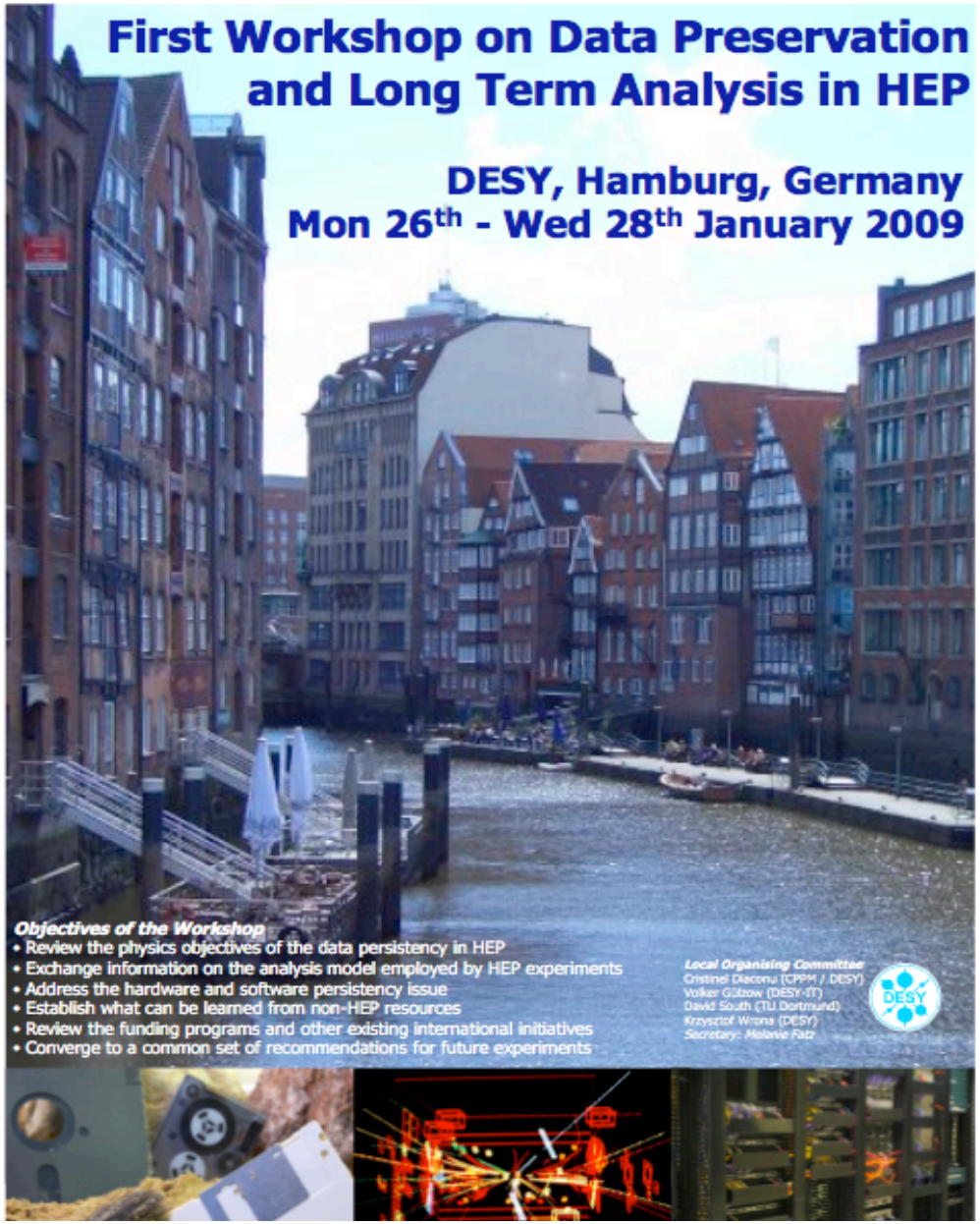
January 26, 2009

# Welcome to the




Cristinel Diaconu
CPP Marseille and DESY

# The context

Digital Preservation is a known issue

"the digital black hole", SMOP, SMOR, economic models etc.

