Brief Report on DPHEP Activities

(borrowed almost entirely from C. Diaconu ICFA Panel Report of 3/12/21)

O. Rind

SDCC Staff Technical Meeting

04/08/21

Data Preservation in High Energy Physics

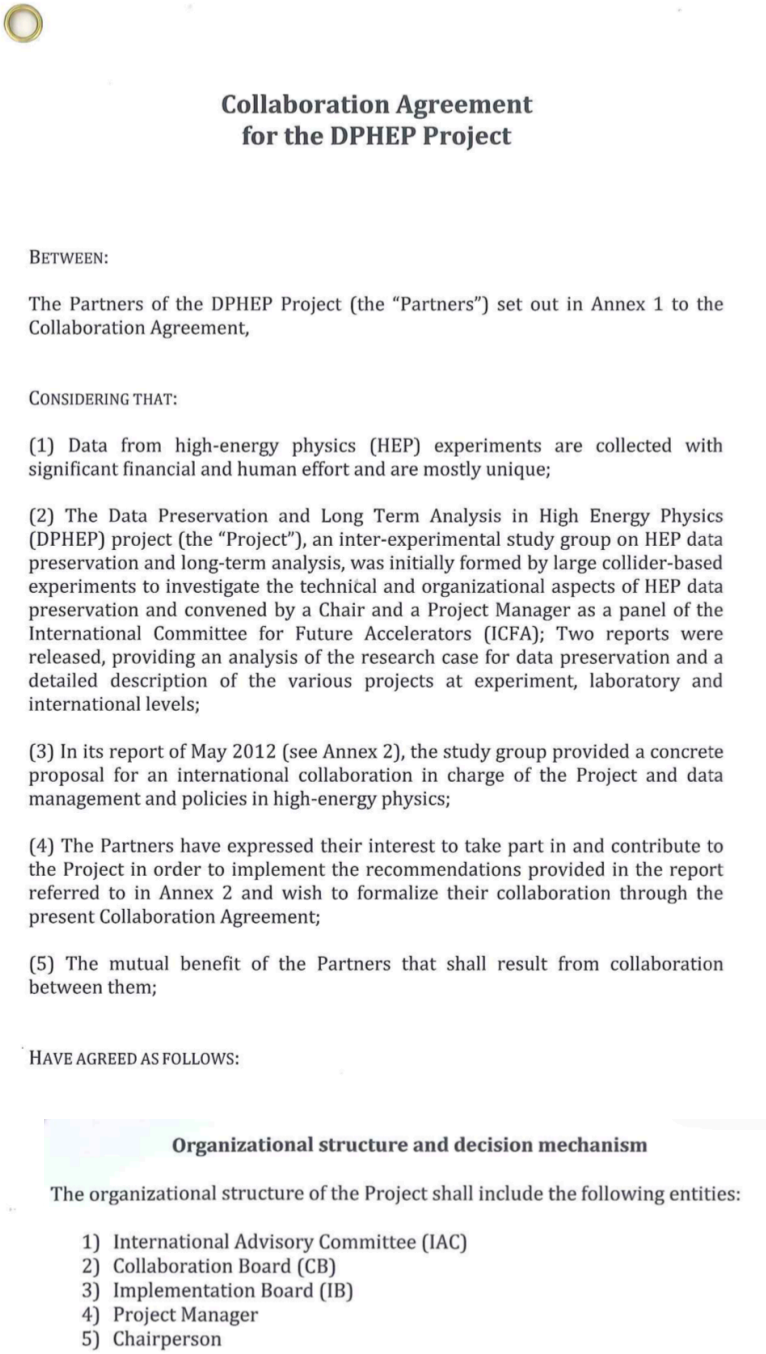
ICFA Panel Report 12/03/2021

DPHEP Study Group for Data Preservation and Long Term Analysis in High Energy Physics

Cristinel DIACONU
CPPM/CNRS/Aix-Marseille University

The DPHEP Collaboration

- Collaboration Agreement was signed in 2014
  - Give a clear sign of the will of labs to collaborate in this common challenge
- Members:
  - 2014: CERN, DESY, HIP, IHEP, IN2P3, KEK, MPP
    - 2015 IPP/Canada , 2017 UK/STFC
  - Active labs from US, Italy have not formally joined, but are represented in the Collaboration Board.
- The DPHEP collaboration continue to act as an ICFA panel, as indicated in the Collaboration Agreement
  - About 60 contact persons FA, Labs, experiments
- DPHEP Activity
  - Global reports 2009(whitepaper), 2012 (blueprint), 2015, 2017 (global reports)
  - Collaboration meetings: 2015, 2017
  - Remote panel discussion March 2nd 2021



Collaboration Agreement for the DPHEP Project

BETWEEN:

The Partners of the DPHEP Project (the "Partners") set out in Annex 1 to the Collaboration Agreement,

CONSIDERING THAT:

(1) Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique;

(2) The Data Preservation and Long Term Analysis in High Energy Physics (DPHEP) project (the "Project"), an inter-experimental study group on HEP data preservation and long-term analysis, was initially formed by large collider-based experiments to investigate the technical and organizational aspects of HEP data preservation and convened by a Chair and a Project Manager as a panel of the International Committee for Future Accelerators (ICFA); Two reports were released, providing an analysis of the research case for data preservation and a detailed description of the various projects at experiment, laboratory and international levels;

(3) In its report of May 2012 (see Annex 2), the study group provided a concrete proposal for an international collaboration in charge of the Project and data management and policies in high-energy physics;

(4) The Partners have expressed their interest to take part in and contribute to the Project in order to implement the recommendations provided in the report referred to in Annex 2 and wish to formalize their collaboration through the present Collaboration Agreement;

