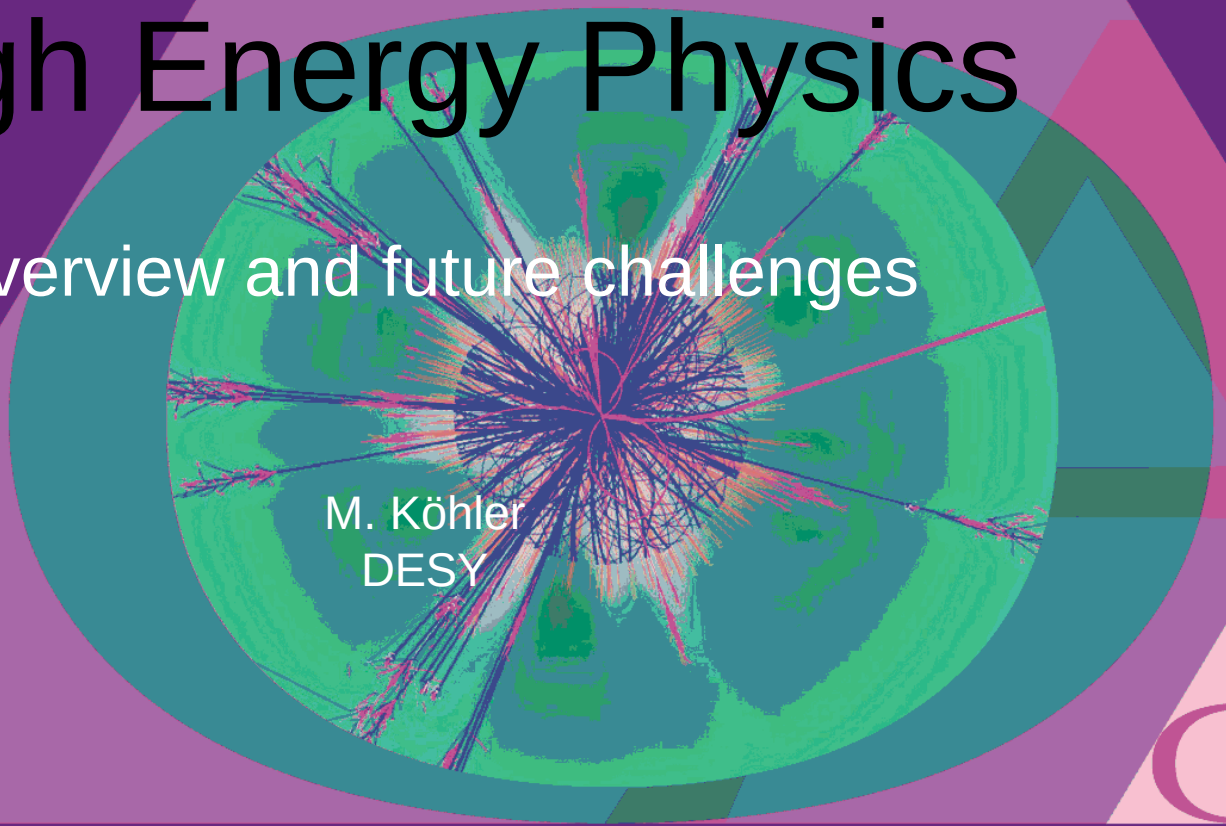
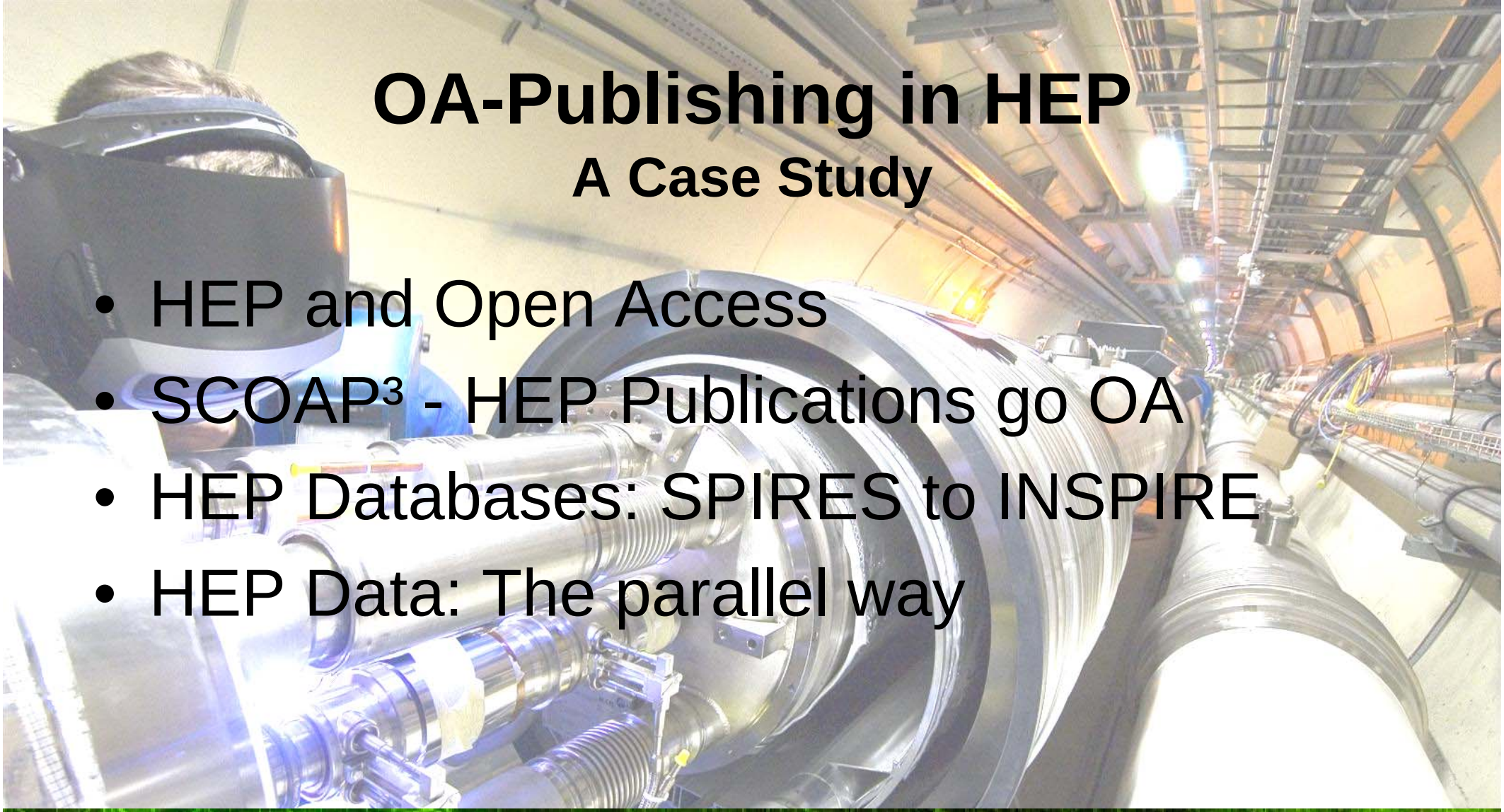
Open Access in High Energy Physics

