# Blueprint

- Start the production of a detailed document on data preservation
  - Gets in details of the individual projects
  - Proposes working directions
  - Define the organization
- Destination: funding requests
  - For individual need (Babar from SLAC, HERA from DESY, Tevatron from DOE/Fermilab)
  - For common projects
- Method: split in chapters, break in subgroups, get together for common sessions
  - Leave CERN with version 0.2

Blueprint of
Data Preservation in High-Energy Physics

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

http://dphep.org
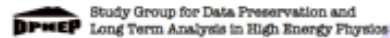
Abstract

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. At the same time, HEP has no coherent strategy for data preservation and re-use. An proposes an International Organization devoted to the data preservation in high-energy physics. The organization is structured around an inter-experimental Study Group supervised by the International Committee for future Accelerators (ICFA). The present document presents the motivation for such an organization, collects example of initiatives at experiment level, defines common R&D projects and draw the main lines of the

Content

- Chapter 1: Executive Summary and General remarks

    - (ICFA document)

- Chapter 2: Physics Motivation: detailed cases, simulations

- Chapter 3: Experiments archival models

- Chapter 4: DPHEP Common projects

- Chapter 5: DPHEP Organization

Chapter 2: The Scientific Potential of the Data Preservation in High Energy Physics

- Prospects for the utility of the whole enterprise

- Past: JADE+DIS+find others.

- Examples of "Gedanken experiments" on archived data
  - I.e. where are the non-experimental weak points
  - Quantify what it needs to re-do the analysis
  - What the result would look like if some errors are reduced

- Give 2-3 striking examples from each experiment:
  - A Measurement (strong coupling, branching ratio, jet cross sections)
  - A search for new physics
  - A complementary analysis or a combination with a future/existing project

- A summary of policies/practices in other fields

- Examples
  - Precision analyses: treat sensitive cases
  - ALEPH list of "embargo" analyses
  - Investigations of discrepancies (Nutev 3sigma, re-analyses)
  - Fixed target fermilab
  - Separate Preservation from Access
  - Examples of LASS k-beam
  - Hadronic contribution to g-2 (Babar)
  - 35 paper (open) for the archival
  - 10-12 from H1
  - Tevatron b-physics area
  - LEP examples (long tail)
  - Analyses enabled by full precision/data sets/reprocessing
  - HERA: jet analyses/alpha_s/instantons
  - Babar example of a flagship measurement that is bound to happen in the archival mode

Chapter 3: Experiments DP projects
A: project/hardware/resources

- The present computing model
  - Description of data model
  - Event size
  - Data sample: events, size
  - MC simulation: size, production
  - Reprocessing strategy
  - Analysis resources
  - Team/Person power
- The archival system (description)
  - Level (1 to 4)
  - Description of the transition steps
  - Timescales
  - Lifetime?
  - Virtualisation

Chapter 3: Experiments DP projects
B: governance, international scene
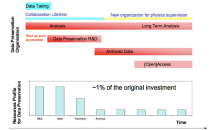
- Organisation of the Physics Supervision (connection to DPHEP)
  - Membership, transition, timelines
  - Functions of the "governance body": supervise the archivists activity
  - Access
  - Mandate from the collaboration on data changes, archival system evolution
  - Selection committee for the archivist
  - Interface with the host laboratory: technical support
- Guidelines for information publishing/disclosure
  - What will be really for public consumption (comments in notes/documentation)
  - Consistence across collaborations? Across time?
- Costs/benefits statement
  - Physics potential versus archival costs
- Interface to DPHEP

– Validation Suites?

– Super-structure for archival systems?

- DPHEP-RD2: Outreach

  – Standard formats, tools, communication techniques

- DPHEP-RD3: Extension of the public documentation

  – Recovery of paper material (with libraries/archives)

  – "Internal" notes (search, understand, attach)

  – (t)wikis (dark archive, embargo, snapshot, choice of material, structuring)

  – Talks (ingest, classify, search)

  – "private" additional information paper-by-paper

  – Figures, Ntuples, mathematica, root, xml for outreach… (who does the stewardship?)

- DPHEP-RD4: Standards

  – Procedures? Part of archivist job? Common formats?

  – Start from the outreach and work backwards (testbed)

  – What can DPHEP learn from the way ASTRO does it (cost/benefit invest now and preserve with less cost). Part of funding surge.

Chapter 5: DPHEP

- ## Modus Operandi
  - management definitions, meetings, reports

- ## Connections levels
  - Experiments

    Supervisions committees => DPHEP
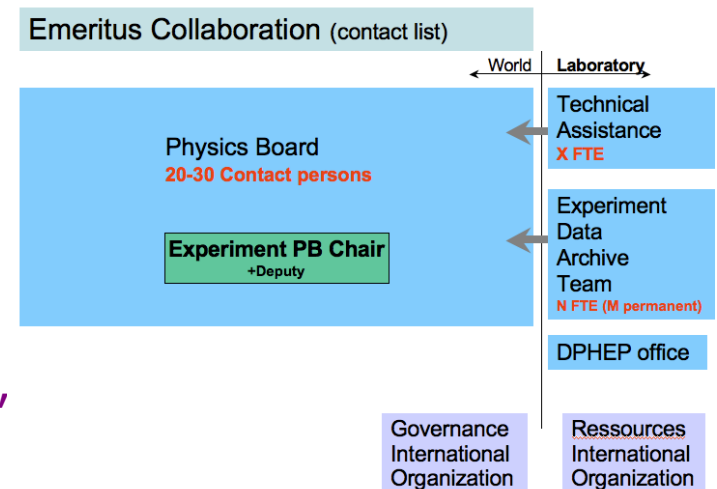 - Laboratories