(5) The mutual benefit of the Partners that shall result from collaboration between them;

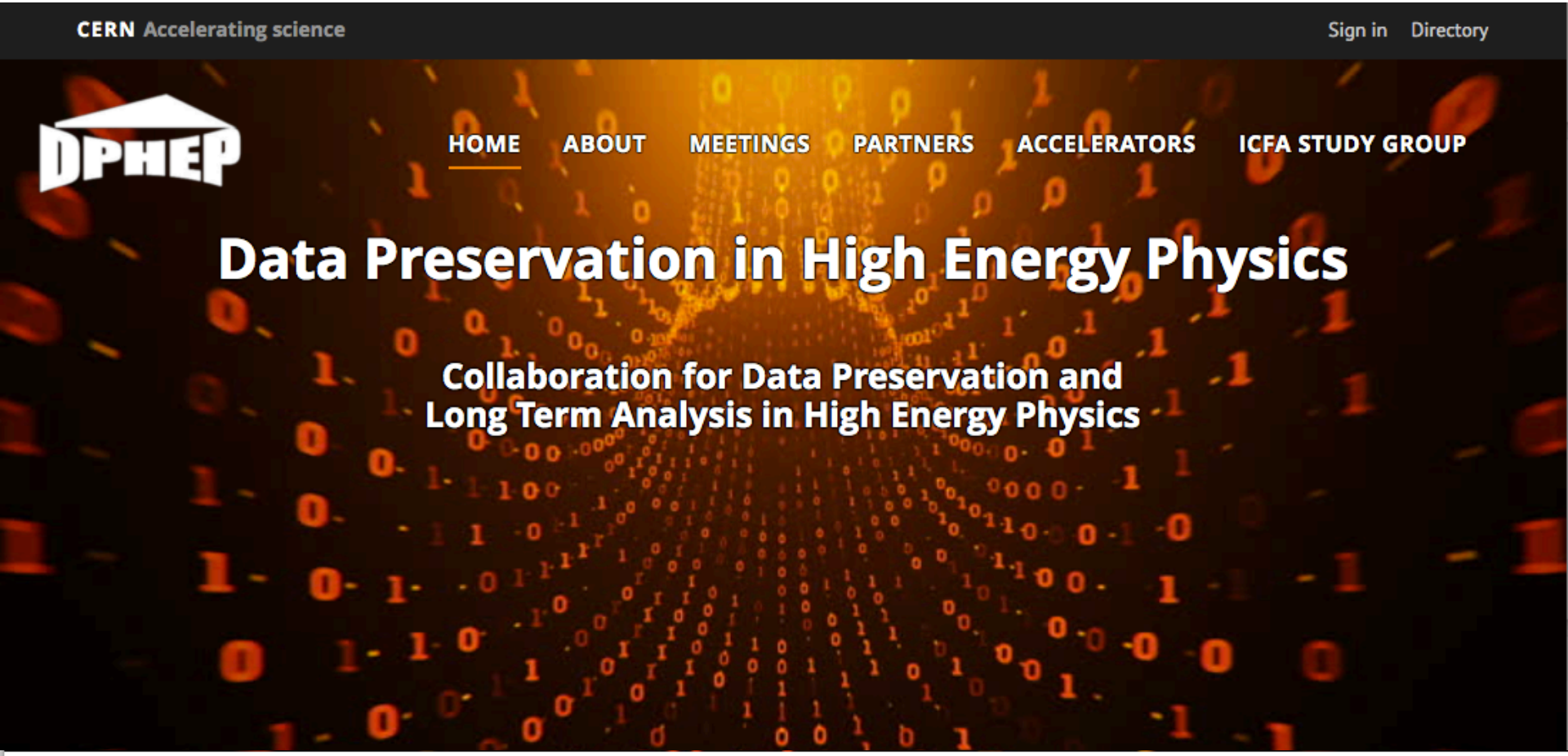
HAVE AGREED AS FOLLOWS:

Organizational structure and decision mechanism

The organizational structure of the Project shall include the following entities:

1) International Advisory Committee (IAC)
2) Collaboration Board (CB)
3) Implementation Board (IB)
4) Project Manager
5) Chairperson

http://dphep.org DPHEP Portal part of "DPHEP 2020 Vision"



WELCOME

DPHEP is "Data Preservation in HEP" and is an International Collaboration of Institutes, Experiments, Funding agencies and other interested parties to implement the recommendations of the DPHEP study group. These are detailed in the Blueprint document available here.

As with the former study group of the same name, DPHEP reports to ICFA.