Overview and future challenges



M. Köhler
DESY



OA-Publishing in HEP

A Case Study

- HEP and Open Access
- SCOAP³ - HEP Publications go OA
- HEP Databases: SPIRES to INSPIRE
- HEP Data: The parallel way

HEP and Open Access: a synergy

- **HEP is used in thinking Open Access:**
  - **Paper preprints shipped worldwide by HEP institutes for 40 years (at institute expenses!)**
  - **HEP launched arXiv (1991), the archetypal Open Archive**
  - **The first free peer-reviewed electronic HEP journals:**
    - ***Journal of High Energy Physics* (1997)**
    - ***Physical Review Special Topics Accelerators and Beams* (1998)**
- **Small and connected community (≈20'000 scientists)**
- **Small scientific output (≈ 8'000 articles/year)**
- **Small publishing landscape (≈ 10 main journals)**
- **Reader and author communities largely overlap**
- **Open Access, second nature: posting on arXiv before even submitting to a journal is common practice.**
  - **No mandate, no debate. Author-driven. Evident benefits**
  - **Revised version post peer-review routinely uploaded**

Is it all about vocal librarians?
Strong support form the LHC collaborations

*"We, the _*_ Collaboration, strongly encourage the usage of electronic publishing methods for _*_ publications and support the principles of Open Access Publishing, which includes granting free access of our _*_ publications to all. Furthermore, we encourage all _*_ members to publish papers in easily accessible journals, following the principles of the Open Access Paradigm."*

5400 scientists
building the largest
scientific instruments ever

_*_ {
ATLAS; approved on 23rd February 2007
CMS; approved on 2nd March 2007
ALICE; approved on 9th March 2007
LHCb; approved on 12th March 2007

What is Open Access?

