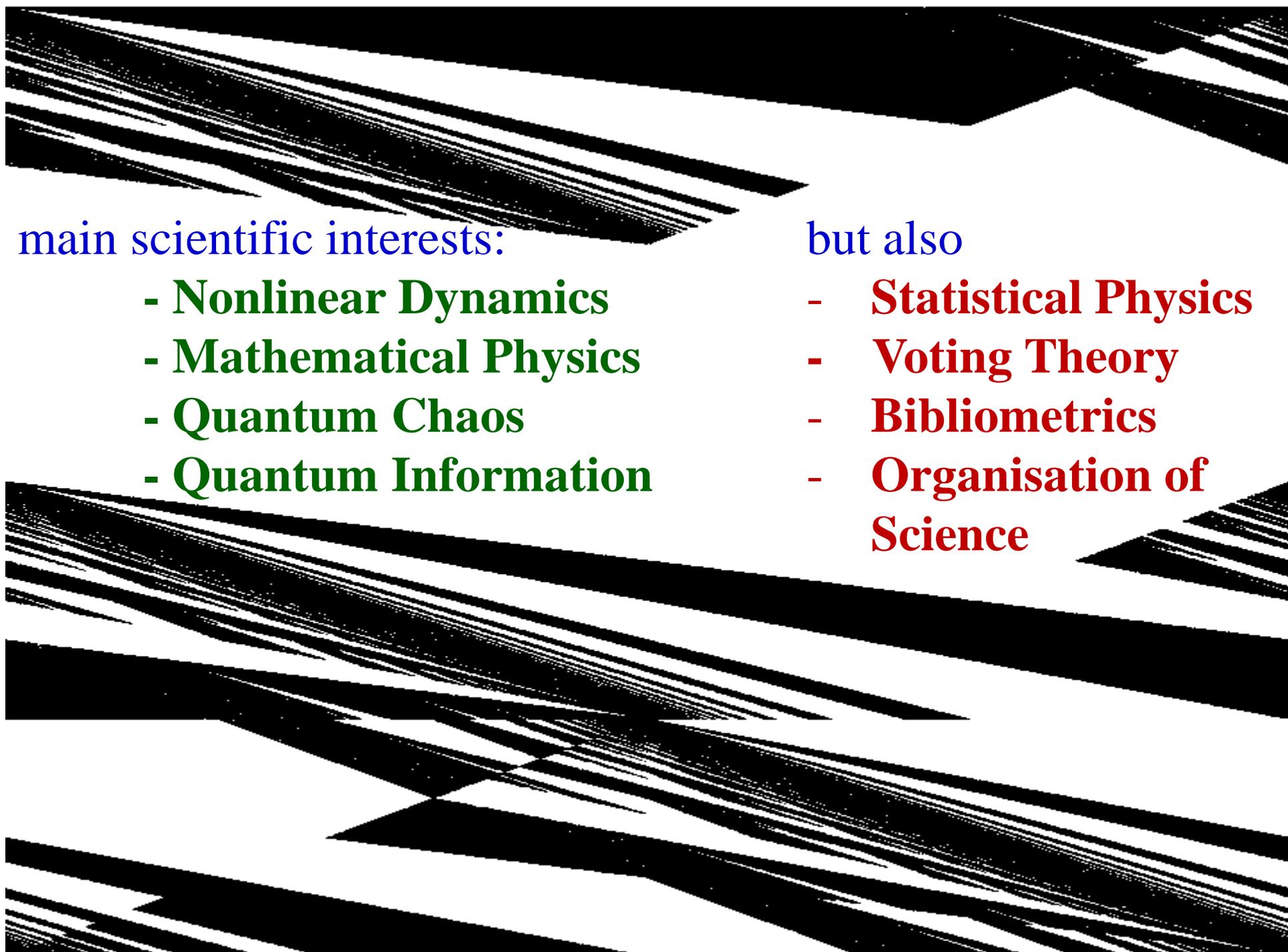
# **Open Access Journals:** **A physicists point of view**

**Karol Życzkowski**

**(Jagiellonian University and  
Center for Theoretical Physics, PAS)**

*Politechnika Krakowska*

*September 26, 2016*



main scientific interests:

- **Nonlinear Dynamics**
- **Mathematical Physics**
- **Quantum Chaos**
- **Quantum Information**

but also

- **Statistical Physics**
- **Voting Theory**
- **Bibliometrics**
- **Organisation of Science**

# Why we do explore space



?

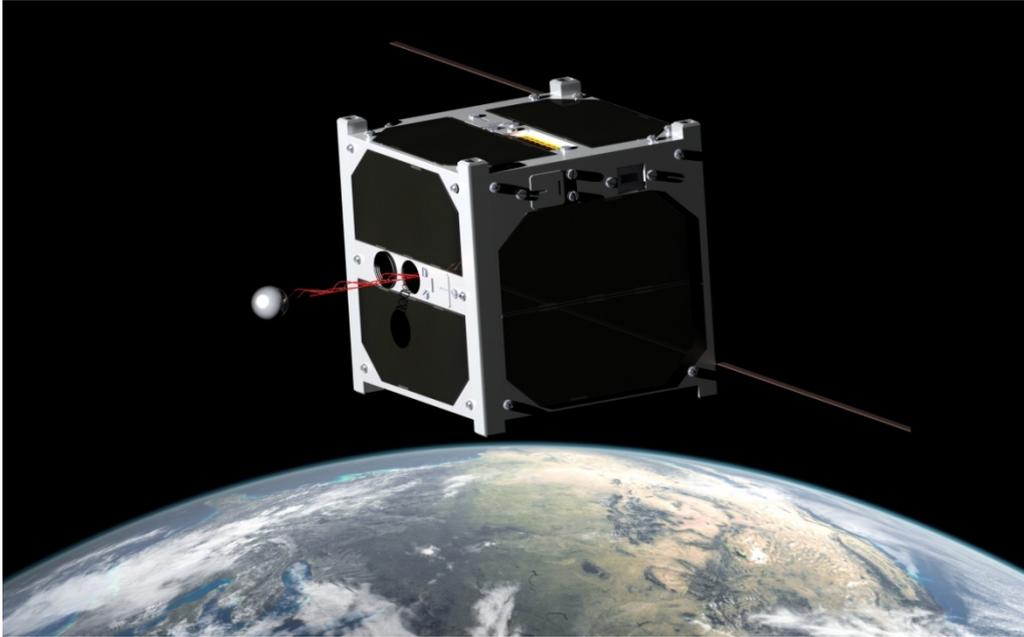


# Why we do explore space



**Beyond standard key goals  
we benefit numerous side effects:  
new technologies, new materials ...**

# New technologies become accesible

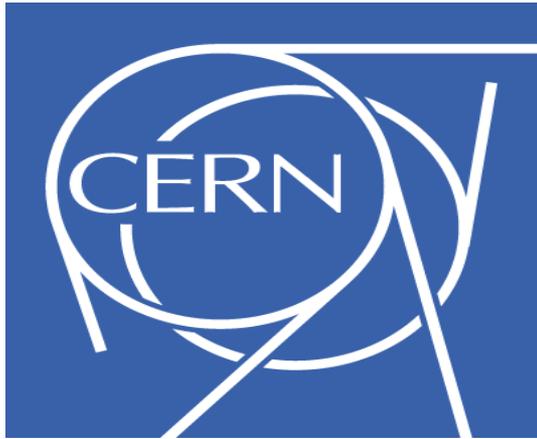


**Mini satellites (*cubesats* - mass around 1 kg) constructed by amateurs can be nowadays sent into the space for circa **10 000 US \$****

**What is physics  
(and physicists)  
good for**



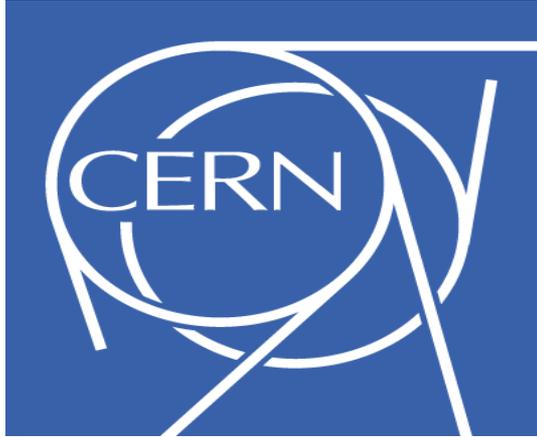
# What is CERN good for



**Conseil  
Européenne  
pour la Recherche  
Nucléaire**

**Geneva  
Switzerland**

# What is CERN good for ?



**Not only to find new elementary particles  
(Higgs boson !)  
and discover laws of high energy physics,  
but also...**

# What is CERN good for



**Not only to find new elementary particles**

**(Higgs boson !)**

**and discover laws of high energy physics,  
but also...**

**to make ground-breaking**

**side discoveries !**

# CERN

the place where **www** was born !