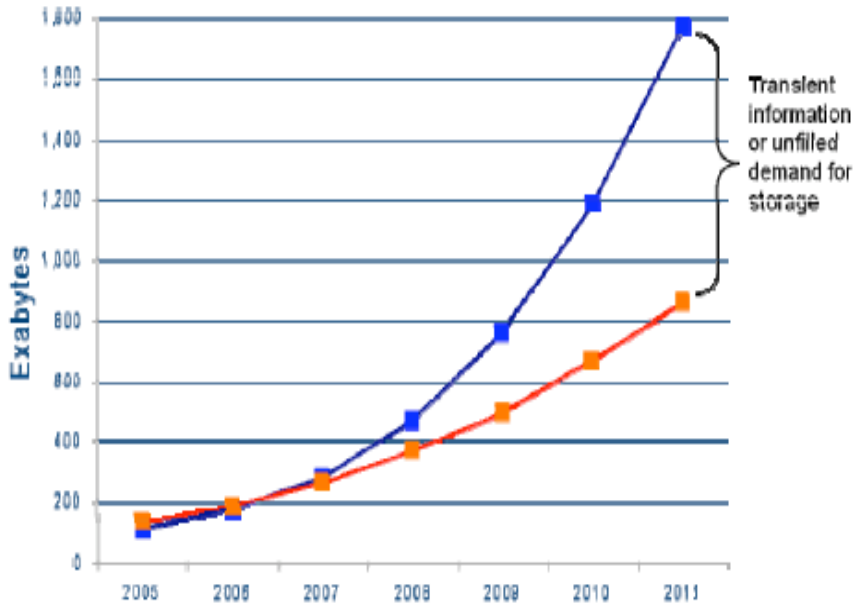
Task forces already in place to address this issue in a generic way (e.g. Blue Ribbon, APA, DPC,eSciDir…)