- Definition of Open Access (OA) research literature
- Legal basis of OA
- Motivation *or* What do we all get out of it ?
- Roads towards OA
- HEP and it's journals

Definition

- **“*Open Access* (OA) is free online availability of digital contents” (Wikipedia)**
- ***OA literature* is digital, free of charge, immediate and permanent online access to the full text of research articles for anyone, webwide made possible by consent of author and copyright holder**
- **OA is compatible with**
  - **peer review**
  - **publishing in non-OA journals**
- **OA is not free to produce**
  - **There is much debate about the economics of funding**
  - **Business models depend on the way OA is delivered (see below)**
- **International statements on OA**
  - **Budapest Open Archive Initiative (2002)**
  - **Berlin Declaration (2003) (signed e.g. by Helmholtz Gemeinschaft)**

Roads Towards OA

- **„Golden“ Road: Open access publishing**

  OA Journals make their articles available immediately on publication

  (see e.g. http://www.doaj.org/ for a list of ≈ 2500 OA journals)

  - **Processing fees (Page charge, author charge, …)**
  - **Sponsoring**
  - **…**

- **„Green“ Road: Open access self-archiving**

  Authors make copies of their own (published) work

  - **pre-print (before peer-review)**
  - **post-print (after peer-review with revisions; content as published)**
  - **publishers version (the version, that is published in a journal, incl. layout etc.)**

  accessible on an institutional or subject repository

  (see the SHERPA/RoMEO list at http://www.sherpa.ac.uk/romeo.php for publisher copyright policies & self archiving)

HEP and its journals

- Journals (with their vaguely anachronistic page and figure limits) are on the way to lose (lost?) a century-old role as vehicle of scholarly communication.
- Still, evaluation of institutes and (young) researchers is based on high-quality peer-reviewed journals.
- The main role of journals is to assure high-quality peer-review and act as keepers-of-the-records
- The HEP community needs high-quality journals as their "interface with officialdom"
- Implicitly, the HEP community supports this role by purchasing subscriptions, as it reads off arXiv anyhow
- As an "all-arXiv discipline" HEP is at high risk to see its journal canceled by large multidisciplinary university libraries (when not already happened)

OA Experimente

- **Sponsoring model** ⇨Institutions pay

  No author charges, content OA

- **Hybrid model** ⇨ pay to make individual articles OA
  - Access to other articles remains restricted
  - Subscription costs *might* be reduced
  - Price 975$ - 3.000$ - Little success
  - More expensive in general

- **Author charges** ⇨ Whole journal OA

  Publishing fee, due for accepted articles
  - Successfull in life science,
    long term sustainability ?

- **Institutional membership** ⇨ for small additional fee: All articles OA for authors at institutes
  - SLAC, Fermilab, DESY, CERN and France

Springer Open Choice

APS physics

SPONSORED ARTICLE

DEUTSCHE PHYSIKALISCHE GESELLSCHAFT | IOP Institute of Physics
New Journal of Physics
The open-access journal for physics

PHYSICAL REVIEW LETTERS

PMC Physics A

Journal of High Energy Physics
A refereed journal, written, run and distributed by electronic means
About JHEP | for Readers | for Authors | for Referees | for Editors

Journal of Instrumentation
an IOP and SISSA journal
About JINST | for Readers | for Authors | for Referees | for Editors

The SCOAP^3 model

Sponsoring Consortium for Open Access Publishing in Particle Physics

A practical approach: How to publish OA about 8’000 articles/year, produced by a community of about 20’000 scientists?

http://scoap3.org/files/Scoap3ExecutiveSummary.pdf
http://scoap3.org/files/Scoap3WPReport.pdf

SCOAP³ in one line

A consortium sponsors HEP publications and makes them OA by re-directing subscription money.