**W**orld **W**ide **W**eb created in 1990 by



**NeXTcube at CERN –  
the first Web server ever**

# CERN

the place where **www** was born !

**W**orld **W**ide **W**eb created in 1990 by



**NeXTcube at CERN –  
the first **Web server** ever**



**Sir Tim Berners-Lee  
(2008)**

# CERN

the place where **www** was born !

**W**orld **W**ide **W**eb created in 1990 by

http =

**H**yper

**T**ext

**T**ransfer

**P**rotocol

the first web server

<http://info.cern.ch>



Sir Tim Berners-Lee  
(2008)



# What is LANL good for ?

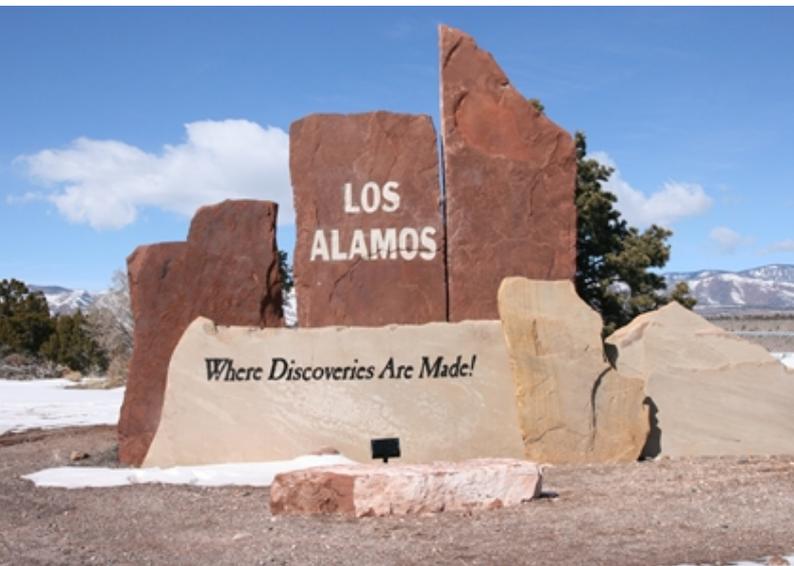


**Los  
Alamos  
National  
Laboratory**



*Los Alamos, New Mexico*

**Not only to produce  
atomic bomb and to work on  
all problems of physics,  
but also...**



# What is LANL good for ?

?

**Los**

**xxx.lanl.gov**

**Alamos**

**National**

**Laboratory**

The logo for arXiv.org, featuring the text "arXiv.org" in white lowercase letters on a solid red rectangular background.

**... to create arXiv (originally lanl preprint world-known repository of electronic preprints, initially devoted to high – energy physics using TeX file format**

# LANL preprint archive

August 14, 1991

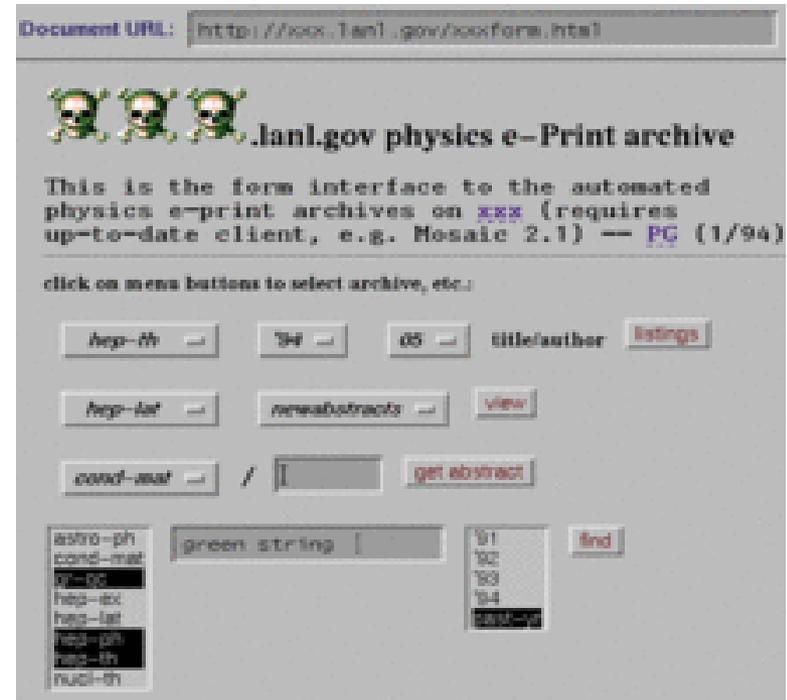
– start of the **lanl.gov**

October 2008:

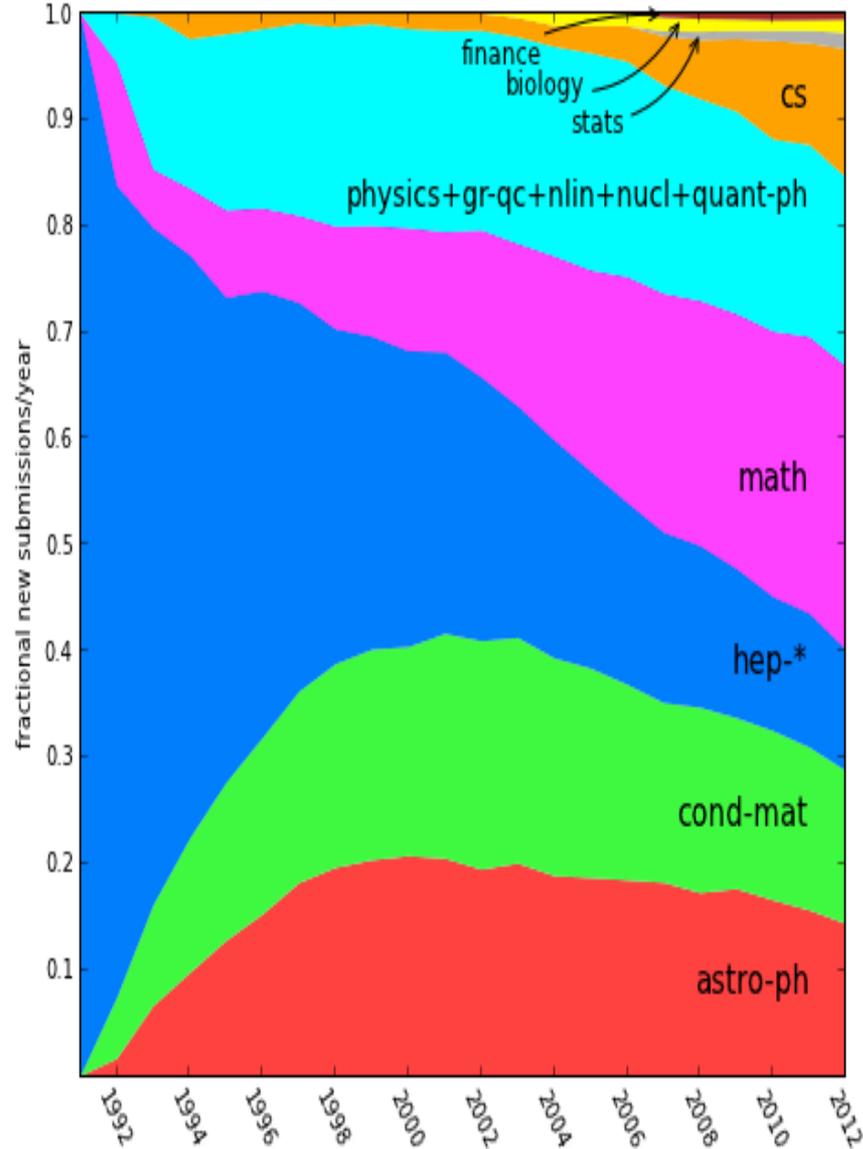
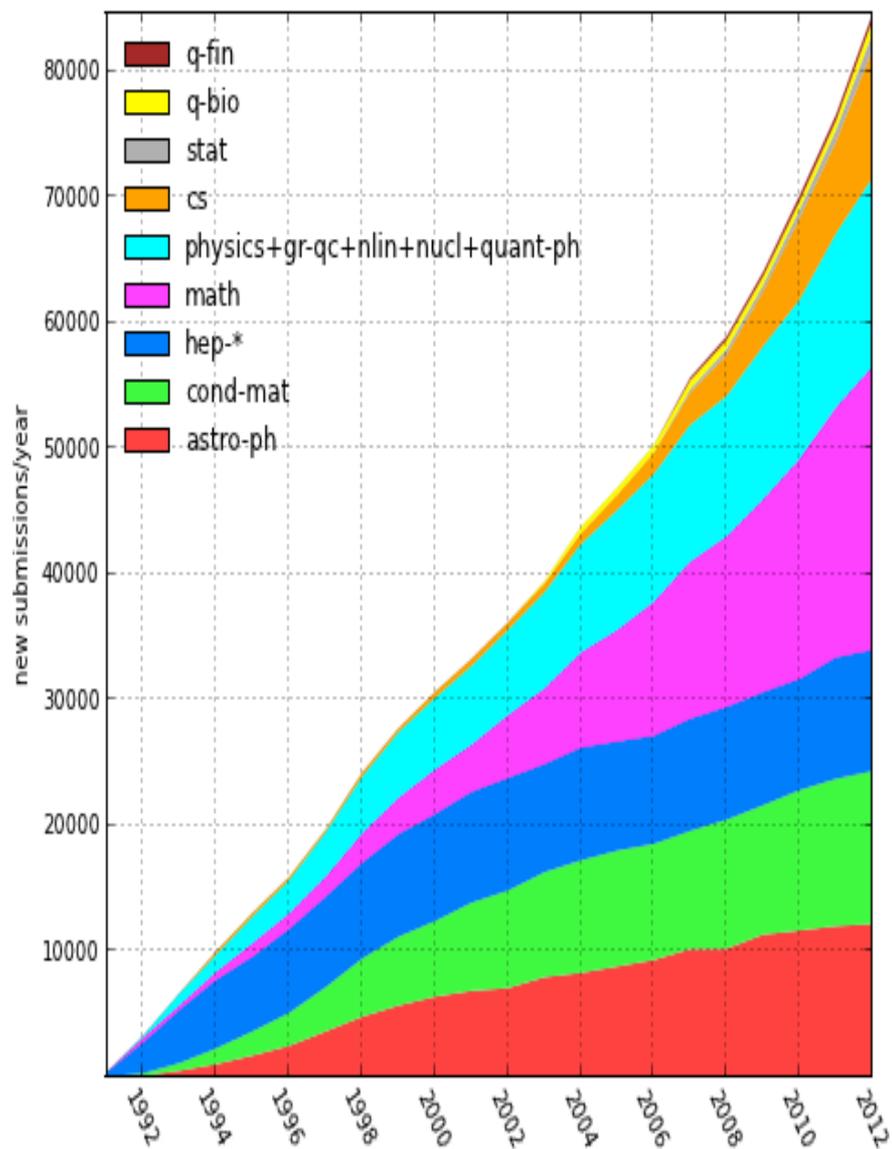
half-million article

