A snapshot from High-Energy Physics (HEP)

2nd Global Research Library 2020 Workshop

Tirrenia, Italy, 26 -28 March 2008

Jens Vigen, CERN Scientific Information Service

Shaping a roadmap for the future of Research Libraries

Research Libraries Requirements & Trends

GRL2020
A Vision for Global Research Libraries

The HEP Community

- Well organized
- 20'000 active scientist
  - ~50% theorists, ~50% experimentalists
- Distributed across a few big international labs and hundreds of universities
- Experiments with a long life circle
- Scientific production
  - ~5'000 articles per year
  - 90% of the articles in the field of theory

The long awaited LHC machine to start this summer!

Communication pattern anno 1960

- HEP scientists cannot wait ~1 year for their articles to reach their peers through journals
  - *Preprint* are the main vehicle of information in HEP: final version of articles as sent to journals
  - Researchers (of affluent institutions) mass-mail preprints to hundreds of (prestigious and therefore affluent) institutions
  - At CERN preprints get indexed and displayed (and often discarded once published)
  - The weekly “new preprints” display is a big event

L. Goldschmidt-Clermont, 1965. http://eprints.rclis.org/archive/00000445/02/communication_patterns.pdf

GRL2020
A Vision for Global Research Libraries

Communication pattern anno 1965 2008

- HEP scientists cannot wait ~1 year for their articles to reach their peers through journals
  - *Preprint* are the main vehicle of information in HEP: final version of articles as sent to journals
  - Researchers (of affluent institutions) mass-mail preprints to hundreds of (prestigious and therefore affluent) institutions
  - At CERN preprints get indexed and displayed (and often discarded once published)
  - The weekly "new preprints" display is a big event

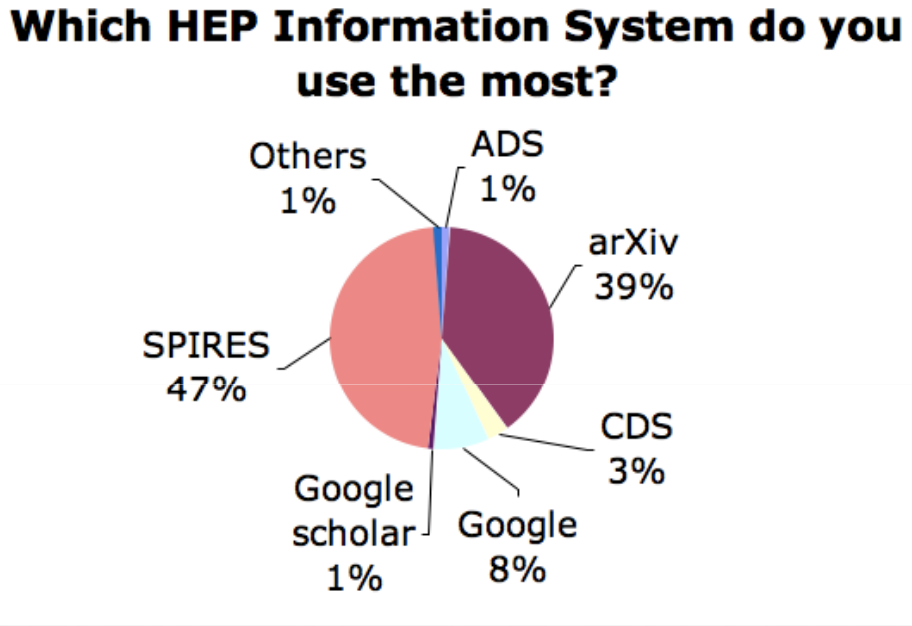
post on arXiv

CDS&SPIRES

daily email from arXiv

L. Goldschmidt-Clermont, 1965. http://eprints.rclis.org/archive/00000445/02/communication_patterns.pdf

Accessing HEP information



A poll of the HEP community >2000 answers (10% of the population!)