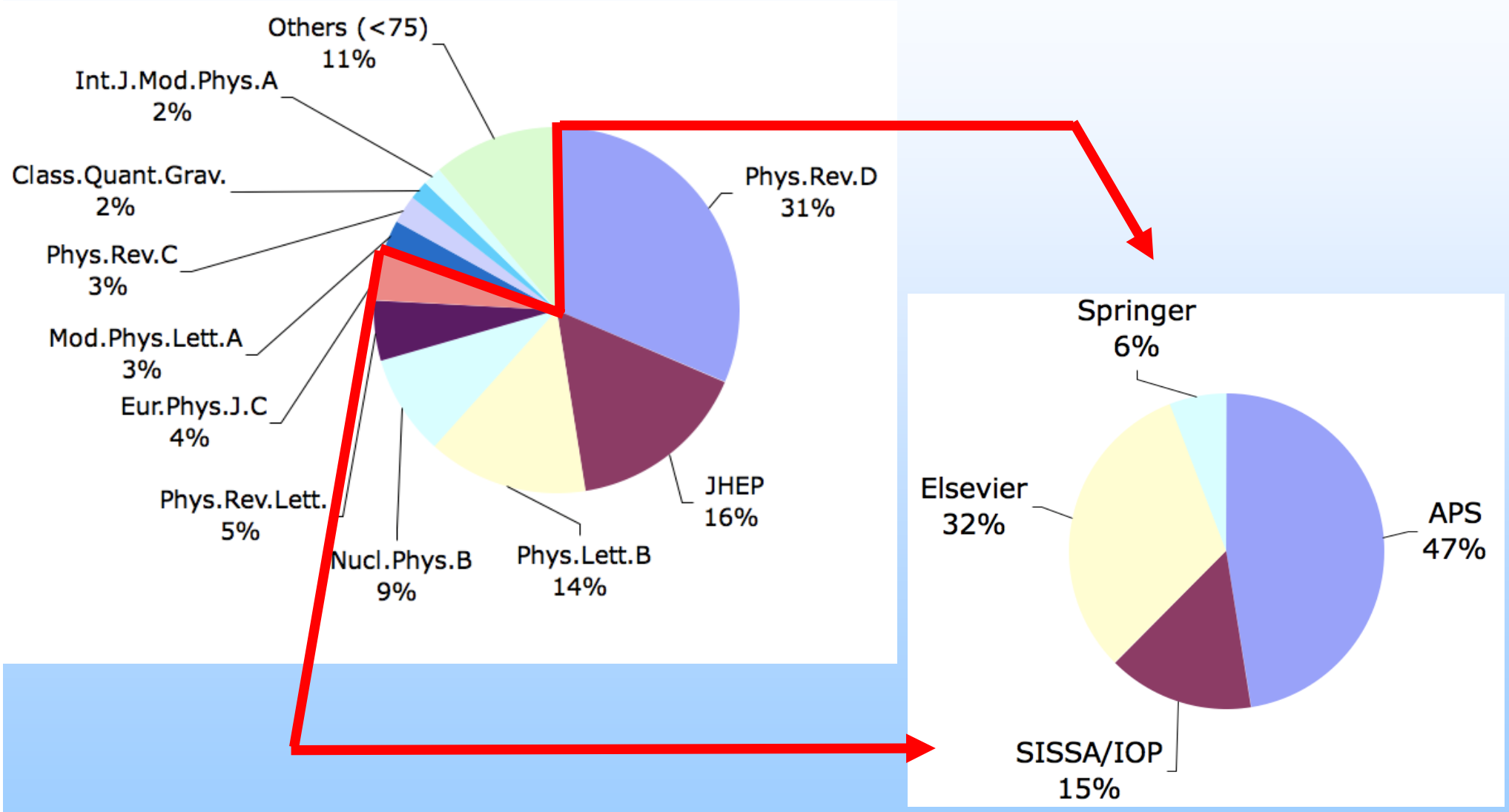
Today: (funding bodies through) libraries buy journal subscriptions to support the peer-review service and to allow their patrons to read articles.

Tomorrow: funding bodies and libraries contribute to the consortium, which pays centrally for the peer-review service. Articles free to read for everyone.

A mix of sponsoring and institutional membership,
on a world-wide scale, with a 10M€/year price tag.

Potential initial partners of SCOAP^3

Journals where HEP researchers mostly publish today



How to organize this?