3/2/21 DPHEP remote preparatory discussion

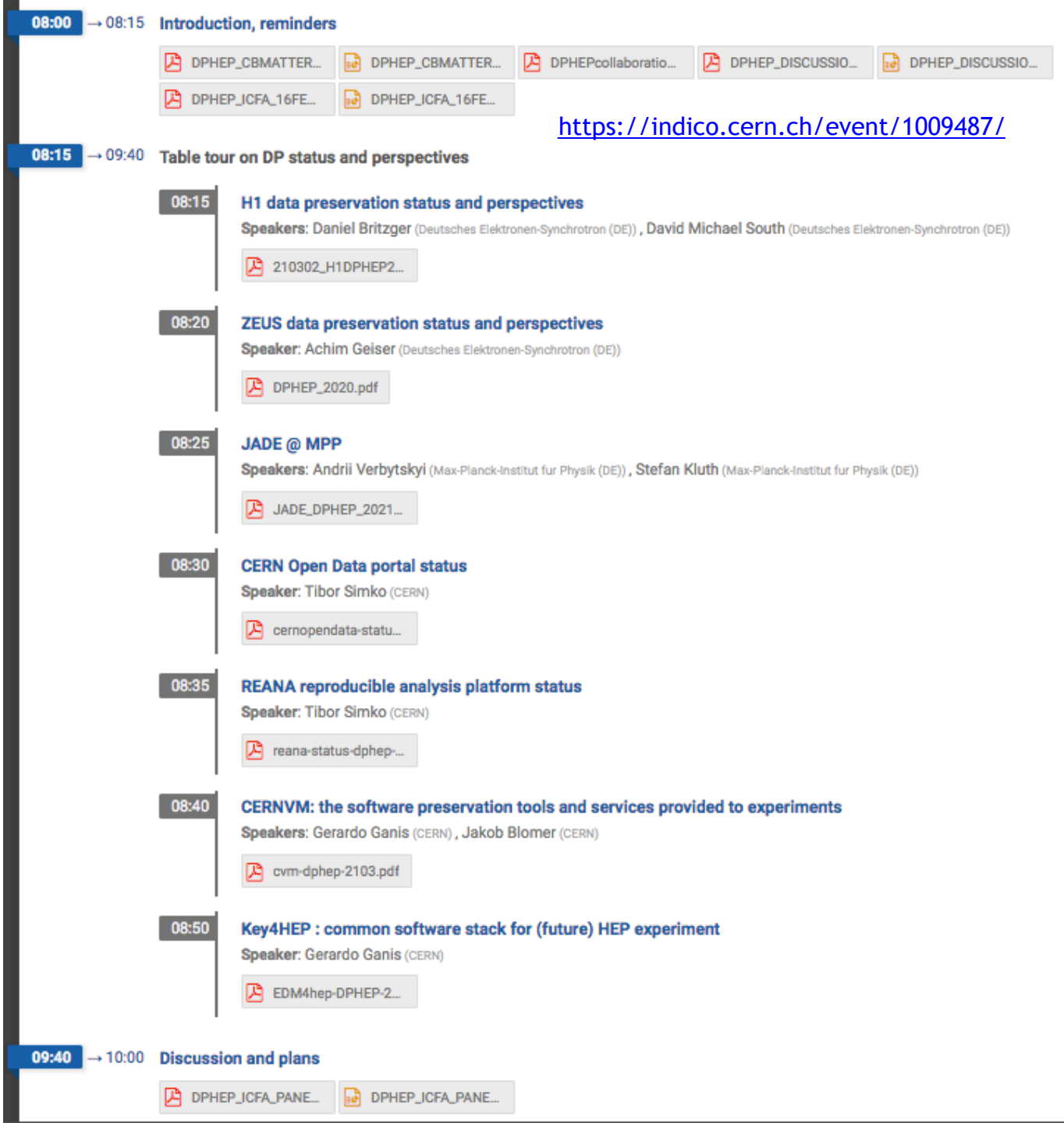
Dirk Duellmann (CERN) joined the collaboration executive team and will take over from **Jamie Shiers** as project manager.

In the past years the HEP data management landscape have seen spectacular changes. Some of those concrete actions have been taken in line with the DPHEP recommendations. 12 years since the first discussions give us a nice time perspective and deserve a common reflection and feedback from/to the community.

The next ICFA meeting on March 11-12 will include a panel tour where it would be useful to present a viewpoint (the last report was presented in spring 2018).