**http://brtf.sdsc.edu (intermediate report)**

**http://www.alliancepermanentaccess.eu**

Scientific Data is a major component of the ongoing efforts

**Data Preservation in HEP?**



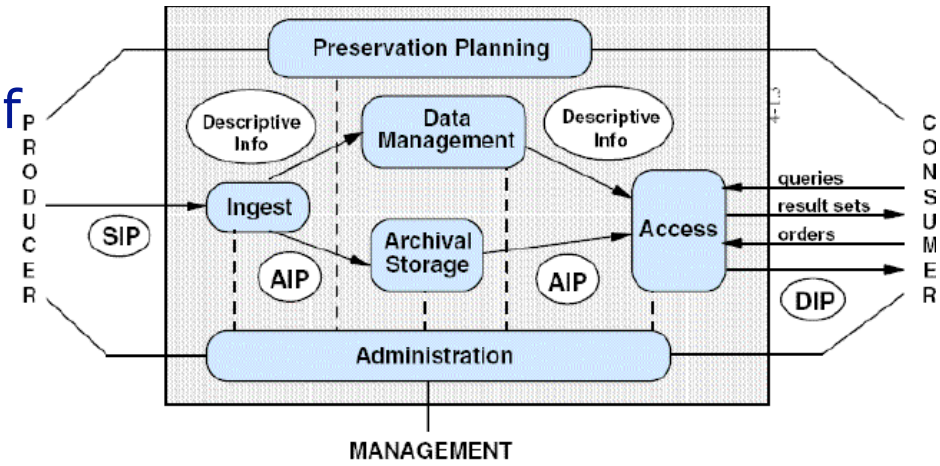



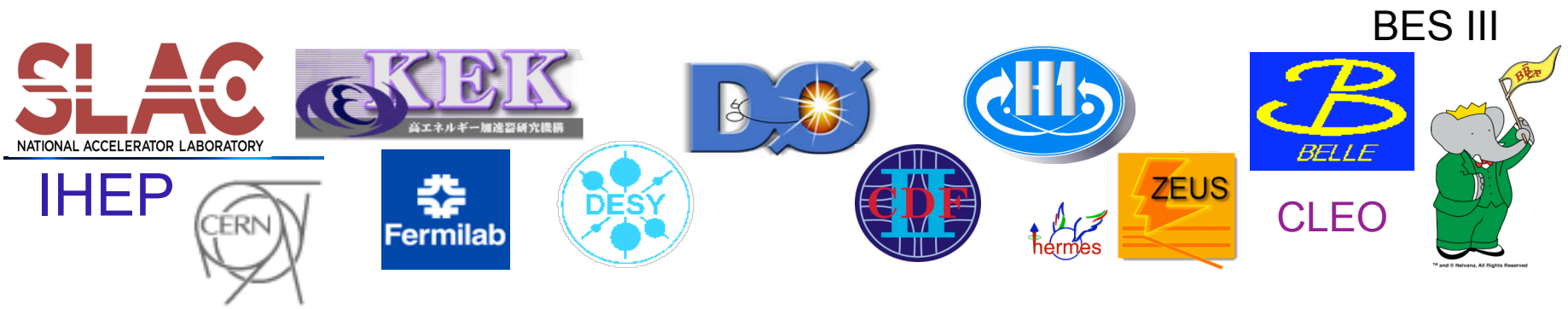

FIGURE 2.1: **The OAIS Reference Model**
http://public.ccsds.org/publications/archive/650x0b1.pdf, Page 4-1.
*Source: Consultative Committee for Space Data Systems January 2002.*

# Data Persistency and Long Term Analysis in HEP as of 2009

- Several large HEP experiments have large/full data sets and head towards final analyses
  - HERA: H1,ZEUS,HERMES (end of collisions July 2007)
  - Tevatron: CDF,D0 (end of collisions in 2-3 years)
  - PEP III: BaBar (end of collisions March 2008)
  - KEKB: Belle (2009 + upgrade)
  - CLEO: end 2008
  - BES III: start 2008
- The complexity of data sets is similar, driven by technological choices (trigger) convoluted with the complexity of the measured final states
- The conservation of the data and possibility of long term analysis are important issues

# An Inter-Experiment Study Group

- Common enterprise of experiments and associated computing centres
  - Experiments H1/ZEUS/Hermes/CDF/D0/Babar/Belle/BES/CLEO and computing centers from DESY/SLAC/KEK/FNAL/IHEP/CERN
  - Contacts with DOE, UE-FP7(PARSE)
  - Do not try to cover everything as "use case" (keep the focus)



# An Inter-Experiment Study Group

**Composition: HEP scientists and IT experts**

Representatives of the commited HEP experiments and Computing Centers

Other experts and contact persons

Preparatory discussions started (first phone discussion end of September 2008):

- First (quick) exchange on data status and typology
- Converged on a set of objectives

Aspects:

- Technology (are we ready for the "long term"?)
- Organizational matters (including the economical model)
- Intellectual property and open access matters

# Committees

## **International Steering Committee**:

Supervise the report, make sure that the recommandations are aligned with real life

Spokespersons of experiments

Directors of the computing centers

DESY-IT: Volker Gülzow (DESY)
H1: Cristinel Diaconu (CPPM/DESY)
ZEUS: Tobias Haas (DESY)
FNAL/DoE: Amber Boehnlein (DoE)
FNAL-IT: Victoria White (FNAL)
D0: Dmitri Denisov (FNAL), Darien Wood (FNAL)
CDF: Jacobo Konigsberg (FNAL), Robert Roser (FNAL)
IHEP-IT: Gang Chen (IHEP)
BES III: Yifang Wang (IHEP)
KEK-IT: Takashi Sasaki (KEK)
Belle: Masanori Yamauchi (KEK), Tom Browder (Hawaii)
SLAC-IT: Richard Mount (SLAC)
BaBar: Francois Le Diberder (LAL/SLAC)
CERN-IT: Frederic Hemmer (CERN)
CERN/PARSE: Salvatore Mele (CERN)
CLEO: David Asner (Carleton)

## **International Advisory Committee:** [In the process of being defined]

Analyse the report and endorse it

Promote the report in HEP centers and instances (e.g. ICFA)

Chaired by Jonathan Dorfan (SLAC) and Siggi Bethke (MPI Munich)

# The Objectives

- Review and document the physics objectives of the data persistency in HEP.
- Exchange information concerning the analysis model: abstraction, software, documentation etc. and identify coherence points
- Address the hardware and software persistency status
- Get some external (non-HEP, non-research etc.) input
- Review possible funding programs and other related international initiatives
- Converge to a common set of specifications in a document (blueprint) that will constitute the basis for future collaborations

# Steps Towards Long Term Analysis in HEP

1) Experiment-wise preparation/organisation for proper conservation of the data/knowledge
   - Proper planning and (new) projects required **(hot topic!)**
2) Common framework for similar experiments
   - Similar experiments converge on data release policy/format
   - Enable (further) combined analyses
3) Open access to expert community
   - Require sufficient knowledge encapsulation, is a natural and necessary result of the previous steps.
4) Open access to a wider community:
   - educational projects, outreach etc.a