December 2014:

million article



**arXiv:** repository of self-archived preprints  
in physics, mathematics, astronomy,  
statistics, computer science,  
also quantitative biology, quantitative finance

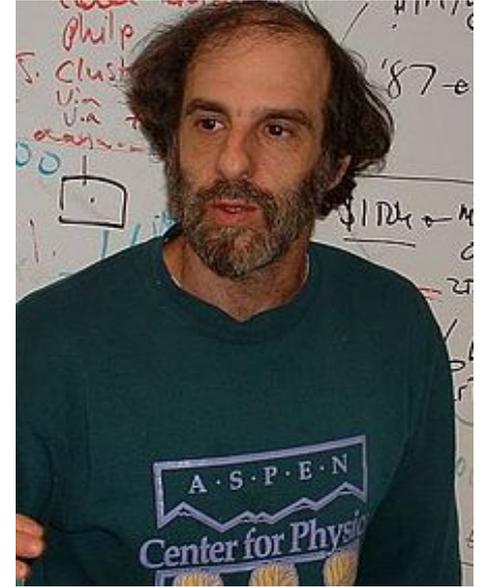


time dependence of number of **e-prints**  
in various **fields**

# arXiv was created at LANL

by (theoretical physicist)

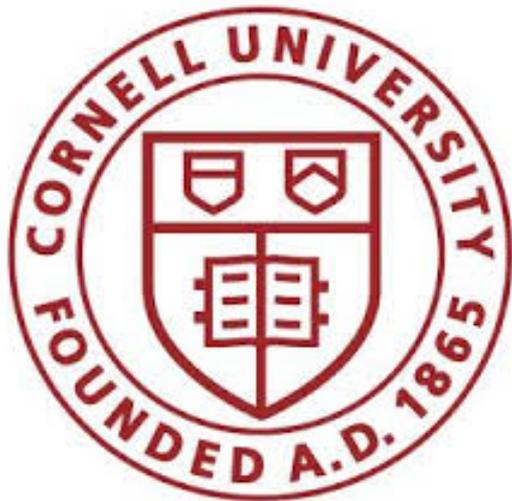
**Paul Ginsparg**



As he moved  
in 2001 to Cornell University

(Ithaca, N. Y.)

(2006)



Cornell University Library  
took overall responsibility for  
**arXiv's** operation, finances  
and administration



**arXiv** server in  
the early 1990s:  
a computer that  
helped to change  
the world of  
physics

**ArXiv at 20**  
Paul Ginsparg *Nature* 476,  
145–147 (2011)

arXiv.org

# Arxiv.org finances : (open)

arXiv.org

**Arxiv: annual budget: 870 k\$**

Cornell University Library	75 k\$
Simons Foundation	100 k\$ +300 k\$
member institutions (191)	300 k\$
ERC	120 k\$

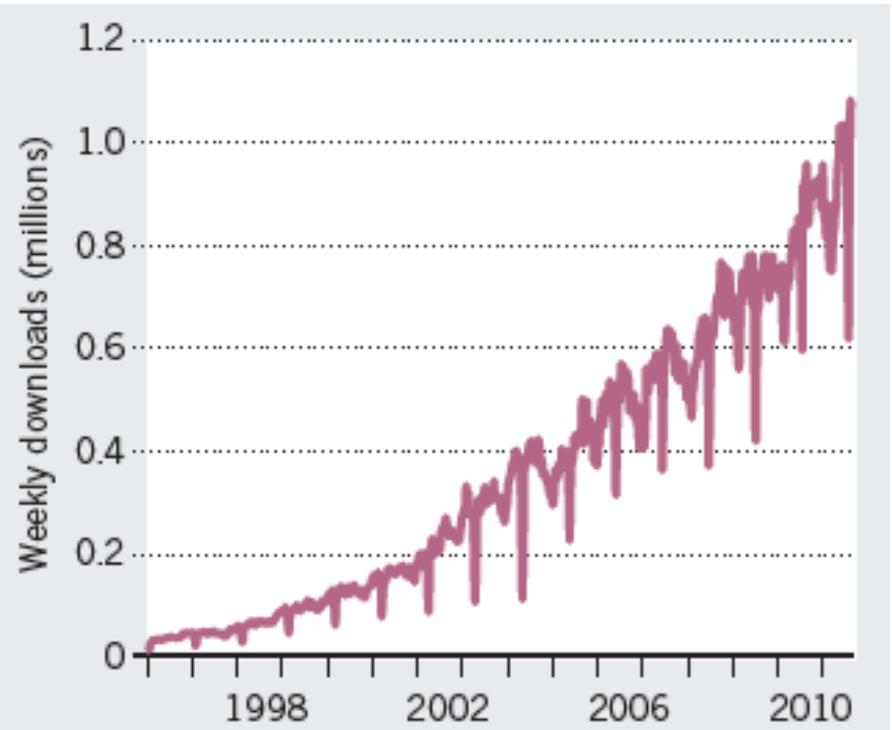
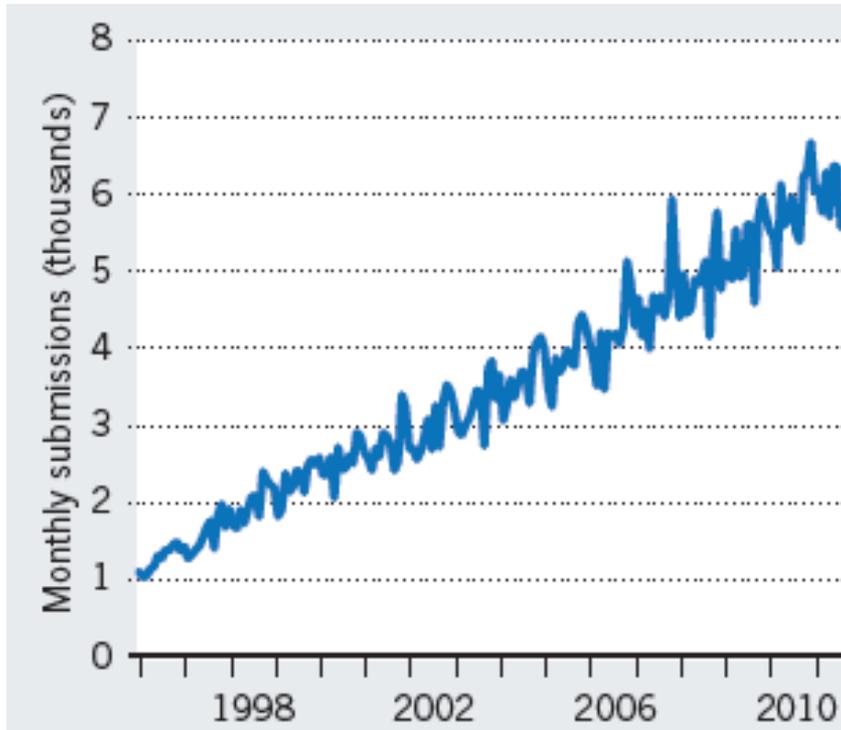
**Employing:**