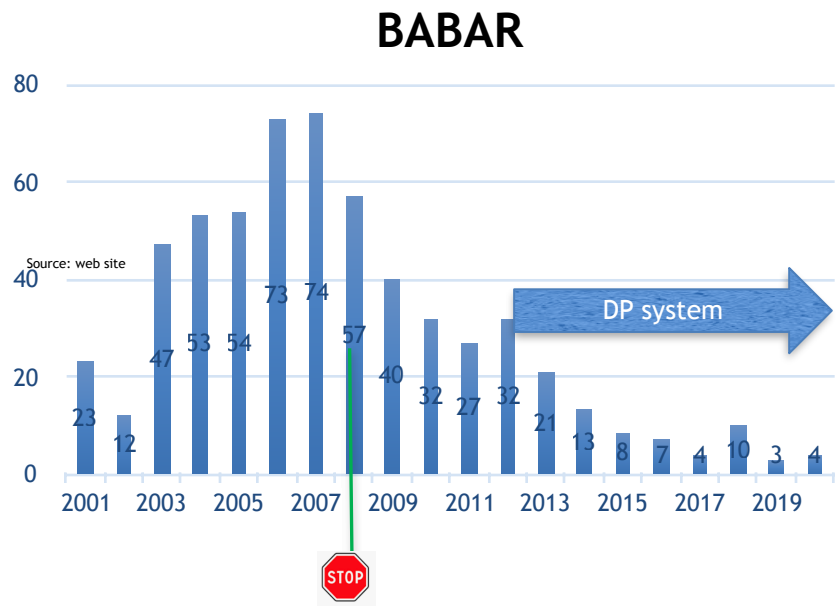
https://icfa.fnal.gov/

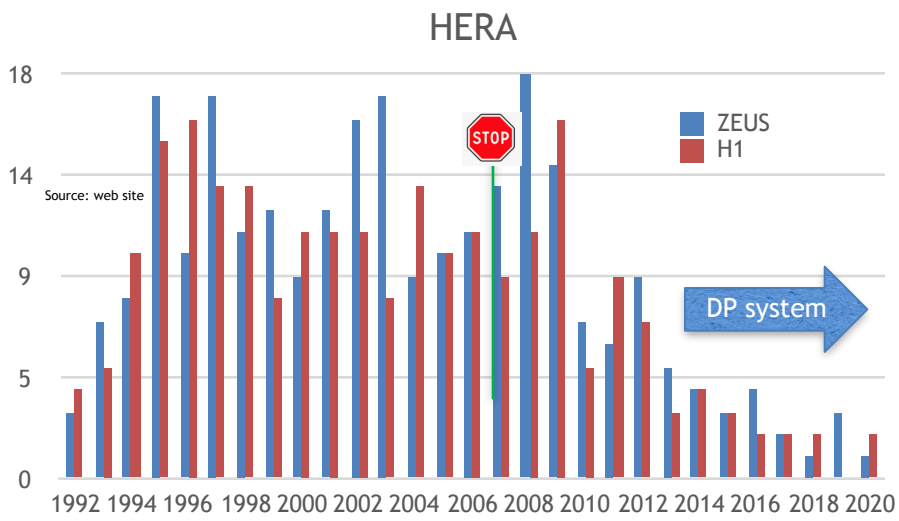
ICFA

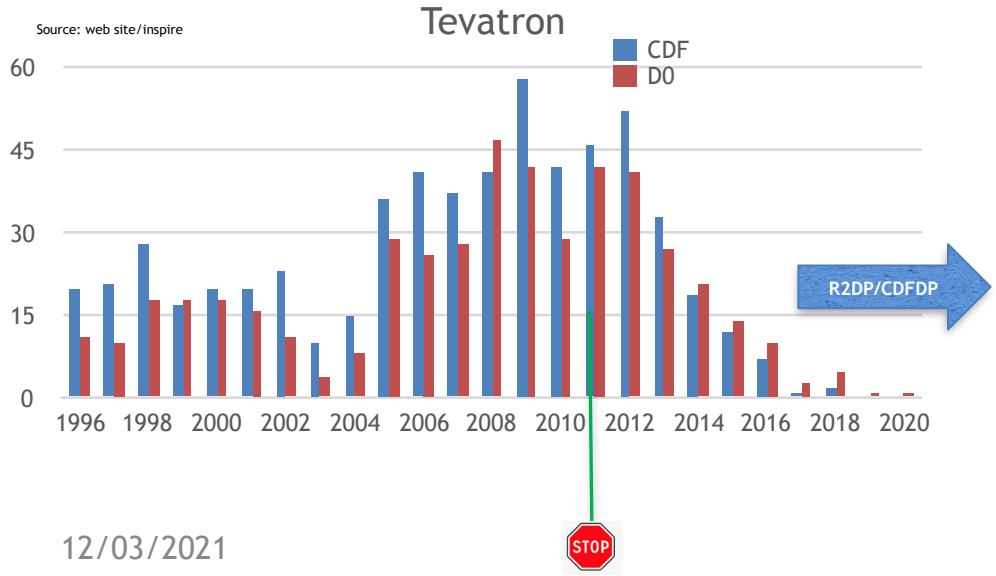


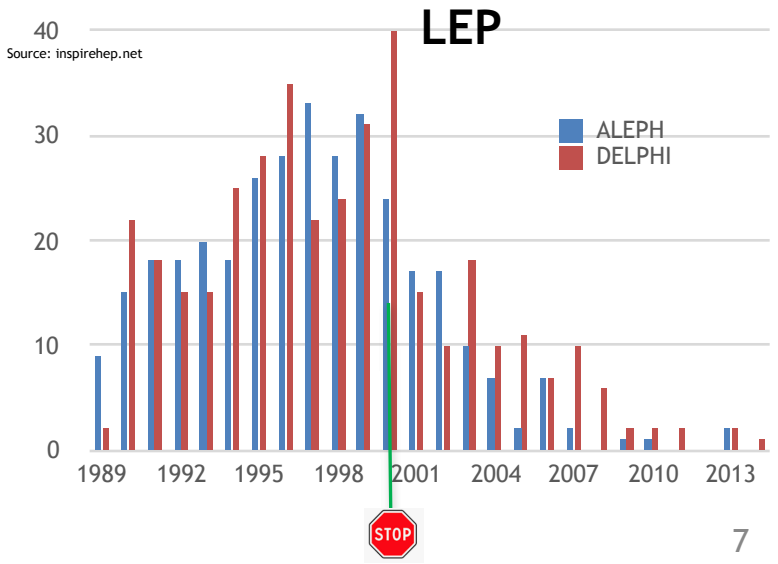
Panel remote discussion: March 2nd

Scientific output from preserved data







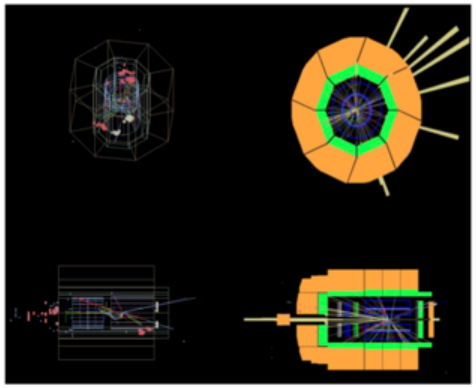


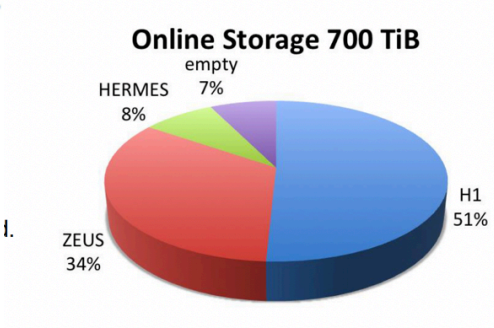
Data Preservation projects labs: recent update