1% Commercial services

- <0.1% pay databases
- 1% free publisher portals

9% Internet search engines

- 8% Google
- 1% Google scholar

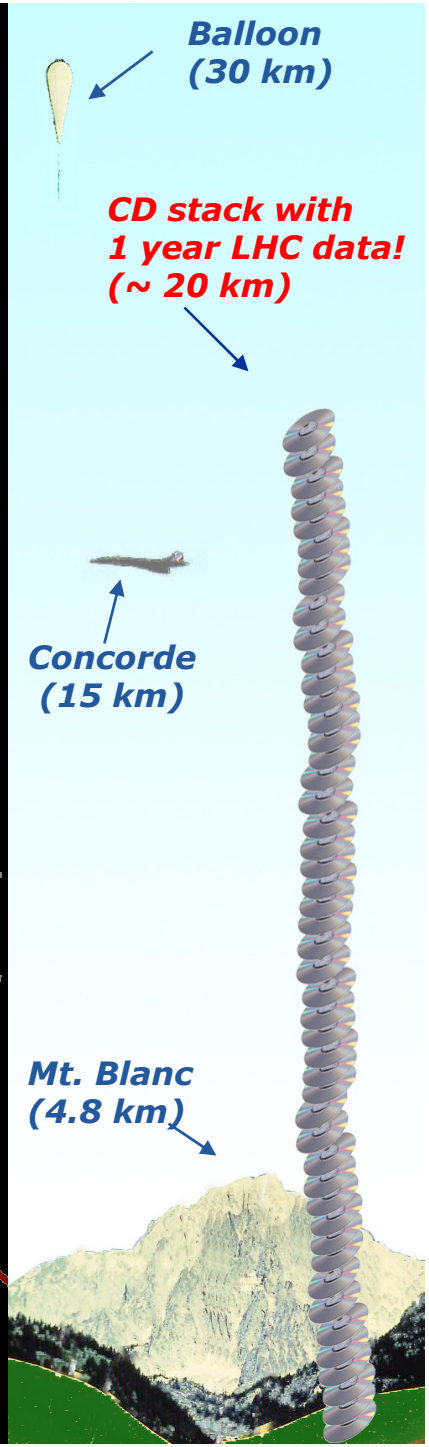
90 % Community services

- 39 % Subject repositories
- 51 % Lab-supported databases

HEP scholars use in-house systems rather than commercial services

What about HEP data?

- The HEP data model is a highly complex one. Data are difficult to be made Open Access (*i.e. usable)* as in Astronomy or Climate science.
- Raw data → calibrated data → skimmed data → high-level objects → physics analyses → results.
- All of the above duplicated for *in-silico* experiments, necessary to interpret the 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

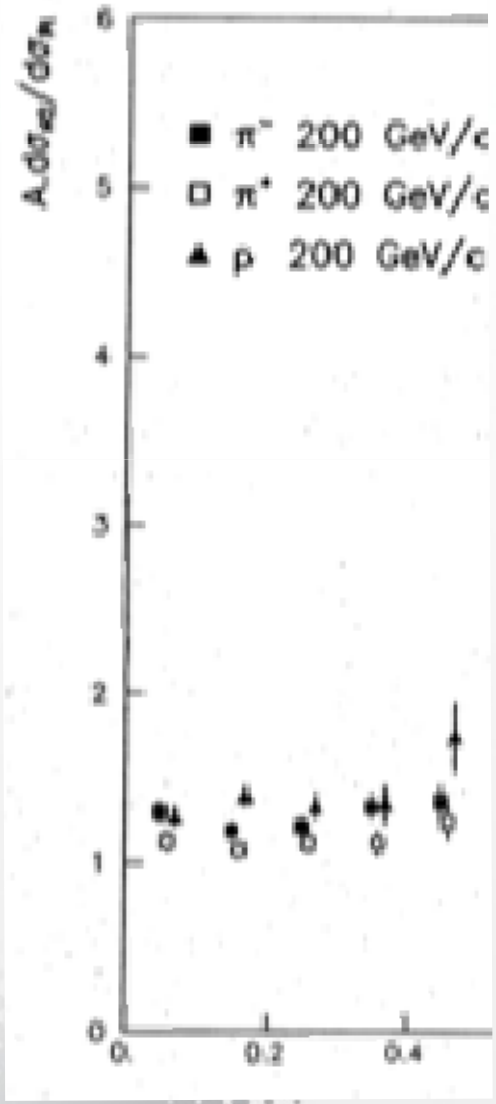


GRL2020
A Vision for Global Research Libraries

EC study under way; including HEP

PARSE.Insight (Access, Preservation, and Curation) aims to help define the infrastructure needed to preserve and use the digitally encoded information on which our society increasingly depends, but which is so fragile. These digital resources contain the intellectual inheritance we leave across time for future generations and which we need to exploit to the fullest, right now, across domains.

eScience anno 2008



From: Senior CERN physicist
Date: March 19, 2008
To: Jens Vigen
Subject: help with values from figure

Dear Jens,

I need your help :-)) I need the numerical values (including errors) of a very old figure...It is from the paper Z. Phys. C 20 (1983) 101, by the NA3 experiment. It is figure 2, only the right side plot and only one of the three data sets in there, the one represented with triangles. In fact, I only want the numbers of the first points, those in the range xF= 0.0 -- 0.4.The paper exists online but the resolution is lousy and I wanted to have a good measurement of the values from the figure. My question is: could you (or one of your expert colleagues) do a high resolution scan of that figure, in the region of the four leftmost points,and send it to me as a PDF file? I will then print it and measure the values with a pencil and a ruler :-((Thanks a lot !!!

GRL2020
A Vision for Global Research Libraries

How to get there …?

- HEP repositories - to be streamlined further
  - Collecting preprints/alerting peers: arXiv
  - Peer review/Open Access publishing: SCOAP^3
  - Curation and preservation: CDS+SPIRES=INSPIRE
- HEP data - waiting for outcome of PARSE.Insight
  - Founding member of the Alliance for Permanent Access
  - Start from tables and work back towards primary data
  - How much additional work? 1%, 5%, 10%?
- 2nd HEP Information Summit in Hamburg May ‘08
  - Comprises all stake holders; scientists, librarians and publishers
  - Motivated by user driven developments vs. “technology push”
  - First “simple step”: Introduce a unique author-authority list

GRL2020
A Vision for Global Research Libraries

Exciting times are ahead!

- With 50 years of preprints and soon 20 years of repositories and web, HEP has spearheaded (Open) Access to Scientific Information
- Time is ripe for an e-Infrastructure for HEP Scientific Communication
  - Build a complete HEP information platform
  - Enable text- and data-mining applications
  - Demonstrate and deploy Web2.0 applications
- The next challenge is the preservation of HEP data
- Eventually, a sandbox for grid-backed applications