**6 full-time employees**

**(administration, user support)**

**2 full – time programing experts**

# arXiv statistics 1996 - 2011



**monthly submissions (k)**

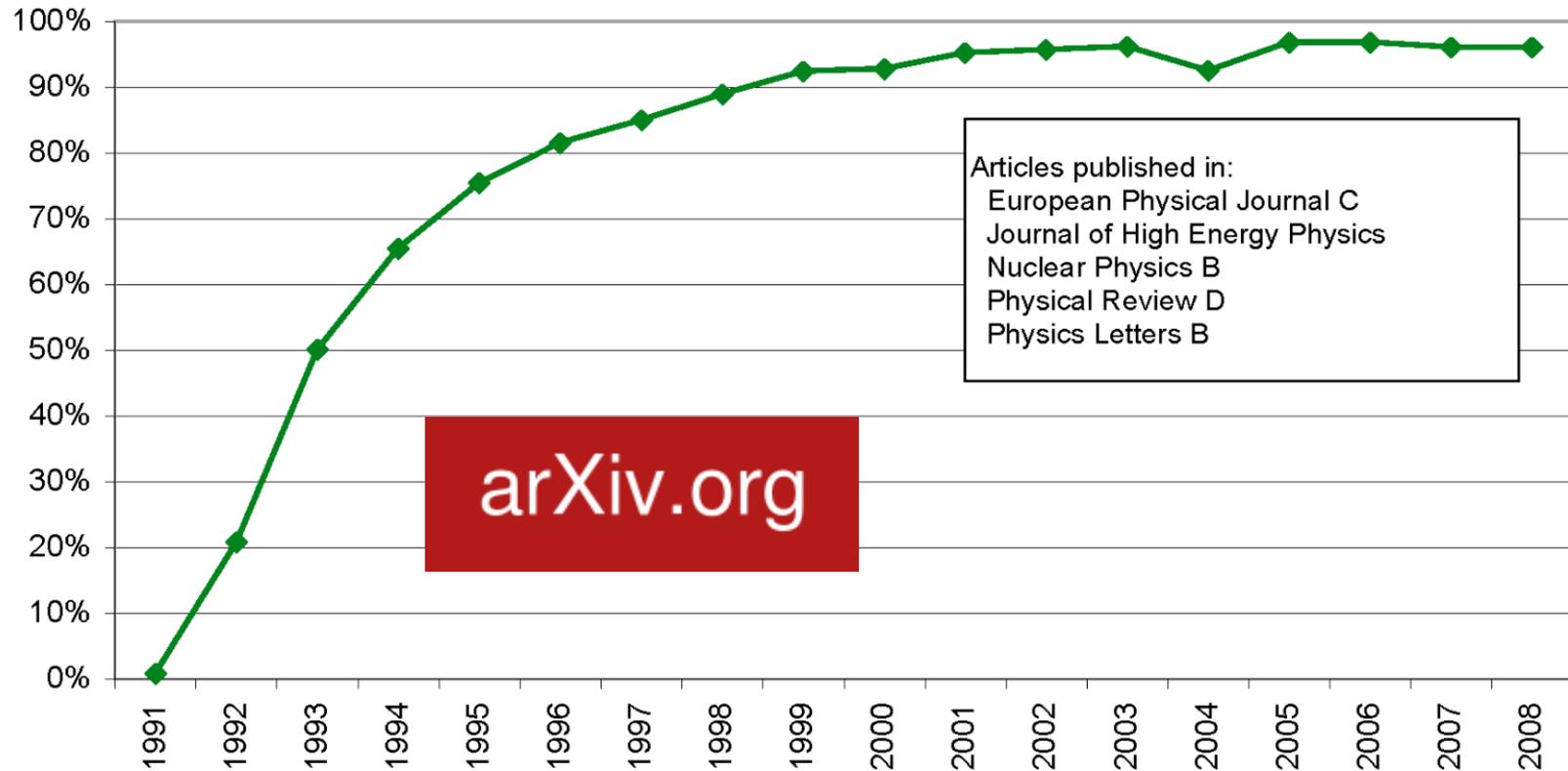
**weekly downloads (M)**

# Reasons to post in the arXive

arXiv.org

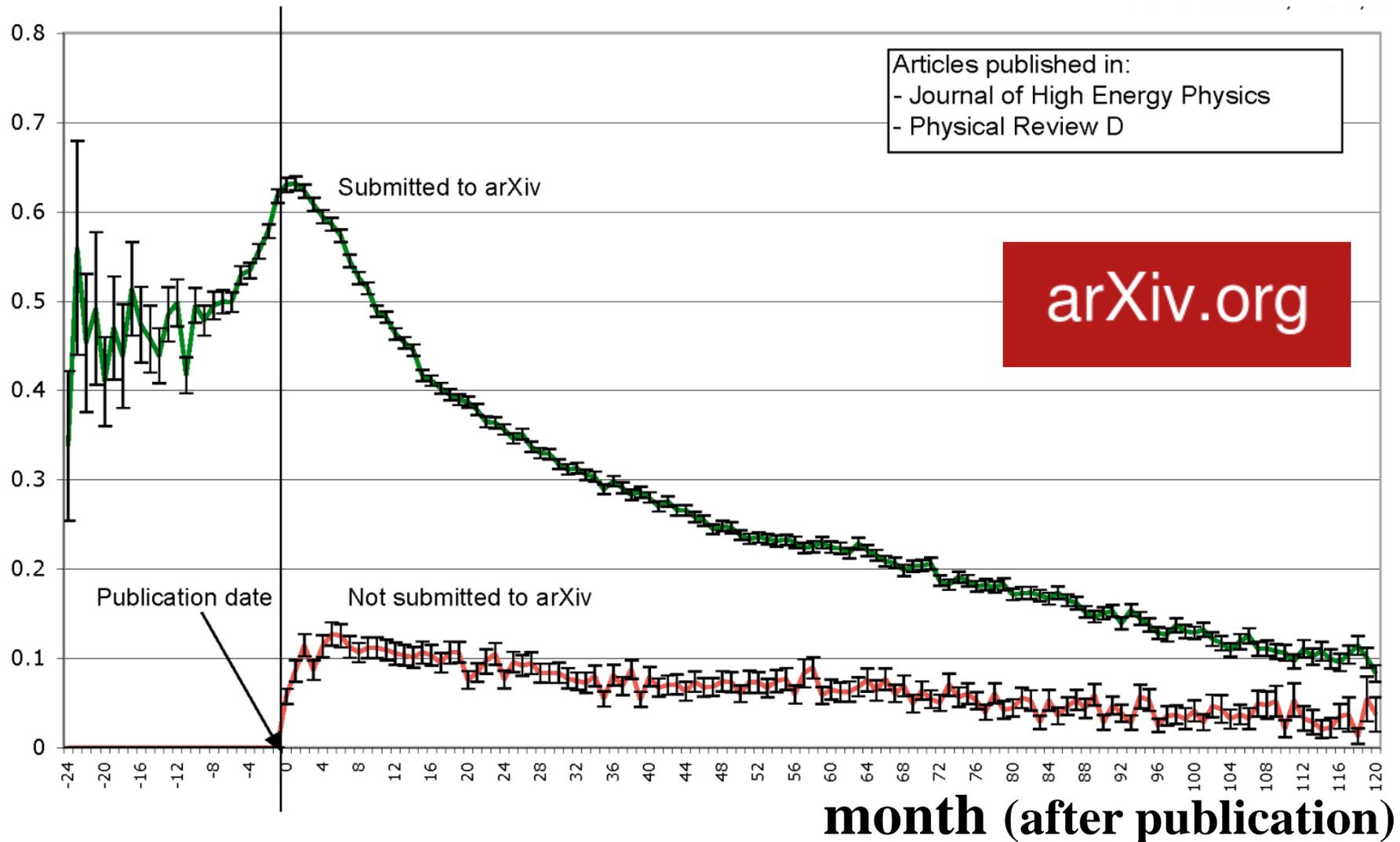
1. Your paper is accessible to everybody
2. Posting attracts the readers
3. Posting establishes priority (**time stamp**).
4. It is conventional. In some fields (as **theoretical physics**) the critical mass is reached and *not* posting to the arXive looks strange