HEP is used to large collaborations

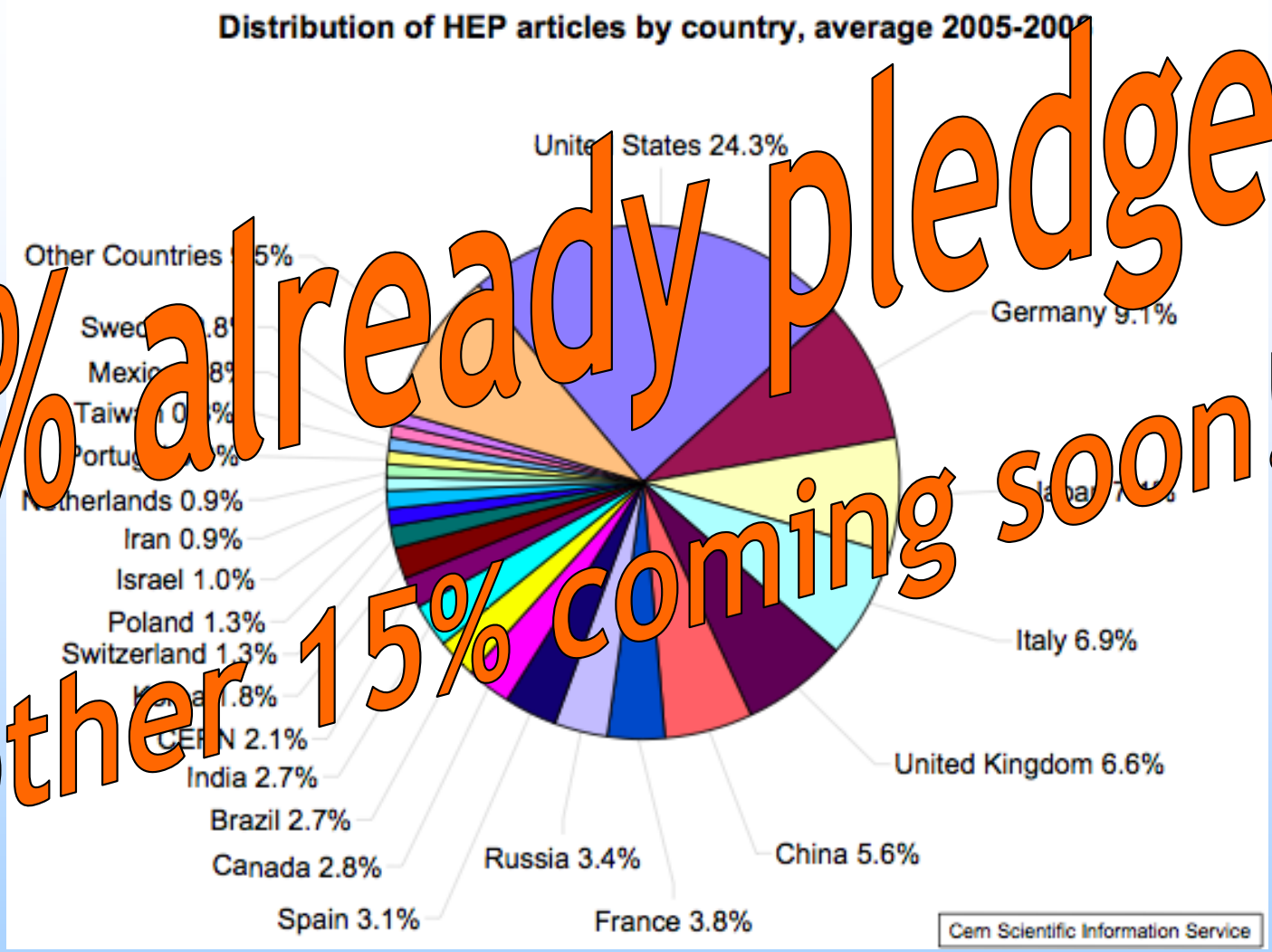
It works already on a much bigger scale

Establish OA with the same structure

SCOAP^3 fund-raising

- **SCOAP^3 financing to be distributed according to a "fair-share" model based on the distribution of HEP articles per country, accounting for co-authorship.**
- **Make a 10% allowance for developing countries who at the beginning might not contribute to the scheme.**
- **Once a sizeable fraction of budget is pledged send a tender to publishers and determine final budget**
- **The model is viable only if every country is on board!**
  Allowing only SCOAP^3 partners to publish Open Access simply replicates the subscription scheme.
- **Goal: SCOAP^3 operational for the first LHC articles!**

SCOAP^3 fund-raising



38% already pledged!

another 15% coming soon!

J.Krause, C.M.Lindqvist,S.Mele CERN-OPEN-2007-014

Germany, Italy, France, CERN, Sweden, Denmark, Norway, Greece, Slovakia, Austria, the Netherlands have already joined. Most European countries expected to join in the near future. Intense discussions in Asia and the Americas. Leading US libraries signing up.

