**DPHEP-2015-001**
**August 2015**

# Status Report of the DPHEP Collaboration: A Global Effort for Sustainable Data Preservation in High Energy Physics

www.dphep.org

**Abstract**

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. An inter-experimental study group on HEP data preservation and long-term analysis was convened as a panel of the International Committee for Future Accelerators (ICFA). The group was formed by large collider-based experiments and investigated the technical and organisational aspects of HEP data preservation. An intermediate report was released in November 2009 addressing the general issues of data preservation in HEP and an extended blmueprint paper was published in 2012. In July 2014 the DPHEP collaboration was formed as a result of the CA signature by seven large funding agencies. This paper presents the status report of the DPHEP collaboration and elaborates on the progress registered during the past 3 years.

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

**The DPHEP Study Group**

Zaven Akopov, Deutsches Elektronen Synchrotron, DESY, Hamburg
Silvia Amerio, INFN Padova, Italy
David Asner, Pacific Northwest National Laboratory, USA
Eduard Avetisyan, Deutsches Elektronen Synchrotron, DESY, Hamburg
Olof Barring, CERN, Geneva, Switzerland
James Beacham, New York University, USA
Matthew Bellis, Stanford University, USA
Gregorio Bernardi, LPNHE, University Paris VI
Siegfried Bethke, MPI for Physics, Munich, Germany
Amber Boehnlein, Stanford Linear Accelerator Laboratory, SLAC, USA
Travis Brooks, Stanford Linear Accelerator Laboratory, SLAC, USA
Thomas Browder, Hawaii University, USA
Rene Brun, CERN, Geneva, Switzerland
Concetta Cartaro, Stanford Linear Accelerator Laboratory, SLAC, USA
Marco Cattaneo, CERN, Geneva, Switzerland
Gang Chen, IHEP, Beijing, China
David Corney, Rutherford Appleton Laboratory, STFC, UK
Kyle Cranmer, New York University, USA
Ray Culbertson, Fermi National Laboratory, FNAL, USA
Sunje Dallmeier-Tiessen , CERN, Geneva, Switzerland
Dmitri Denisov, Fermi National Laboratory, FNAL, USA
Cristinel Diaconu, CPPM, Aix-Marseille University, France and DESY, Germany
Vitaliy Dodonov, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Tony Doyle, University of Glasgow, SUPA, UK and CERN, Geneva, Switzerland
Gregory Dubois-Felsmann, Stanford Linear Accelerator Laboratory, SLAC, USA
Michael Ernst, Brookhaven National Laboratory, BNL, USA
Martin Gasthuber, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Achim Geiser, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Fabiola Gianotti, CERN, Geneva, Switzerland
Paolo Giubellino, CERN, Geneva, Switzerland
Andrey Golutvin, CERN, Geneva, Switzerland
John Gordon, Rutherford Appleton Laboratory, STFC, UK
Volker Guelzow, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Takanori Hara, KEK, Tsukuba, Japan
Hisaki Hayashii, Nara Women's University, Japan
Andreas Heiss, Karlsruhe Institute for Technology, Germany
Frederic Hemmer, CERN, Geneva, Switzerland
Fabio Hernandez, Computing Centre IN2P3, France
Graham Heyes, Jefferson National Laboratory, JLAB, USA
Andre Holzner, CERN, Geneva, Switzerland
Peter Igo-Kemenes, CERN, Geneva, Switzerland
Toru Iijima, Nagoya University, Japan
Joe Incandela, University of California Santa Barbara, USA and CERN, Geneva, Switzerland
Roger Jones, CERN, Geneva, Switzerland
Yves Kemp, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Kerstin Kleese van Dam, Pacific Northwest National Laboratory, PNNL, USA
Juergen Knobloch, CERN, Geneva, Switzerland
David Kreincik, Cornell University, USA
Kati Lassila-Perini, Helsinki Institute of Physics, Finland
Francois Le Diberder, Laboratoire de l'Accelerateur Lineaire, LAL/IN2P3, Orsay,, France
Sergey Levonian, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Aharon Levy, Tel Aviv University, Israel
Qizhong Li, Fermi National Laboratory, FNAL, USA
Bogdan Lobodzinski, Deutsches Elektronen Synchrotron, DESY, Hamburg

Marcello Maggi, INFN Bari, Italy
Janusz Malka, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Salvatore Mele, CERN, Geneva, Switzerland
Richard Mount, Stanford Linear Accelerator Laboratory, SLAC, USA
Homer Neal, Stanford Linear Accelerator Laboratory, SLAC, USA
Jan Olsson, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Dmitri Ozerov, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Leo Piilonen, Virginia Polytechnic Institute, VPI, USA
Giovanni Punzi, Fermi National Laboratory, FNAL, USA
Kevin Regimbal, Pacific Northwest National Laboratory, PNNL, USA
Daniel Riley, Cornell University, USA
Michael Roney, University of Victoria, Canada
Robert Roser, Fermi National Laboratory, FNAL, USA
Thomas Ruf, CERN, Geneva, Switzerland
Yoshihide Sakai, KEK, Tsukuba, Japan
Takashi Sasaki, KEK, Tsukuba, Japan
Gunar Schnell, University of the Basque Country UPV/EHU, Bilbao, Spain
Matthias Schroeder, CERN, Geneva, Switzerland
Yves Schutz, Subatech IN2P3, Nantes, France
Jamie Shiers, CERN, Geneva, Switzerland
Tim Smith, CERN, Geneva, Switzerland
Rick Snider, Fermi National Laboratory, FNAL, USA
David M. South, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Rick St. Denis, University of Glasgow, UK
Michael Steder, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Jos Van Wezel, Karlsruhe Institute for Technology, Germany
Erich Varnes, University of Arizona, USA
Margaret Votava, Fermi National Laboratory, FNAL, USA
Yifang Wang, IHEP, Beijing, China
Dennis Weygand, Jefferson National Laboratory, JLAB, USA
Vicky White, Fermi National Laboratory, FNAL, USA
Katarzyna Wichmann, Deutsches Elektronen Synchrotron, DESY, Hamburg, Germany
Stephen Wolbers, Fermi National Laboratory, FNAL, USA
Masanori Yamauchi, KEK, Tsukuba, Japan
Itay Yavin, McMaster University and Perimeter Institute, Canada
Hans von der Schmitt, MPI Munich, Germany