- **@DESY**: H1 (migration) and ZEUS (encapsulation) in great shape
  - successful transitions to the DP systems, publication plans continues and includes O(10) papers
  - objective: alive by 2030; New institutes joining (synergy with EIC)
- **@CERN**: **strong LHC activity**, LEP data/sw refreshed, OD/OS standards/technologies, DPHEP portal
  - Need for the continuation of the central management support
- **@MPI**: multi-experiment framework explored (JADE, HERA, OPAL)
  - JADE on a desktop
- **@KEK**: BELLE I data readable in Belle II framework ;
  - objective maintain Belle I data by 2023 (when the precision will be exceeded by the new data)
- **@IHEP/BES3**: The experiment is expected to stop data taking by 2022
  - Data to be preserved for 15 years
  - Strong support to DP national and international activities expressed
- **@BNL/JLAB**: DP activity ongoing (ATLAS, EIC), discussed with NP Community
- **@Babar:** LTDA supported analysis since 2012. SLAC support ended in February. Data almost entirely copied to CERN/GridKa.
  - Data saved at CERN/GridKa: ~ 1.2 PB+ 0.5 PB (ongoing), Minimal user infrastructure for ongoing analyses and documentation hosted at U. of Victoria.
- **@FNAL**: (indirect news this time) transition to a DP system for both CDF (CDFDP) and D0 (R2DP)
  - Data stored/saved @FNAL+Italy, 500th paper from D0 in 2021

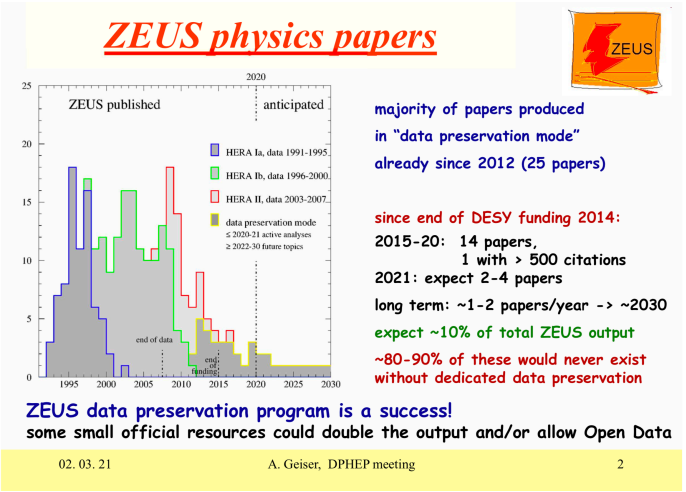
HERA: successful DP, towards open data

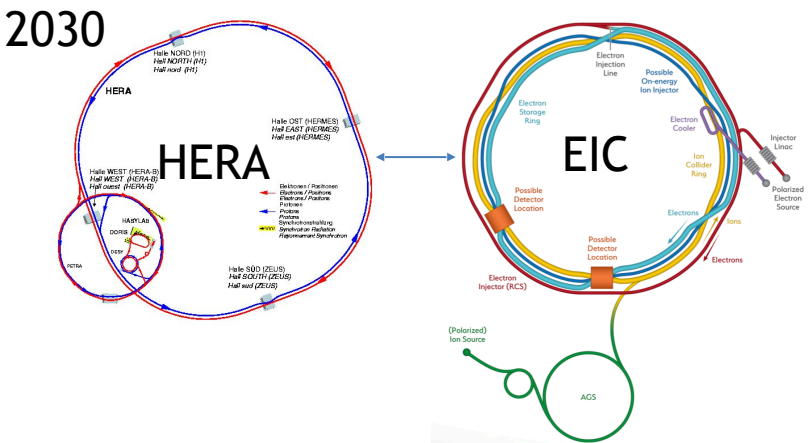
- H1: “Level 4” DPHEP strategy
  - All data, full migration, including regular recompilation/validation
  - Recent “technology jump” successful : in line with modern tools
    - “LHC”-like tools, ready for open data





- ZEUS : “Level 2/3” DPHEP strategy
  - Root ntuples produced in the preparatory phase
  - easy to maintain/use/test/open





- New topics/collaborators (EIC)

JADE

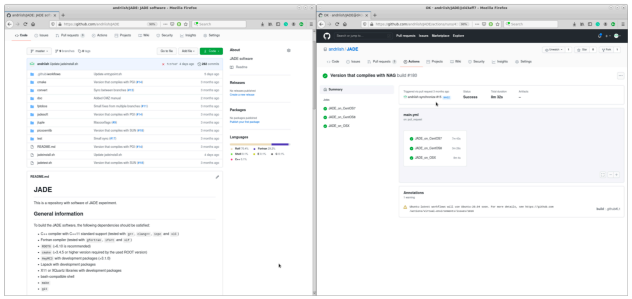
- The JADE experiment was located at the PETRA e+e− storage ring at DESY in Hamburg, Germany. The experiment took data between 1979 and 1986 in the center-of-mass range between 12 and 46.6 GeV.
- The size of JADE data is much smaller (600Gb) than of modern experiments → one can use MPCDF ownCloud to store/ access it
- “The main thing about the JADE software: it is Fortran”
- JADE DP stack is based on open standards, does not rely on specific SW and is extremely portable. One can run it completely on desktop.
  - “JADE – collider experiment on your desktop”.

Data Preservation model *circa* 1980s

JADE software: recent developments

More portability, testing and documentation.