**Steps 2-4 imply a policy for open access to the HEP data (status?)**

# Workshops of the Study Group

- First Workshop organized in DESY (January 26-28, 2009)
  - http://indico.cern.ch/conferenceDisplay.py?confId=42722
  - Prepared on December 3rd: agenda, template-talk for experiments
- Exploratory, but with clear agenda/objectives
  - Initiate further work/working groups on technical and organizational matters
  - Define the editorial procedure for the document (blue print) for data preservation and long term analysis in HEP.
- A second workshop mid-2009 (SLAC)
  - Objective: prepare a status report/document for LP2009
- A path for further collaborations/meetings is expected to occur

# The Plan

One day to get into the business

- -we want to see what the size of the problem is
- -what is the status in the field

One day to put our problem in the context

- -past experiences
- -further ideas, new technology
- -other fields, programs and funding

One day to summarise, gather ideeas, prepare the Study Group for further work

- -start sub-groups and define future directions
- -draft the plan for the blue-print

Minutes: Andre Holzner
Proceedings: to be defined

# The Agenda

Robert Roser, Homer Neal

**Monday, 26 January 2009**

| | |
|---|---|
| 09:00 | [32] **Welcome and Introduction to the Workshop**<br>(Seminar 4 (EVO): 09:00 - 09:20) |
| | [0] **H1 Analysis and Computing Model**<br>by David SOUTH (Technische Universität Dortmund)<br>(Seminar 4 (EVO): 09:20 - 09:40) |
| | [1] **ZEUS Analysis and Computing Model**<br>by Janusz SZUBA (DESY)<br>(Seminar 4 (EVO): 09:40 - 10:00) |
| 10:00 | [2] **CDF Analysis and Computing Model**<br>by Robert ROSER (Fermilab)<br>(Seminar 4 (EVO): 10:00 - 10:20) |
| | [3] **D0 Analysis and Computing Model**<br>by Qizhong LI (Fermilab)<br>(Seminar 4 (EVO): 10:20 - 10:40) |
| | Coffee<br>(10:40 - 11:00) |
| 11:00 | [6] **Belle Analysis and Computing Model (via EVO)**<br>by Prof. Nobuhiko KATAYAMA (HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION)<br>(Seminar 4 (EVO): 11:00 - 11:20) |
| | [7] **BES-III Analysis and Computing Model (via EVO)**<br>by Dr. Gongxing SUN (INSTITUE OF HIGH ENERGY PHYSICS)<br>(Seminar 4 (EVO): 11:20 - 11:40) |
| | [5] **BaBar Analysis and Computing Model**<br>by Homer NEAL (Physics Department)<br>(Seminar 4 (EVO): 11:40 - 12:00) |
| 12:00 | [28] **CLEO Analysis and Computing Model**<br>by Daniel RILEY<br>(Seminar 4 (EVO): 12:00 - 12:20) |

Lunches: we have a reserved area in the cantine

# Monday Afternoon

Volker Guelzow, Frederic Hemmer

| Monday, 26 January 2009 | | | |
|---|---|---|---|
| 14:00 | [8] **A Multi-purpose Computing Centre: CC-IN2P3 (Lyon, France)**<br>by Mr. Fabio HERNANDEZ (IN2P3/CNRS Computing Centre)<br>(Seminar 4 (EVO): 14:00 - 14:25) | | |
| | [9] **A Multi-purpose Computing Centre: FNAL (USA)**<br>by Stephen WOLBERS (Fermilab Computing Division)<br>(Seminar 4 (EVO): 14:25 - 14:50) | | |
| 15:00 | [34] **A Distributed Computing Centre (NDGF, Nordics)**<br>by Erik Mattias WADENSTEIN<br>(Seminar 4 (EVO): 14:50 - 15:15) | | |
| | Coffee<br>(15:15 - 15:45) | | |
| 16:00 | [14] **Storage Systems: Status and Perspectives**<br>by Mr. Martin GASTHUBER (DESY); Volker GUELZOW<br>(Seminar 4 (EVO): 15:45 - 16:10) | | |
| | [13] **Virtualisation**<br>by Yves KEMP<br>(Seminar 4 (EVO): 16:10 - 16:35) | | |
| | [38] **Discussion Experiments/Computing Centers**<br>(Seminar 4 (EVO): 16:35 - 16:55) | | |
| 17:00 | **Workshop Discussions: ee Experiments (Parallel)**<br>(Seminar 4 (EVO): 17:00 - 18:00) | **Workshop Discussions: ep Experiments (Parallel)**<br>(Seminar 3a: 17:00 - 18:00) | **Workshop Discussions: pp Experiments (Parallel)**<br>(Seminar 5: 17:00 - 18:00) |
| 18:00 | | | |