# In some fields arXiv become a standard:



**Gentil-Beccot, Mele, Brooks, [arXiv:0906.5618](https://arxiv.org/abs/0906.5618)**

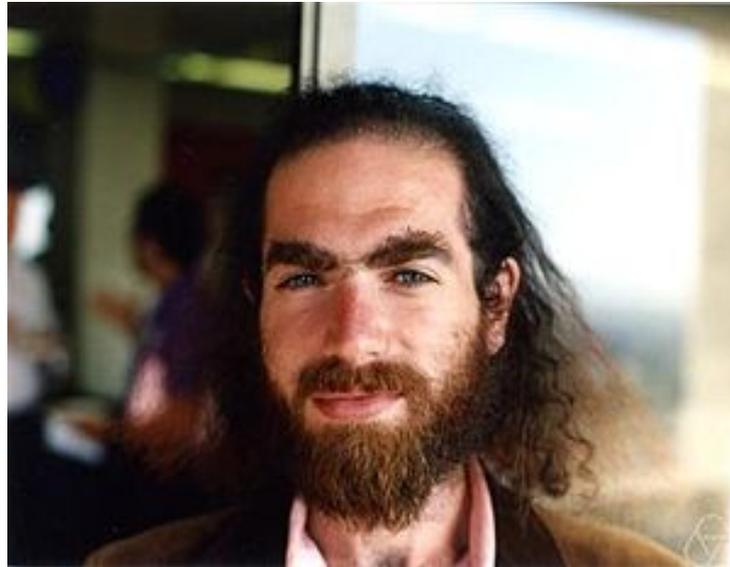
# Posting in arXiv increases citations



Gentil-Beccot, Mele, Brooks, **arXiv:0906.5618**

**Some important papers are posted in  
arXiv and are never published !**

**Example: Grigori Perelman –  
proof of the Poincaré conjecture (Nov. 2002)**



**G. Perelman - Fields Prize (2010), declined**

# rules of arXiv

arXiv.org

- access is free to all users
- researchers deposit their articles free
- an *endorsement* system is used
- moderators review submissions to verify that „they follow accepted standards of scholarly communication”
- what is posted in the arXiv will remain there for good, but new versions of each preprint can be uploaded



Cornell University  
Library

**K. Ż: a personal  
perspective**

**arXiv.org Search Results**

**Showing results 1 through 25 (of 145 total)**

**for au:[zyczkowski](#)**

1. arXiv: **160601991** (June 2016)

*Operational approach to Bell inequalities: applications to qutrits*

...

145. arXiv:chao-dyn/**9501019** (**January 1995**)

*Parametric Spectral Correlations of Disordered Systems  
in the Fourier Domain*

# other archives and repositories



**standards in various fields do differ!**

# E-prints in the archives and publishing in journals

**Different standards in various fields !**

**some examples:**

**Physics (APS, AIP)**

**Mathematics (AMS)**

**compatible !**

**(AACR) Am. Assoc.**

**Cancer Research**

**(JCI) J. Clinical**

**Investigation**

**incompatible !**

# other arXiv and repositories

**viXra.org**

*(created by people who disagree  
with arXiv endorsement policy)*

**open to anybody –**

**including amateur scientists**

**viXra**

„highlights Cornell University's  
unacceptable censorship policy”

# snarXiv - a random high-energy theory paper generator

(David Simmons-Duffin, Princeton)

basing on the title it is difficult to distinguish a scientific paper from a fake one:

what is the science about??

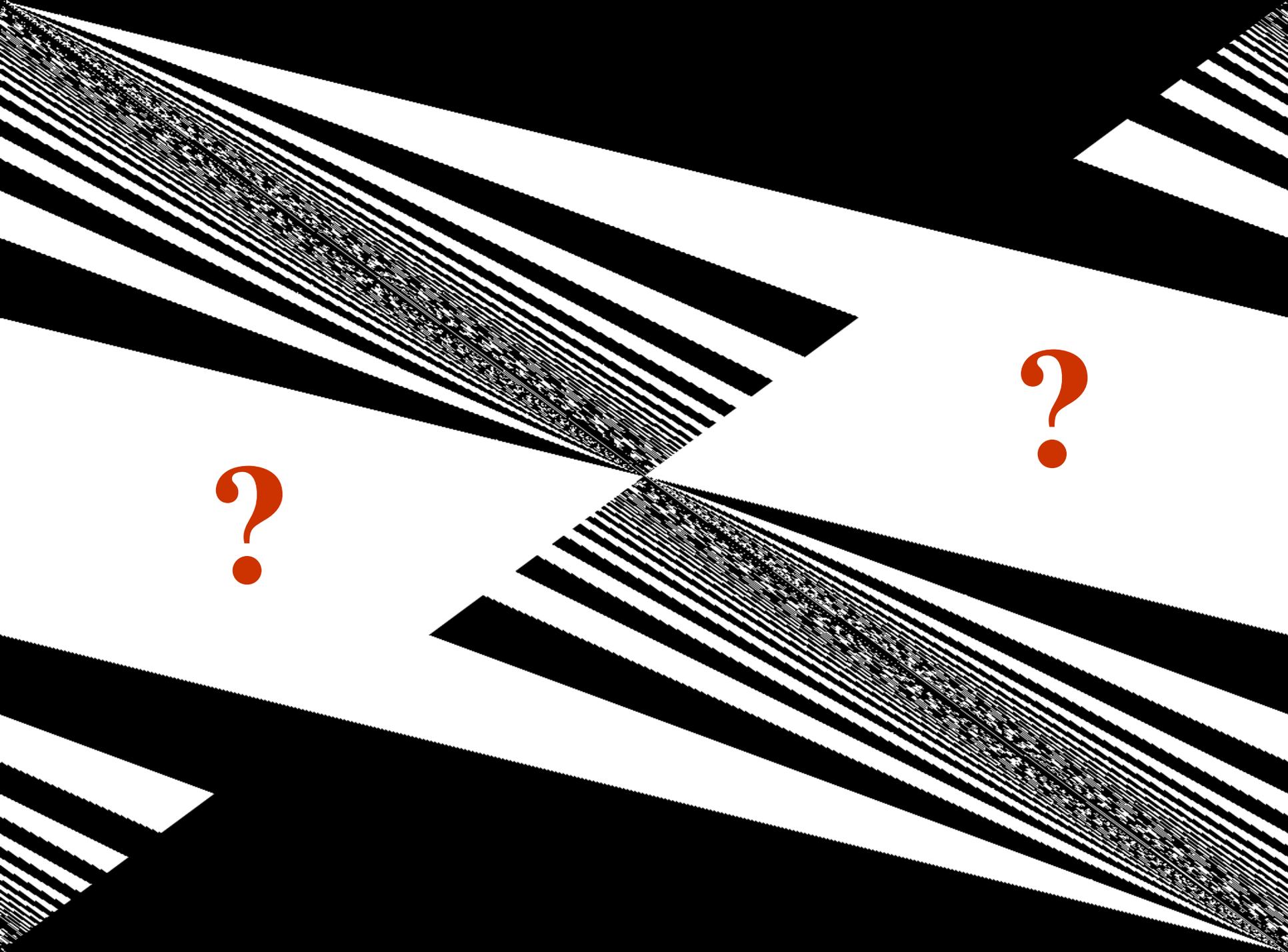


Gerbs in Topological String Theory  
and a  $G_2$  Singularity

The Reaction  $S\text{Pi} N O \text{Pi} \text{Pi} N\text{S}$  Above  
Threshold in Chiral Perturbation  
Theory

# How to distinguish

- 1) A sound scientific paper from a fake one ?**  
(e.g. produced by a computer!)
- 2) A sound scientific journal from a „predatory” open access journals ?**
- 3) Quality scientific conference from a meeting organized for profit only ?**



# How to distinguish

- 1) A sound scientific paper from a fake one ?  
(e.g. produced by a computer!)  
**in some cases only expert can help**
- 2) A sound scientific journal from a  
„predatory” open access journals ?  
**list of Jeffrey Beal**  
**of predatory publishers and journals**
- 3) Quality scientific conference from a  
meeting organized for profit only ?