- GNU and IBM toolchains support extended with preliminary Intel[NEW] and NAG[NEW]. GNU is still the most stable one.
- More CI tests[NEW].
- Updated the site and documentation[NEW].
- Support for CentOS8[NEW] and MacOSX10.15+ on x86_64[NEW]



LHC Data Preservation

- Data Preservation and Open Access policies (already since 2012-2014)
  - DP is a « specification » included in the computing models and plans for upgrades
  - HEP Software Foundation Roadmap
- Strong initiative on Open Data and Open Science policy
- Concrete implementation and technology-oriented survey
  - Very active multi-experiment projects
  - data re-use, reanalysis, reinterpretation, outreach etc.
    - OpenData, Analysis Preservation, REANA…

A Roadmap for HEP Software and Computing R&D for the 2020s

arXiv:1712.06982

CERN announces new open data policy in support of open science

A new open data policy for scientific experiments at the Large Hadron Collider (LHC) will make scientific research more reproducible, accessible, and collaborative

11 DECEMBER, 2020

nature physics

Explore Content | Journal Information | Publish With Us

nature › nature physics › perspectives › article

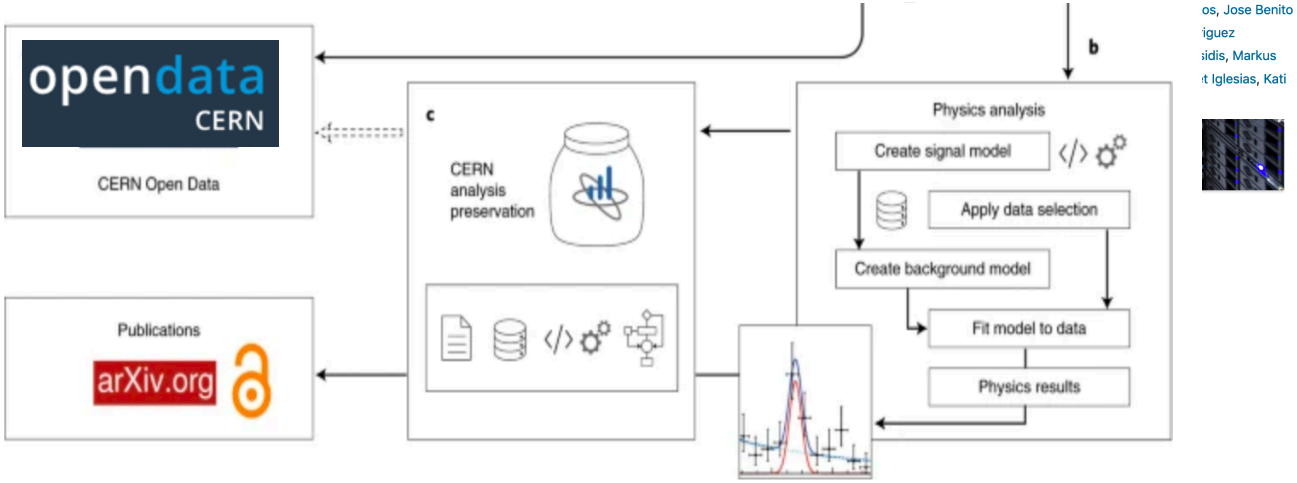
https://www.nature.com/articles/s41567-018-0342-2

Perspective | Open Access | Published: 15 November 2018

Open is not enough

os, Jose Benito
iguez
idis, Markus
t Iglesias, Kati

Filter by experiment	2017	2021
ALICE	15	15
ATLAS	101	109
CMS	878	1167
LHCb	3	4
OPERA		904



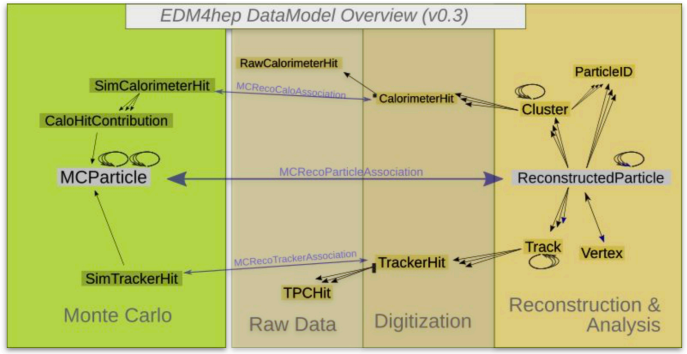
Other experiments expressed clear intention to join : LEP, JADE, H1/ZEUS, BaBar (HR is an issue)

Towards more standards

- Key4hep - low maintenance, customizable, common turnkey software stack, needs…..
  - Driven by ILC, FCC….
- Requires a common event data model with a common language: EDM4hep
- In active development
  - See three contributions at vCHEP21

EDM4hep: the common language

- The Event Data Model describes the structure of the data
  - Challenge: can we have the same for all HEP experiments? LCIO shared by ILC and CLIC
- Heavily inspired by LCIO and FCC-edm



key4hep / EDM4hep and DPHEP?

- Key4hep / EDM4hep: framework with longer perspective than a single experiment
  - Not just *another data format*, but one that might become a standard
- Requires “migration”, which may be a pain or not even possible
  - Workpower / Experts missing
    - *Encapsulation* may help here, both for migration and validation
- For LEP data, FCC-ee may provide a unique opportunity
  - Share to center-of-mass energies: 91.2 GeV, 160 GeV
    - Clear advantage in looking at what real data look like to understand bottle necks and limitations
    - Possible student projects
  - ALEPH: early investigations promising
    - ALPHA++ provides the relevant code for migration
    - Several ALEPH *experts* involved in FCC-ee studies

Situation and trends

- Significant/measurable impact of dedicated DP projects @expts./labs
  - Production of high quality and unique scientific results at very low (non-zero) cost
    - 10% output for less than 1% investment: ✓
  - Signs of re-invigorating collaborations in the context of new projects
    - HERA-EIC; LEP-FCCee
  - Case for longer term preservation: data sets parking
    - CDF, D0, Babar, LEP, Jade : carefully follow the usability in time
- LHC exps. very active in DP and strongly linked to **Open Data/Science**:
- The (DP)HEP future is also considered
  - FCC, EIC : transfer of knowledge in DP from LHC/oldies
- And more is possible on:
  - Education, training, outreach….
    - open data projects are an opportunity to reinforce these aspects as well