Common Dinner in DESY Bistro

# Tuesday Morning

Tobias Haas, Takashi Sasaki

**Tuesday, 27 January 2009**

09:00

[11] **Experience from re-analysis of LEP and PETRA Data**
by Siegfried BETHKE (Max-Planck-Institut fur Physik)
(Seminar 4 (EVO): 09:30 - 09:55)

10:00

[30] **Experience from the LEP Higgs Working Group**
by Peter IGO-KEMENES (Physikalisches Institut)
(Seminar 4 (EVO): 09:55 - 10:20)

[12] **Data Conservation at LEP**
by Andre Georg HOLZNER (Eidgenossische Technische Hochschule Zurich/ETH (ETH))
(Seminar 4 (EVO): 10:20 - 10:45)

Coffee
(10:45 - 11:15)

11:00

[29] **Management of Astronomical Data Archives and their Interoperability through the Virtual Observatory Standards**
by Prof. Fabio PASIAN (INAF, Trieste)
(Seminar 4 (EVO): 11:15 - 11:40)

[10] **Challenges in Long Term Computing Models and ROOT Solutions**
by Rene BRUN
(Seminar 4 (EVO): 11:40 - 12:15)

12:00

Photograph
(12:15 - 12:30)

# Tuesday Afternoon

Siggi Bethke, Richard Mount

**Tuesday, 27 January 2009**

| | |
|---|---|
| 14:00 | [15] **The HEP Survey for Long Term Data Persistency and Open Access**<br>by Dr. Salvatore MELE (CERN)<br>(Seminar 4 (EVO): 14:00 - 14:25) |
| | [18] **EU/FP7 Policies and Programmes**<br>by Dr. Salvatore MELE (CERN)<br>(Seminar 4 (EVO): 14:25 - 14:50) |
| 15:00 | [17] **US/DoE Policies and Programmes**<br>by Dr. Amber BOEHNLEIN (FERMI NATIONAL ACCELERATOR LABORATORY)<br>(Seminar 4 (EVO): 14:50 - 15:15) |
| | Coffee<br>(15:15 - 15:45) |
| 16:00 | [37] **UK/STFC Policies and Programmes**<br>by David CORNEY (STFC)<br>(Seminar 4 (EVO): 15:45 - 16:10) |
| | [19] **Report from SPIRES**<br>by Travis BROOKS (SLAC)<br>(Seminar 4 (EVO): 16:10 - 16:35) |
| 17:00 | [39] **Discussion Long Long Term Governance, Models, Open Access**<br>(Seminar 4 (EVO): 16:35 - 17:05) |

Bus leaves at 18h30, Museum visit, Dinner 20h00

**Working Dinner**
(Altonaer Museum: 18:30 - 23:00)

# Wednesday

| Wednesday, 28 January 2009 | | |
|---|---|---|
| 09:00 | [25] **Report from ee Experiments**<br>(Seminar 4 (EVO): 09:30 - 09:50) | Homer Neal, Takashi Sasaki |
| 10:00 | [26] **Report from ep Experiments**<br>(Seminar 4 (EVO): 09:50 - 10:10) | David South, Janusz Szuba |
| | [27] **Report from pp Experiments**<br>(Seminar 4 (EVO): 10:10 - 10:30) | Qizhong Li, Robert Roser |
| | Coffee<br>(10:30 - 11:00) | |
| 11:00 | [21] **Scenarios for Long Term Analysis (Summary)**<br>(Seminar 4 (EVO): 11:00 - 11:20) | Stephen Wolbers |
| | [22] **Working Directions (Discussion)**<br>(Seminar 4 (EVO): 11:20 - 11:40) | Homer Neal |
| | [23] **Blueprint Plans (Discussion)**<br>(Seminar 4 (EVO): 11:40 - 12:00) | Cristinel Diaconu |
| 12:00 | [24] **Next Workshop**<br>(Seminar 4 (EVO): 12:00 - 12:15) | Richard Mount |

Summary talks collecting the concrete data and proposals,
some “volunteers” already (preliminarily) assigned