# Critical analysis of scholarly open-access publishing



**Jeffrey Beal**  
**UoC (Denver)**

- Beal's black lists of**
- a) Predatory publishers**
  - b) Predatory journals, which**
    - **have no individual editor named,**
    - **concocted editorial boards,**
    - **artificial name e.g.**
      - „Swiss Journal of ....“**
    - **use spam e-mails to look for potential authors (= victims)**
    - **display fake impact factor numbers**
    - **...**

# Beal's lists of **predatory publishers**

- **Academia Publishing**
- **Academia Research**
- **Academia Scholarly Journal**
- **Academic Journals, Inc.**
- **American Scientific Publishers**
- **American Scientific Research Journals**
- **Canadian Science and Technology Press**
- **Cardiology Academic Press**
- **Open Access Journals**
- **Open Access Library**
- **Open Access Publishing Group**

*and many other*

# Beal's lists of **predatory journals**

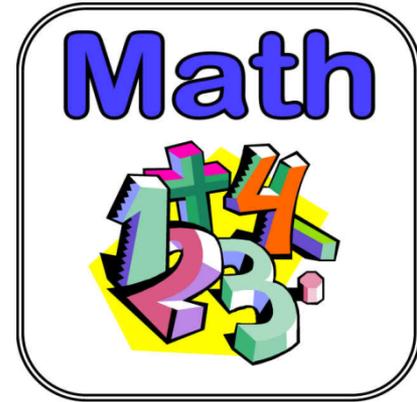
- **Academic Exchange Quarterly**
- **Academic Research Reviews**
- **Acta Kinesiologica**
- **Acta Medica International**
- **Actual Problems of Economics**
- **International Journal of Computer Science and Network (IJCSN)**
- **Progress in Physics**
- **Quarterly Physics Review**

*and many other ...*



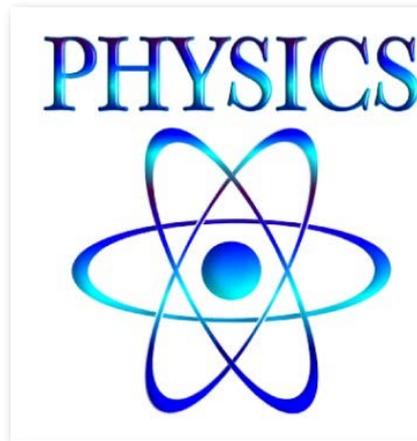
# Is Impact factor a universal solution ?

Standards in various fields differ!



## Mathematicians:

- publish **less** papers a year
- quote **less** references in each article
- the average number of the authors of a paper is **smaller** in comparison with **physicists** !



## Conclusion:

One **should not** compare bare data from **different fields** !  
(but rather use data **rescaled** to the average...)

# Comparison of different fields I

## (Data: USA 2006 - 2011)

Field	Articles per year	Scientists (USA)	Authors per article	Articles per year per author	References per article	Citations per year per scientist	Citations compared to Math.
<b>Mathematics</b>	4190	37000	2.0	0.2	15.0	3.5	1
<b>Physics</b>	18227	49000	5.3	2.0	27.8	54.8	15.9
<b>Chemistry</b>	16430	86000	4.3	0.8	31.7	26.0	7.5
<b>Computer Sci.</b>	2188	20000	3.0	0.3	17.4	5.8	1.7
<b>Engineering</b>	14609	144000	3.8	0.4	18.7	7.3	2.1
<b>Space Science</b>	3187	5000	5.9	3.8	41.2	154.9	44.9
<b>Geosciences</b>	11621	21000	4.0	2.2	33.1	73.0	21.2
<b>Agriculture</b>	3469	22000	4.3	0.7	20.8	14.2	4.1
<b>Biological Sci.</b>	49614	193000	5.3	1.4	41.1	55.8	16.2
<b>Medical Science</b>	58664	45000	5.6	7.3	43.2	315.3	91.4
<b>Psychology</b>	9805	114000	3.2	0.3	46.5	13.0	3.8
<b>Social Science</b>	12020	100000	1.9	0.2	26.8	6.2	1.8

# Comparison of different fields II

## Citations of 100-th highly cited scientist

*(Web of Science, ESI 2002-20012)*

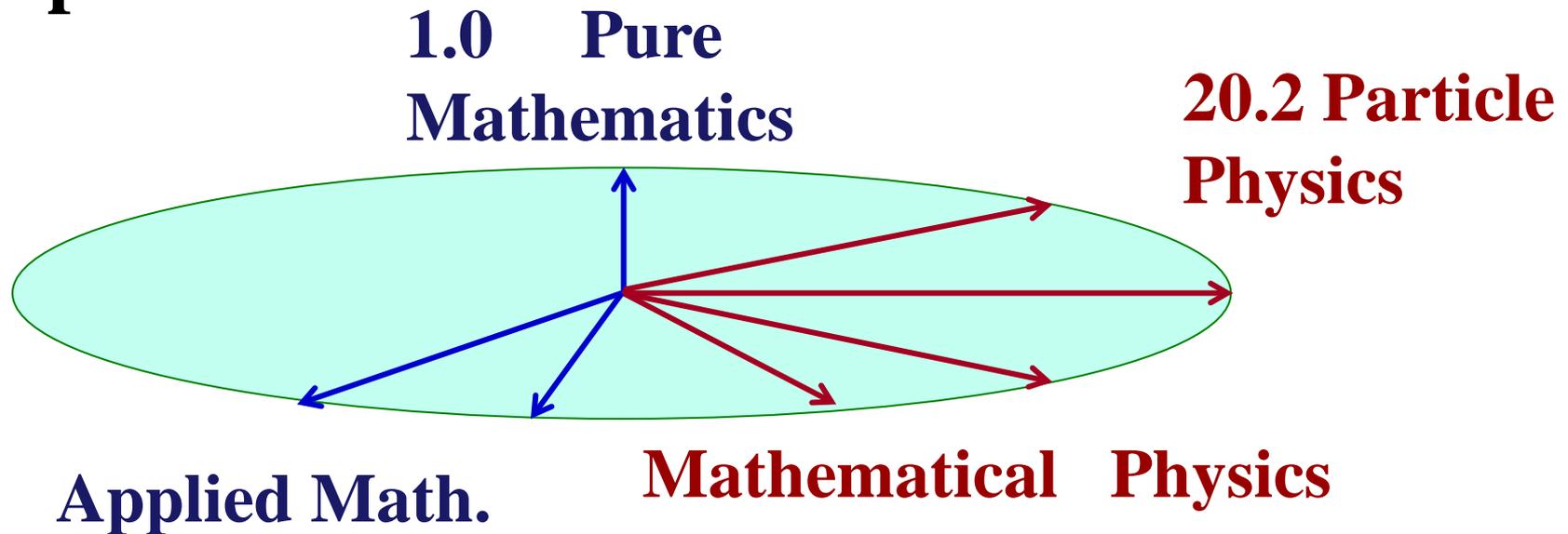
Field	Citations of 100-th scientist	Relative to Mathem.
<b>Mathematics</b>	733	<b>1.0</b>
<b>Physics</b>	14772	<b>20.2</b>
<b>Chemistry</b>	12420	<b>16.9</b>
<b>Computer Sci.</b>	1247	<b>1.7</b>
<b>Engineering</b>	3165	<b>4.3</b>
<b>Space Science</b>	9700	<b>13.2</b>
<b>Geosciences</b>	3571	<b>4.9</b>
<b>Agricultural Sci.</b>	1288	<b>1.7</b>
<b>Biology &amp; Bioch.</b>	6092	<b>8.3</b>
<b>Psychiat./Psycho.</b>	3256	<b>4.4</b>
<b>Social Sciences</b>	948	<b>1.3</b>

Field	Citations of 100-th scientist	Relative to Mathem.
<b>Clinical</b>		
<b>Medicine</b>	17051	<b>23.3</b>
<b>Economics</b>	960	<b>1.3</b>
<b>Ecology</b>	3013	<b>4.1</b>
<b>Immunology</b>	4169	<b>5.7</b>
<b>Materials Sci.</b>	4750	<b>6.5</b>
<b>Microbiology</b>	3244	<b>4.4</b>
<b>Molecular Biol.</b>	9021	<b>12.3</b>
<b>Neuroscience</b>	5781	<b>7.9</b>
<b>Pharmacology</b>	2081	<b>2.8</b>
<b>Plant &amp; Animal</b>	3114	<b>4.3</b>

The above numbers are only approximate and **should not** be treated as **conversion coefficients!**

The reason: differences inside a given field

Example:



Entire spectrum of intermediate cases exists...

# Statistics of citations in scientific papers

## *What are they good for?*

- some people (librarians, publishers, editors) do care about **bibliometric indices**, e.g.

**impact factor (IF-2)**, **E. Garfield 1972**

$$IF(i) = c_i / (Z_{i-1} + Z_{i-2})$$

$c_i$  = the number of **citations** gained in year  $i$  by **articles** published in  $i-1$  and  $i-2$ .

$Z_i$  = the number of **articles** published in  $i$ .

- can we afford the **luxury** to neglect it ??

