Why a Master Class on Open Science ?

Gabriel Chardin

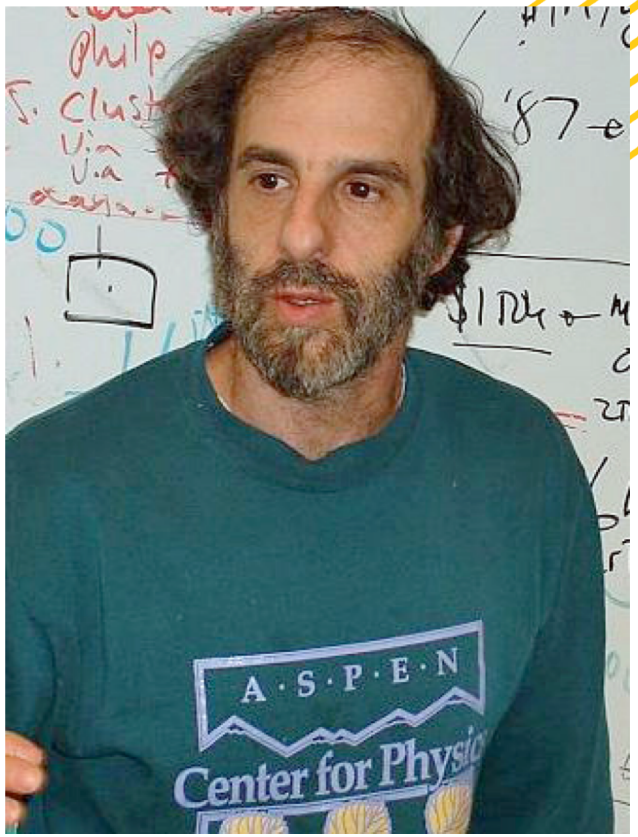
SFP secretary

Société Française de Physique

150 ANS D'ENGAGEMENT POUR LA PHYSIQUE

arXiv et al. (1990-…)

- Paul Ginsparg: develops in 1990 on a single Next workstation what soon becomed arXiv and transformed fundamentally the way we present our results to the community
- Abstracts and preprints of a whole field
- Some selection by arXiv to avoid too crazy papers, but relatively minor censorship
- Not recognized in LANL, Paul left Los Alamos in 2001 for Cornell



arXiv et al. (1990-…)

Next-Generation Implications of Open Access

Paul Ginsparg

CIS and Physics, Cornell University

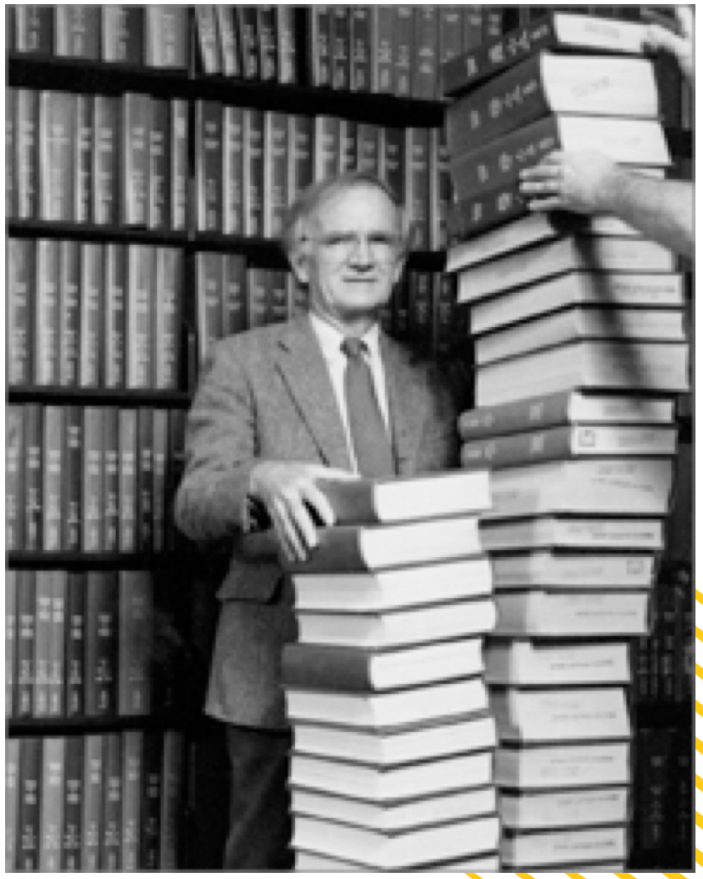
Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Open data policy (for the US tax payer…)

- 2012 : Jim Whitmore, NSF Scientific director for astroparticle and neutrino physics : request for open data, open publications (make access free for the US tax payer…)
- But no additional funding to implement these open data and open publication policies

Attempting to follow the flow of information

- Val Fitch noted the exponential progression of the total amount of publications
- Several new publishing actors, and possibly some predators…
- MDPI publications are now probably of a better and more controlled standard, but…



Impact on various fields

- In astrophysics, publication impact more than a factor two increase after open data release
- Allows to check results, data analysis more reproducible, more efficient use of data
- Particularly important in biology, where more than 50% of data funded by private biofirms (and at some epoch, >80% of those studies funded by Merck…) are non reproducible (although the situation has somewhat improved since this key problem has been noted)

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true.

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on

- Let me thank the organizers and in particular Bart van Tiggelen, Noemi Cobolet, Greg Henning and all the organizers for inviting me to this Master Class

Merci

150 ans

150 ANS D'ENGAGEMENT
POUR LA PHYSIQUE