- The panel expresses the need to keep the issue highly visible on the community's agenda
  - ensure an adequate level of endorsement from FA/Labs/Experiments

Next steps

- DPHEP as a collaboration
  - CERN support needed: focal point of ongoing major experiments/computing standards
  - Reinforce Laboratory and FA contacts
  - DPHEP Workshop : July 2021
    - Collaboration Board meeting, management evolutions needed

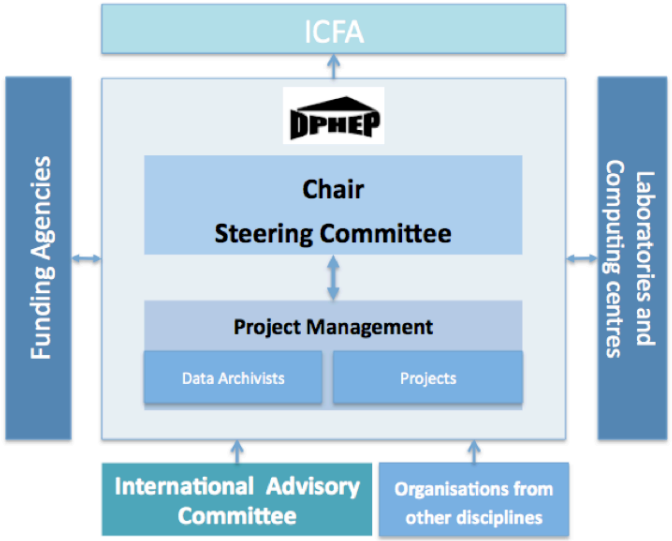
- DPHEP as an ICFA panel:
  - a mandate prolongation is considered as a very useful asset

- Objectives for 2021-2024:
  - improve the awareness and stimulate improvements on DP
    - Scientific motivation, organization, technologies, standards, outreach and education
    - Organize Workshops / issue Global Reports, link to other communities
  - reinforce and support the ongoing laboratory-based projects and their cooperation
    - keep alive data sets that (can) still produce science, keep track on parked data sets
  - support/develop the DP aspects for future experiments and encourage the ToK
  - encourage open data and open science as a way to preserve data and knowledge

BACKUP

The DPHEP Collaboration

- October, 2012: CERN endorses the blueprint and appoints the DPHEP Project Manager (Jamie Shiers)
- Retain the basic structure of the Study Group, with links to the host experiments, labs, funding agencies, ICFA
- The collaboration agreements signed in 2013



Dear Dr. Diaconu,

Following the delivery of the final DPHEP blueprint, various inputs received into the European Strategy for Particle Physics symposium earlier this week and after consultation with my colleagues, I would like to inform you that CERN offers to provide the role of the initial DPHEP project manager.

We would propose to appoint Jamie Shiers in this role for an initial period of 3 years starting 1 January 2013, after which the role may be assumed by another laboratory, as suggested in the blueprint.

We would anticipate that during this period the DPHEP organization will be launched (year 1) and that the initial deliverables defined in the blueprint would be achieved.

CERN would also foresee participation in the other activities described in the document in areas such as R&D into the use of virtual machine technology for data preservation purposes (PH-SFT input to ESPP) and into the management of very large data stores.

Yours sincerely,

Sergio Bertolucci
Director for Research and Computing

The DPHEP Collaboration

- The DPHEP ICFA panel lead to a Collaboration officially started after the Collaboration Agreement was signed in 2014 by several large laboratories and funding agencies
  - Give a clear sign of the will of all labs to co-operate and collaborate in this common challenge
- Members:
  - 2014: CERN, DESY, HIP, IHEP, IN2P3, KEK, MPP
  - 2015: IOP
  - US institutes, UK, Italy have not formally joined, but are represented in the Collaboration Board.
- Retain the basic structure of the Study Group, with links to the host experiments, labs, funding agencies
- The DPHEP collaboration continue to act as an ICFA panel, as indicated in the Collaboration Agreement.