# Executive Summary

**TO UPDATE IT FOR THE ICFA MEETING IN VIENA??**

Main messages:
The emergency to invest in data preservation remains, the problem is not solved , though a lot of progress has been achieved

-Preamble, history, DPHEP collaboration, etc

-Former experiments: mature preservation programs, several paradigms tested, publication flux clearly profited from these projects, local/regional synergies

-LHC: great progress and great initiatives

-Interdisciplinary contacts:

-Technologies

-Tendencies and evolutions in DP and in HEP

# Content

1. Reports from labs and experiments
    a. (we will start by collecting these, 4-6 pages from each contribution, also invited remotely, if necessary)

2. Physics case overview
    a. the research we have gained, the added value

3. Review of DP models
    a. the experience we have gained todo's and nottodo's

4. Review of DP-related technologies
    a. and how they have been used in the past years

5. The resources and costs for DP, experience

6. A critical review of the DPHEP evolutions, the DPHEP Roadmap, The potential for common projects:

The BLUEPRINT content table (reminder)

**EXECUTIVE SUMMARY ..... 4**

**1. INTRODUCTION.....** ERROR! BOOKMARK NOT DEFINED.

**1.1 The Context for Data Preservation in HEP.....**Error! Bookmark not defined.

**1.2 Data preservation in other disciplines.....**Error! Bookmark not defined.

**1.3 A study group on data preservation in HEP** ..... Error! Bookmark not defined.

**2. THE SCIENTIFIC POTENTIAL OF DATA PRESERVATION IN HIGH ENERGY PHYSICS** ..... ERROR! BOOKMARK NOT DEFINED.

**2.1 Physics Potential of Former Experiments** ..... Error! Bookmark not defined.

**2.2 Long-term Physics Potential of Experiments in the Final Analysis Phase** Error! Bookmark not defined.

**2.3 Generic Classification of the Physics Potential of Preserved Data** ..... Error! Bookmark not defined.

**3. MODELS OF HEP DATA PRESERVATION** ..... ERROR! BOOKMARK NOT DEFINED.

**3.1 Data Preservation Levels** ..... Error! Bookmark not defined.

**3.2 Preparing a Data Preservation Project in HEP** ..... Error! Bookmark not defined.

**3.3 Hardware Requirements and Computing Centre Issues** ..... Error! Bookmark not defined.

**4. EXPERIMENTAL LEVEL AND LABORATORY LEVEL STRATEGIES.** ERROR! BOOKMARK NOT DEFINED.

**4.1 The Data Preservation Programme at BaBar** ..... Error! Bookmark not defined.

**4.2 The Data Preservation Programme at H1** ..... Error! Bookmark not defined.

**4.3 The Data Preservation Programme at ZEUS** ..... Error! Bookmark not defined.

**4.4 The Data Preservation Programme at HERMES** ..... Error! Bookmark not defined.

**4.5 The Data Preservation Programme at Belle** ..... Error! Bookmark not defined.

**4.6 The Data Preservation Programme at BESIII** ..... Error! Bookmark not defined.

**4.7 The Data Preservation Programme at CDF** ..... Error! Bookmark not defined.

**4.8 The Data Preservation Programme at DØ** ..... Error! Bookmark not defined.

**4.9 Data Preservation and the LHC Experiments** ..... Error! Bookmark not defined.

**5. DPHEP COMMON PROJECTS** ..... ERROR! BOOKMARK NOT DEFINED.

**5.1 A Generic Validation Framework for HEP Experiments** ..... Error! Bookmark not defined.

**5.2 A Common Project on Archival Systems** ..... Error! Bookmark not defined.

**5.3 The RECAST Framework** ..... Error! Bookmark not defined.

**5.4 Preservation of Documentation and High-level Objects** ..... Error! Bookmark not defined.

**5.5 Outreach** ..... Error! Bookmark not defined.

**6. A GLOBAL ORGANISATION FOR DATA PRESERVATION IN HEP** ERROR! BOOKMARK NOT DEFINED.

**7. OUTLOOK.....** ERROR! BOOKMARK NOT DEFINED.

**APPENDIX A: COMMITTEES OF THE DPHEP STUDY GROUP.** ERROR! BOOKMARK NOT DEFINED.

**APPENDIX B: WORKSHOPS OF THE DPHEP STUDY GROUP..** ERROR! BOOKMARK NOT DEFINED.