**perhaps not...**

**Mean Impact factor  $\langle IF \rangle$ , mean number of citations  $\langle c \rangle$   
 probability  $\langle p \rangle$  that a paper will be quoted in 2-years  
 data from *Journal of Citation Reports* for 1994-2005**

<b>Field</b>	<b>Size</b>	<b><math>\langle IF \rangle</math></b>	<b><math>\langle c \rangle</math></b>	<b><math>\langle p \rangle</math>[%]</b>
<b>Biology</b>	<b>511</b>	<b>4.76</b>	<b>45.8</b>	<b>20.5</b>
<b>Astronomy</b>	<b>25</b>	<b>4.29</b>	<b>38.3</b>	<b>21.5</b>
<b>Medicine</b>	<b>766</b>	<b>2.89</b>	<b>33.9</b>	<b>18.3</b>
<b>Chemistry</b>	<b>145</b>	<b>2.61</b>	<b>33.1</b>	<b>17.0</b>
<b>Physics</b>	<b>503</b>	<b>1.91</b>	<b>24.0</b>	<b>16.7</b>
<b>Economics</b>	<b>159</b>	<b>0.82</b>	<b>30.4</b>	<b>12.1</b>
<b>Comp. Scien.</b>	<b>124</b>	<b>0.63</b>	<b>17.2</b>	<b>19.3</b>
<b>Mathematics</b>	<b>149</b>	<b>0.56</b>	<b>18.4</b>	<b>8.5</b>
<b>History</b>	<b>23</b>	<b>0.41</b>	<b>81.8</b>	<b>10.1</b>

**Althouse, West, Bergstrom, 2008**

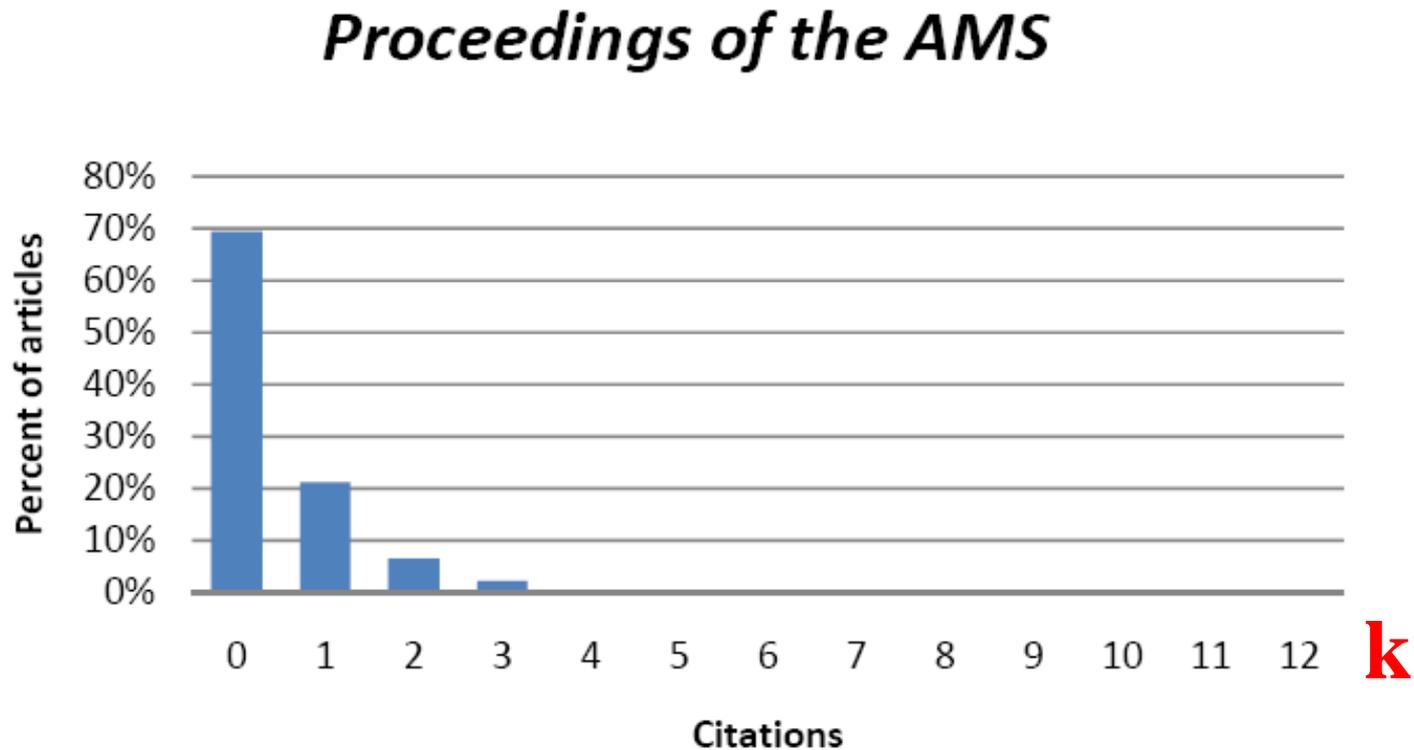
# Traps and pitfalls related to **citation statistics**

## *From the committee charge*

The drive towards more transparency and accountability in the academic world has created a "culture of numbers" in which institutions and individuals believe that fair decisions can be reached by algorithmic evaluation of some statistical data; unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure. This trend calls for comment from those who professionally "deal with numbers" — mathematicians and statisticians.

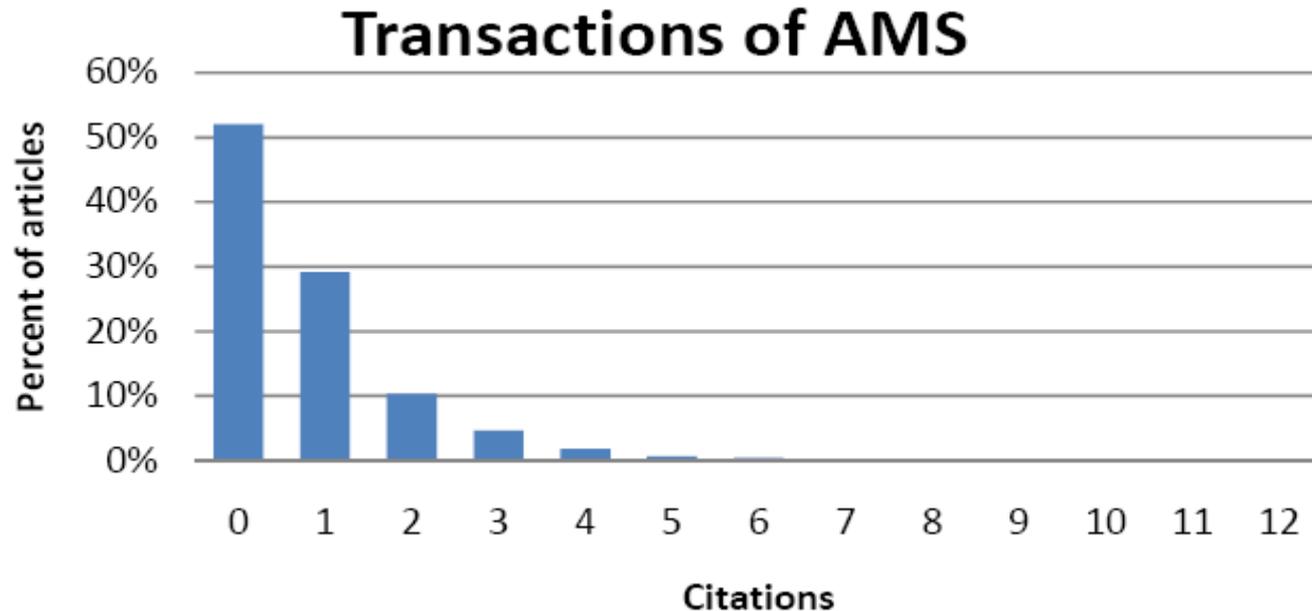
**R. Adler, J. Ewing, P. Taylor**, Report of *Joint Committee on Quantitative Assessment of Research*, November 2008

Distribution of citations for a given paper is given by a **power** law,  $P(k) = a k^{-b}$ , (Garfield 1987)