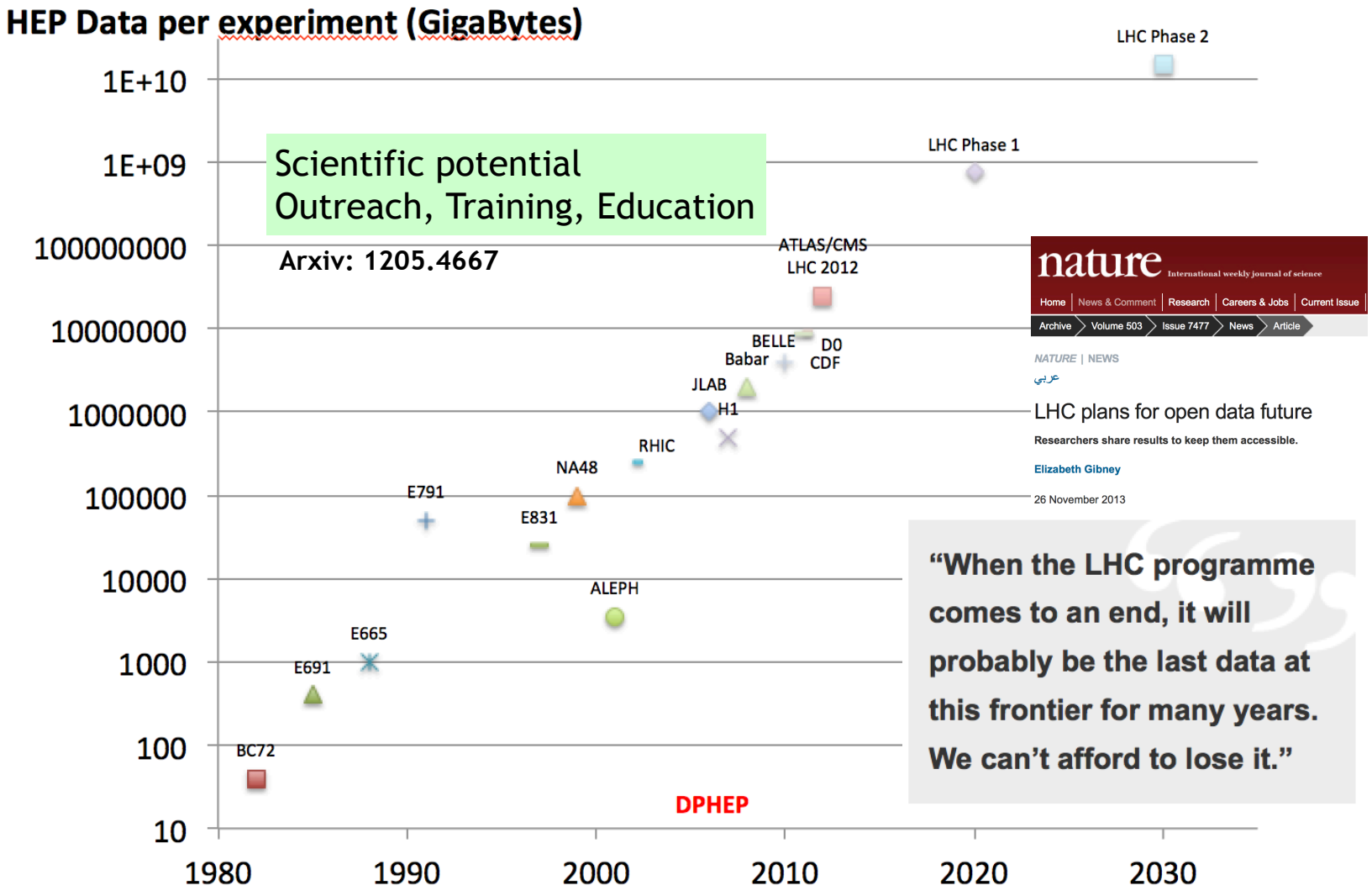
Annex 1: Partners of the DPHEP Project and contact persons 2014

Initial DPHEP Partner	Location	Contact person
European Organization for Nuclear Research, **CERN**	Switzerland	J. Shiers
Deutsches Elektronen-Synchrotron, **DESY**	Germany	D. South
Helsinki Institute of Physics, **HIP**	Finland	K. Lassila-Perini
Institute of High Energy Physics, **IHEP**	China	G. Chen
Institut national de physique nucléaire et de physique des particules, **IN2P3**	France	G. Lamanna
Institute of Particle and Nuclear Studies, High Energy Accelerator Research Organisation, **IPNS, KEK**	Japan	T. Hara
Max Planck Institut für Physik , **MPP**	Germany	S. Kluth

Following institutes are members of the DPHEP Study Group and intend to join formally the DPHEP Collaboration:

Brookhaven National Laboratory, **BNL**	USA	M. Ernst
CSC- IT Center for Science	Finland	N.N.
Fermi National Accelerator Laboratory, **FNAL**	USA	S. Wolbers
Institute of Particle Physics, **IPP**	Canada	R. Sobie Joined 2015
Istituto Nazionale di Fisica Nucleare, **INFN**	Italy	M. Maggi
SLAC National Accelerator Laboratory	USA	C. Cartaro
Science and Technology Facilities Council, **STFC**	UK	J. Bicarregui

HEP Data


HEP Data per experiment (GigaBytes)
1E+10
1E+09
100000000
10000000
1000000
100000
10000
1000
100
10
1980
1990
2000
2010
2020
2030
LHC Phase 2
LHC Phase 1
ATLAS/CMS
LHC 2012
BELLE
D0
Babar
CDF
JLAB
H1
RHIC
NA48
E791
E831
ALEPH
E665
E691
BC72
DPHEP
Scientific potential
Outreach, Training, Education
Arxiv: 1205.4667
nature
International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue
Archive | Volume 503 | Issue 7477 | News | Article
NATURE | NEWS
عربي
LHC plans for open data future
Researchers share results to keep them accessible.
Elizabeth Gibney
26 November 2013
"When the LHC programme comes to an end, it will probably be the last data at this frontier for many years. We can't afford to lose it."

2018 status

DPHEP timelines

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
	Start-up					Consolidation			DPHEP Collaboration		
HEP	HERA stops	Babar stops	LHC starts	Belle I stops	Tevatron stops				LHC Run 2		
DPHEP Group			ICFA Panel		LHC exp. joined	DPHEP Manger appointed at CERN		DPHEP Collaboration Agreements signed	1st DPHEP Collaboration Meeting		2nd DPHEP Collaboration Meeting
DPHEP Docs			DPHEP White Paper			Blueprint Report			DPHEP Status Report 2020 Vision		DPHEP 2017 Status Report
DP Projects within expts.		Babar DP starts		HERA DP starts	BELLE DP starts	CMS DP Policy CDF/D0 DP starts Babar LTDAP operational	ALICE, LHCb, DP Policies	ATLAS DP Policy H1/ZEUS DP systems operational	CERN/LHC Open Data	CERN/LHC Analysis Preservation Tevatron DP operational	