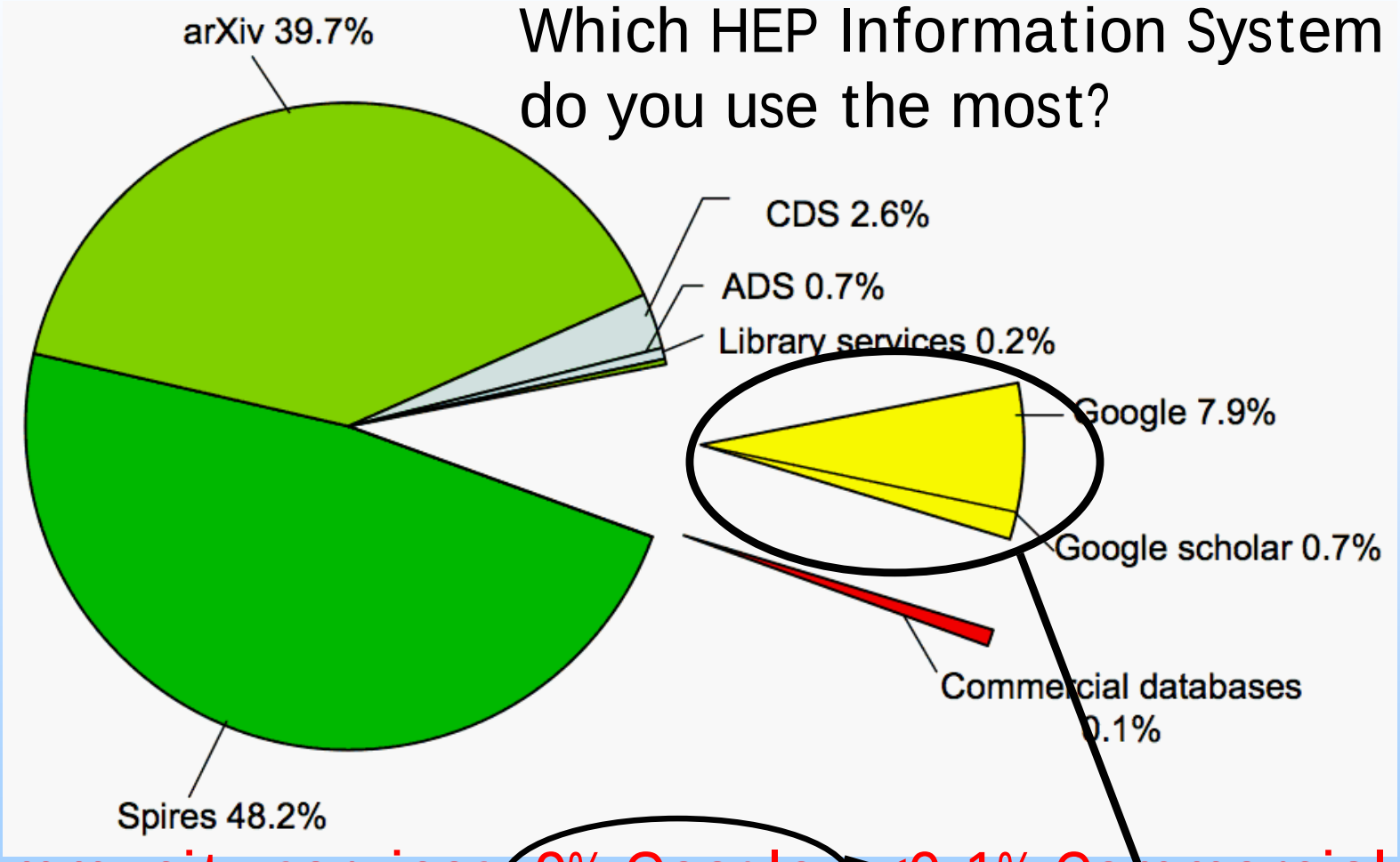
HEP Databases

From SPIRES to INSPIRE

Information discovery in HEP

A poll of the HEP community

>2000 answers (10% of the community!)