**Example a):** 2005 data for *Proceedings of AMS* imply impact factor **IF= 0.43**

**Example b):** 2005 data for *Transactions of AMS* imply impact factor **IF= 0.85**

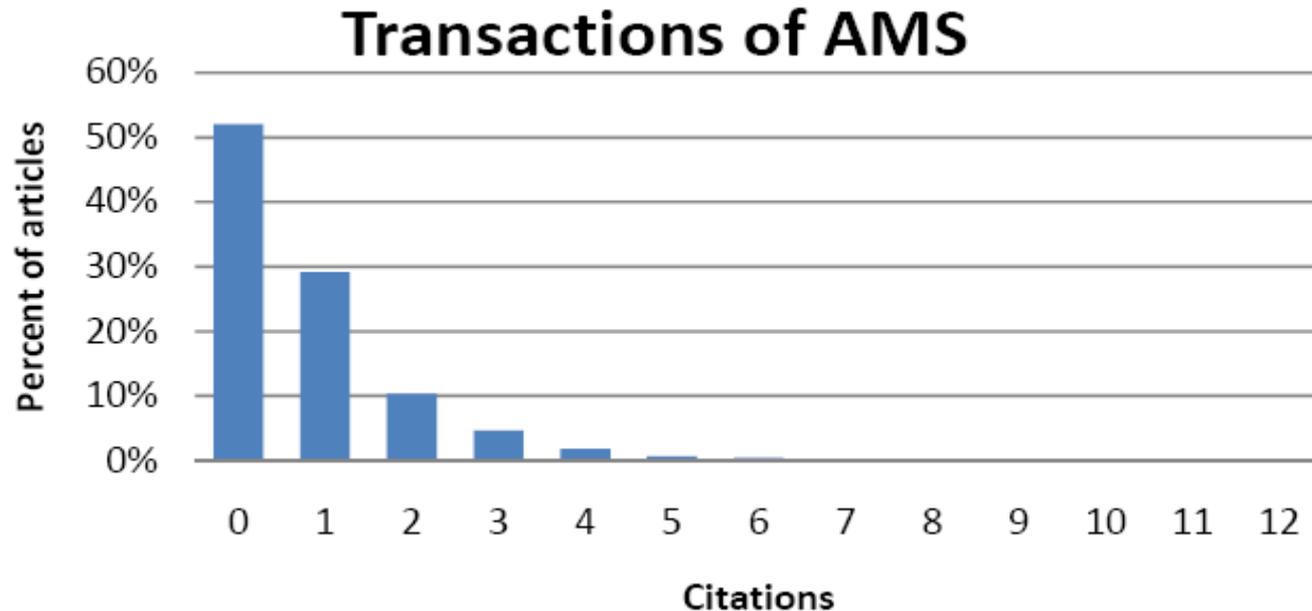


**Question:** Is an average *Transaction* paper **twice as good** as the one published in *Proceedings* ?

- What is the **probability** that a **random** *Proceedings* paper as at least as many citations as the one from *Transactions*?

• ?

**Example b):** 2005 data for *Transactions of AMS* imply impact factor **IF= 0.85**



**Question:** Is an average *Transaction* paper **twice as good** as the one published in *Proceedings* ?

- What is the **probability** that a **random** *Proceedings* paper as at least as many citations as the one from *Transactions*?
- the answer is: **62%**

We are more often **wrong** than **right** !

While it is incorrect to say that the impact factor gives no information about individual papers in a journal, the information is surprisingly vague and can be dramatically misleading.

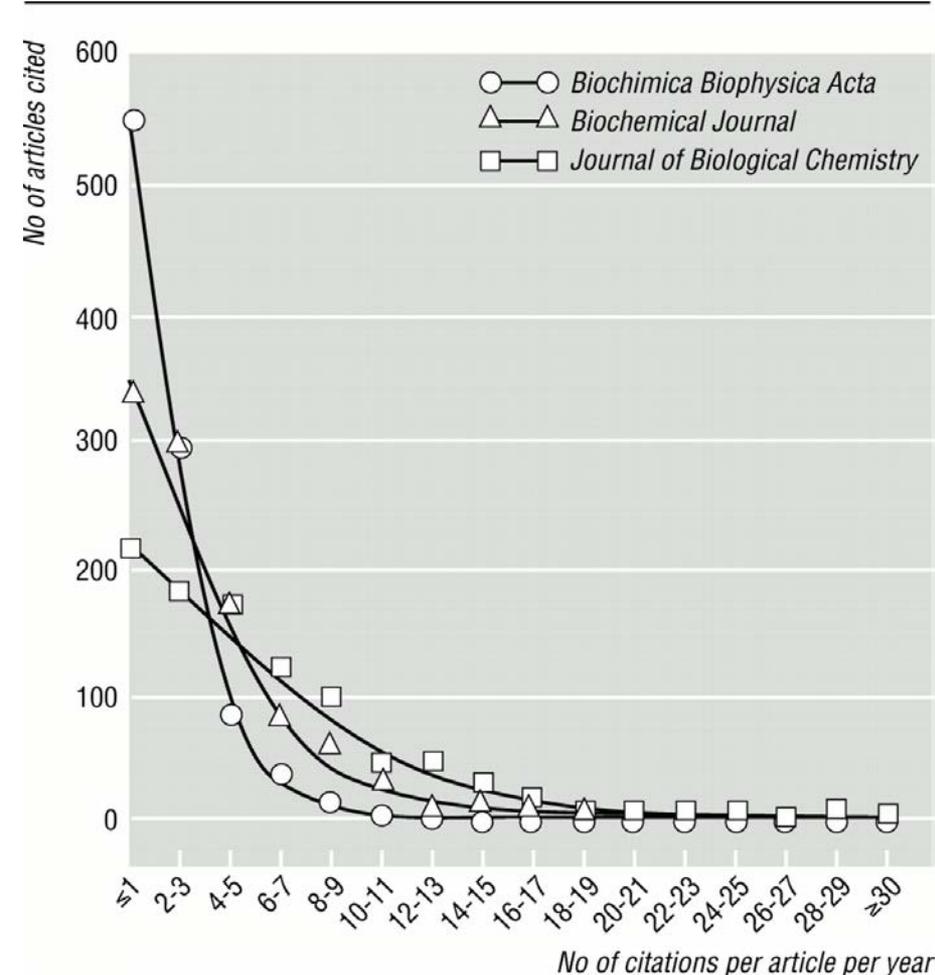
- The **2008** report of **Adler, Ewing, Taylor** criticise sole reliance on the **impact factor**, since the ‘objectivity’ of such numbers can be **illusory**.

They compare the judgments of a **journal** based on its **Impact Factor** alone to using the **weight** of a patient to judge his **health...**

- See also: **E. Falgas and V. Alexiou:**  
*The top-ten in journal impact factor manipulations*  
**Arch. Immunol. Theor. Exp. 2008**

# Why *Impact Factor* of journals should not be used for evaluating research

P. O. Seglen, *British Medical J.* 1997



**IF does not reflect the skewness of the distribution (50% of citations are gained by 10% papers).**

**IF depends on the field.**

**Article citations determine the journal IF, but not vice versa!**

**IF depends on the mean number of references in each article.**

**Adding Impact Factor** of journals in which all papers of a given author were published is **a capital crime** against the rules of bibliometry !! *(would you sum e.g. the \$\$\$ price of each issue?)*

- If one considers **citations** as a **quality** indicator, one should rather care about the average (**true**)

**impact factor of his papers**

(older than 3 years) **defined** as the **sum of citations** his articles published in year  $i$  gained in years  $i+1$  and  $i+2$  .

= *his direct contribution to IF of the journals !*

# Role of the bibliometric approach

- i) **Bibliometric data** have **statistical** character, so they can be used to compare **scientific output** in a single country or progress in a given branch of science.
- ii) **Bibliometric data** data are less useful to evaluate research record of an **individual scientist** and they allow for a reasoning in **one direction** only:
  - an author with a **small** number of **papers and citations** is unlikely to be an influential scientist,
  - a **huge** number of **papers and citations** does not prove that the author is a **prominent scientist** =>

**peer review is required !**

# Concluding remarks

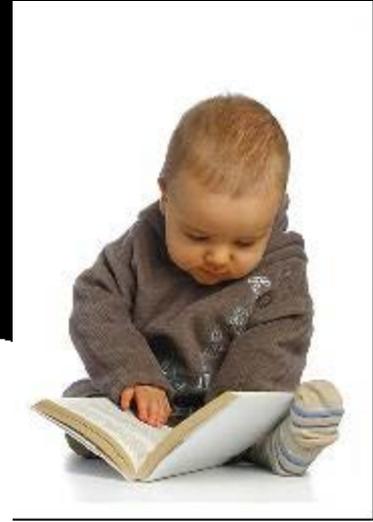
- 1) Preprint depositories (**arXiv**) totally changed publication process in some fields (physics, math) as they open scientific literature to everybody
- 2) Open Access journals play their role in science, but their fate is jeopardized by **predatory** journals, which are difficult to identify by a layman
- 3) **Impact Factor** (and other indicators) helps to distinguish a descent journal, but it should be **ruled out** from process of **evaluation** of scientific record of **individual scientists**
- 4) Comparing two scientists from different fields one **cannot use** bare bibliometric data.

# The very last final remark:

If necessary,

do use **bibliometric data,**

but do it in a **reasonable way!**



**Thank You !**



Letnim wieczorem 1916 roku dwaj młodzi krakowianie,

Stefan Banach i Otton Nikodym,

na ławce na plantach rozmawiali o matematyce.

Do dyskusji włączył się przechodzący obok matematyk,

dr Hugo Steinhaus.

Tak został odkryty niezwykle matematyczny talent Stefana Banacha,

Jednego z najwybitniejszych polskich uczonych.

Otton Nikodym

Stefan Banach

in conversation about mathematics.

This bench <sup>was</sup> their famous meeting with Hugo Steinhaus in the

Planty Garden in summer 1916.

# Stefan Banach and Otton Nikodym



Opening of the *Mathematicians' Bench*  
Planty Garden (Straszewskiego),  
Friday, October 14, at 12.00