    Data Archive Team=> DPHEP
  - Connections from external HEP contributors
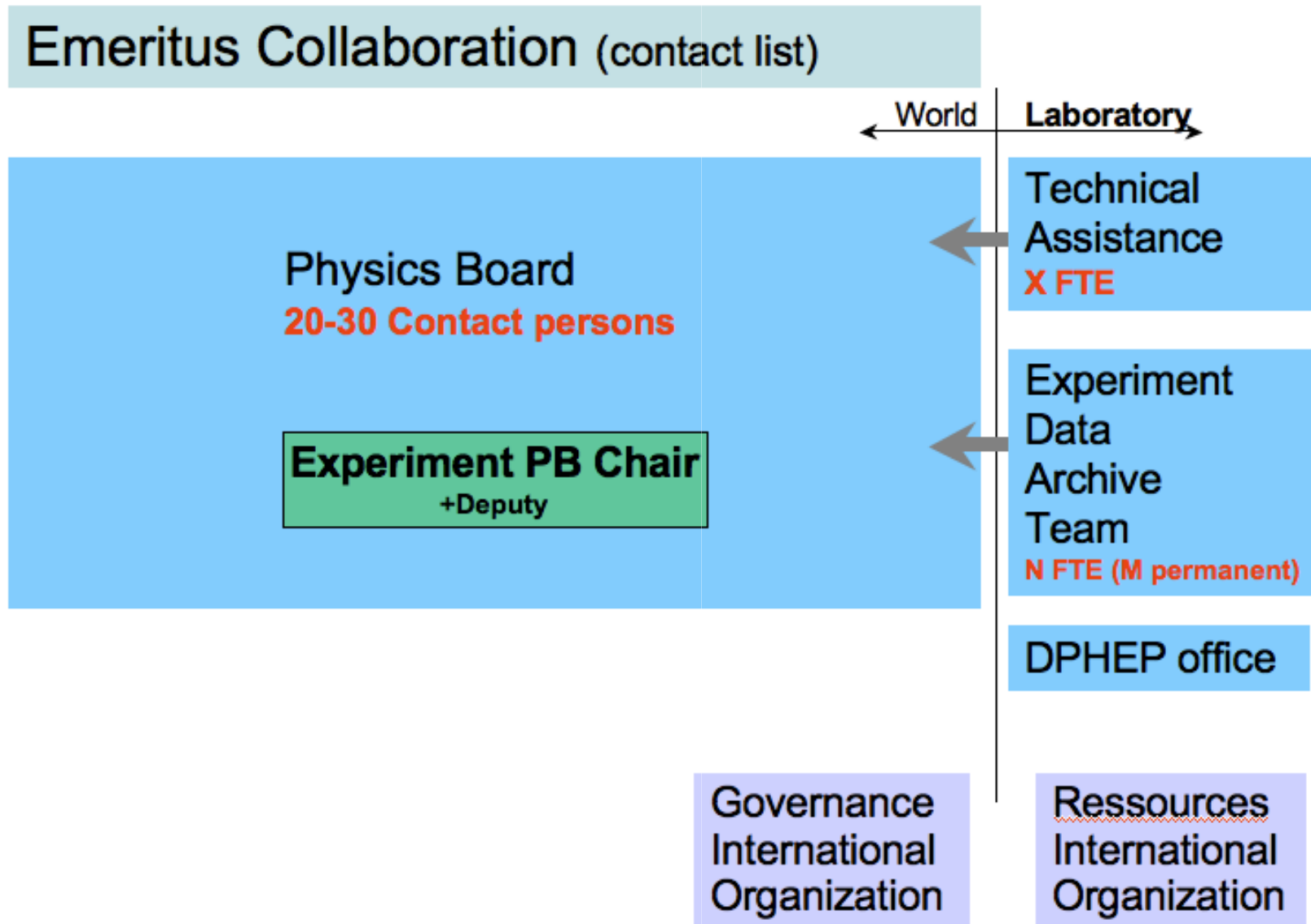
  - Connections to other fields

    NB: IVOA HEP contact person: Fabio Pasian, connect more organizations?)

A long term organization of HEP experiments

A long term organization of HEP experiments

Emeritus Collaboration (contact list)

World | **Laboratory**

Physics Board
20-30 Contact persons

Experiment PB Chair
+Deputy

Technical Assistance
X FTE

Experiment Data Archive Team
N FTE (M permanent)

DPHEP office

Governance International Organization

Ressources International Organization

Chapter 5: DPHEP Organization

- Define/propose the funding model
  - Possible contributors
  - Investigate funding programs

    USA

    Europe