91 % Community services | 9% Google | <0.1% Commercial services

- 40 % Subject repositories
- 51 % Lab-supported databases

6% for scholars > 6 career years
22% for scholars < 2 career years

Vision for an e-Infrastructure for HEP scientific communication

May'07: Kick Off: HEP Information Summit @ SLAC

May'08: Summit @ DESY

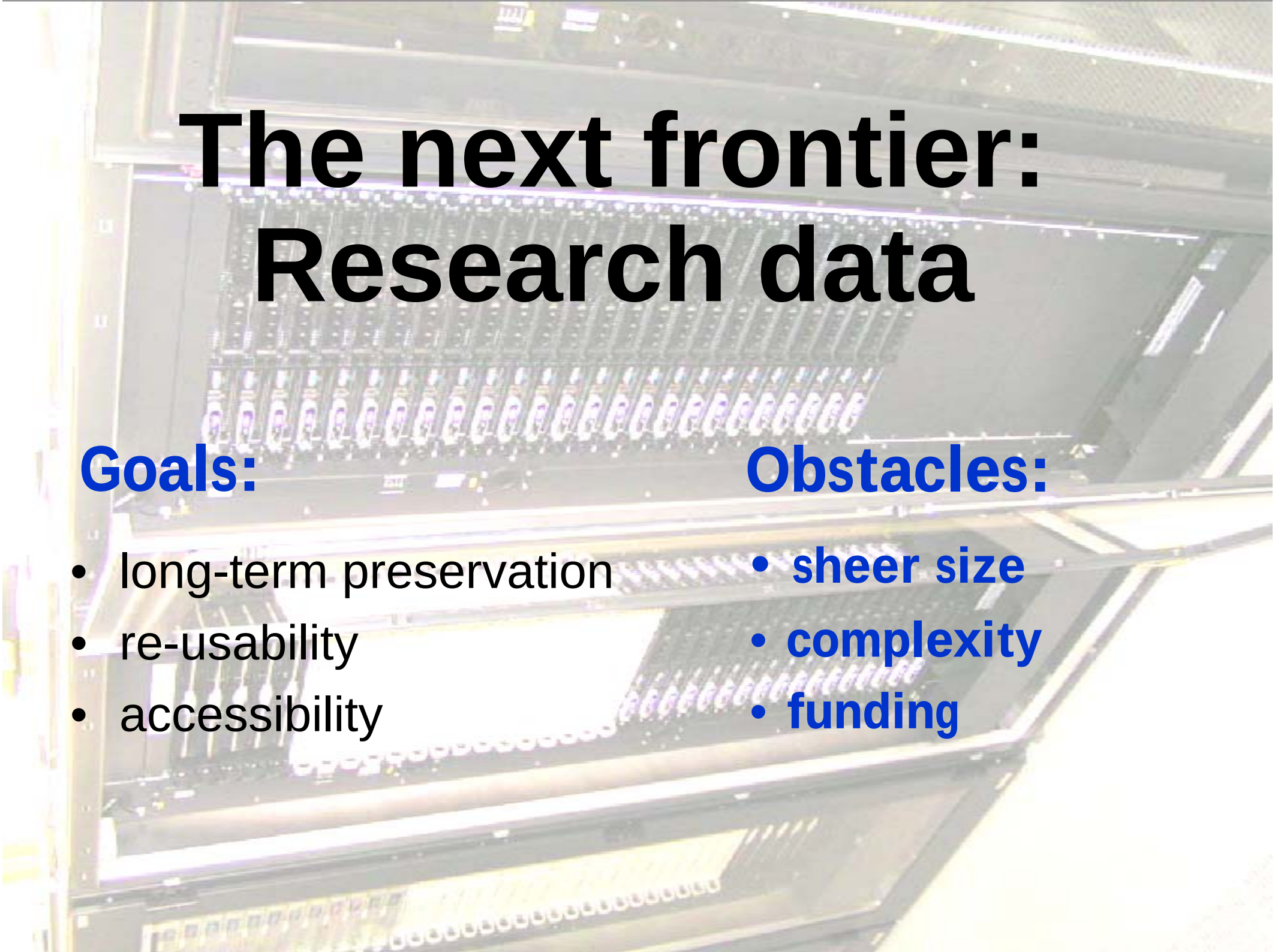
Meeting of all HEP information providers incl. Publishers and neighboring disciplines

EoI signed by CERN, DESY, SLAC and Fermilab to:

- Build a complete HEP e-Infrastructure with:
- Text- and data-mining applications
- Web 2.0 technology

Roadmap

- **Summer/Autumn 08:**
  - **Alpha Test**
- **Next steps:**
  - **Alpha public interface → beta public interface**
    - Live updating from SPIRES
    - Bug fixes etc.
  - **alpha data-administration interface**
    - Simple data input and management
  - **Collaboration with other groups**
  - **Migration of SPIRES**



The next frontier: Research data

Goals:

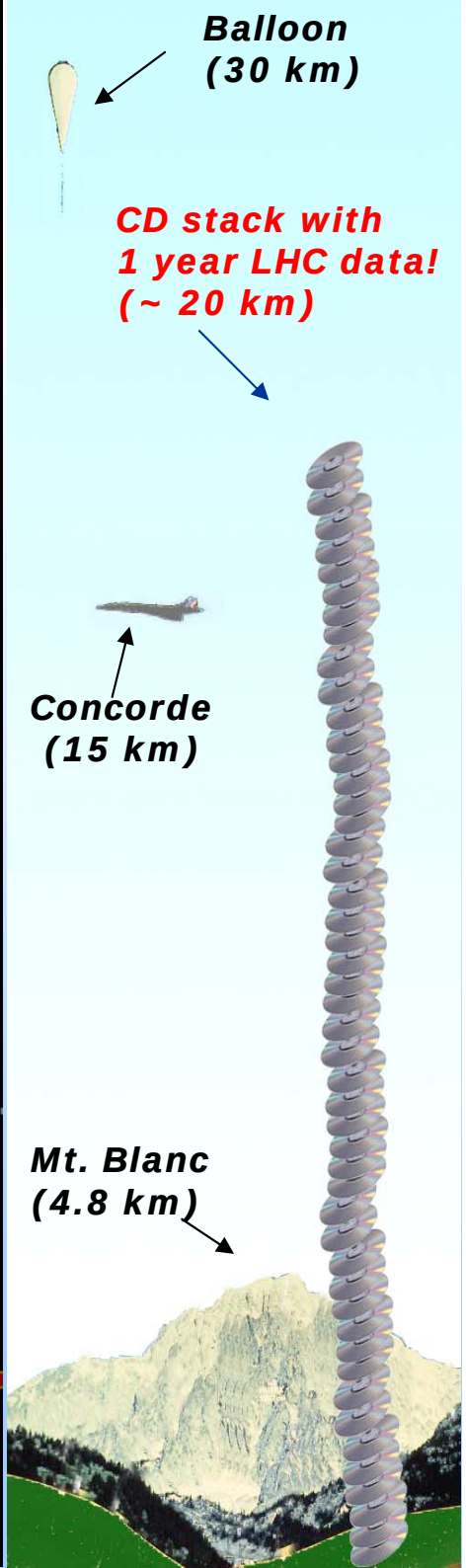
- long-term preservation
- re-usability
- accessibility

Obstacles:

- **sheer size**
- **complexity**
- **funding**

Preserving HEP data?

- The HEP data model is highly complex. Data are traditionally not re-used as in Astronomy or Climate science.
- Raw data → calibrated data → skimmed data → high-level objects → physics analyses → results.
- All of the above needs duplication for *in-silico* experiments, necessary to interpret the highly-complex data.
- Final results depend on the grey literature on calibration constants, human knowledge and algorithms needed for each pass...oral tradition!
- Years of training for a successful analysis



Data archival and re-use

Billions of funds are invested in colliders and experiments all over the world.

If data can not be re-used after the experiment stopped this investment is not exploited to its full capability.

LEP@CERN
HERA@DESY
TEVATRON@FNAL
KLOE@LNF
BABAR@SLAC
BELLE@KEK

- Everything one hasn't thought of or known (new models, better parametrization)
- Combination with future experiments

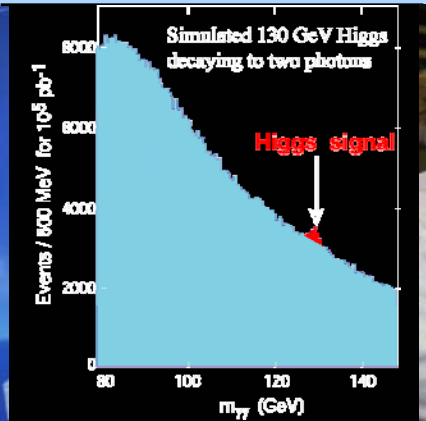
An additional relatively small fraction of the funds preserves a large fraction of the knowledge.

HEP data: The “parallel way” to publish/preserve/re-use/OpenAccess

- In addition to experiment data models, elaborate a parallel format for (re-)usable high-level objects
  - In times of need (to combine data of “competing” experiments) this approach has worked
  - Embed the “oral” and “additional” knowledge
- A format understandable and thus re-usable by practitioners in other experiments and theorists
- Start from tables and work back towards primary data

Issues with the “parallel” way

- A small fraction of a big number still gives a large number
- Activity in competition with research time
- 1000s person-years for parallel data models need enormous (impossible?) academic incentives for realization …or additional (external) funds
- Need insider knowledge to produce parallel data
- Address issues of (Open) Access, credit, accountability, “careless measurements”, “careless discoveries”, reproducibility of results, depth of peer-reviewing
- A monolithic way of doing business needs rethinking



Conclusions

- **After more than 40 years of preprints,16 years of repositories and the web...**
- **Next step ⇨ SCOAP³: A model for Open Access Publishing**
- **Time is ripe for an e-Infrastructure for HEP Scientific Communication ⇨ INSPIRE**
- **The next challenge is the preservation of HEP data**

Thank You !