

YEARS / ANS **CERN**

Open Access, Data, Knowledge

Policies, Strategies & Implementation



International Collaboration for Data Preservation and Long Term Analysis in High Energy Physics



Jamie.Shiers@cern.ch

RECODE - Final Workshop -
January 2015

Overview

- Setting the scene – key characteristics of CERN & High Energy Physics (HEP)
- Some of the parameters – data volumes, Use Cases, duration & lifetimes (**decades+**)
- **Towards Strategies & Implementation(s)**



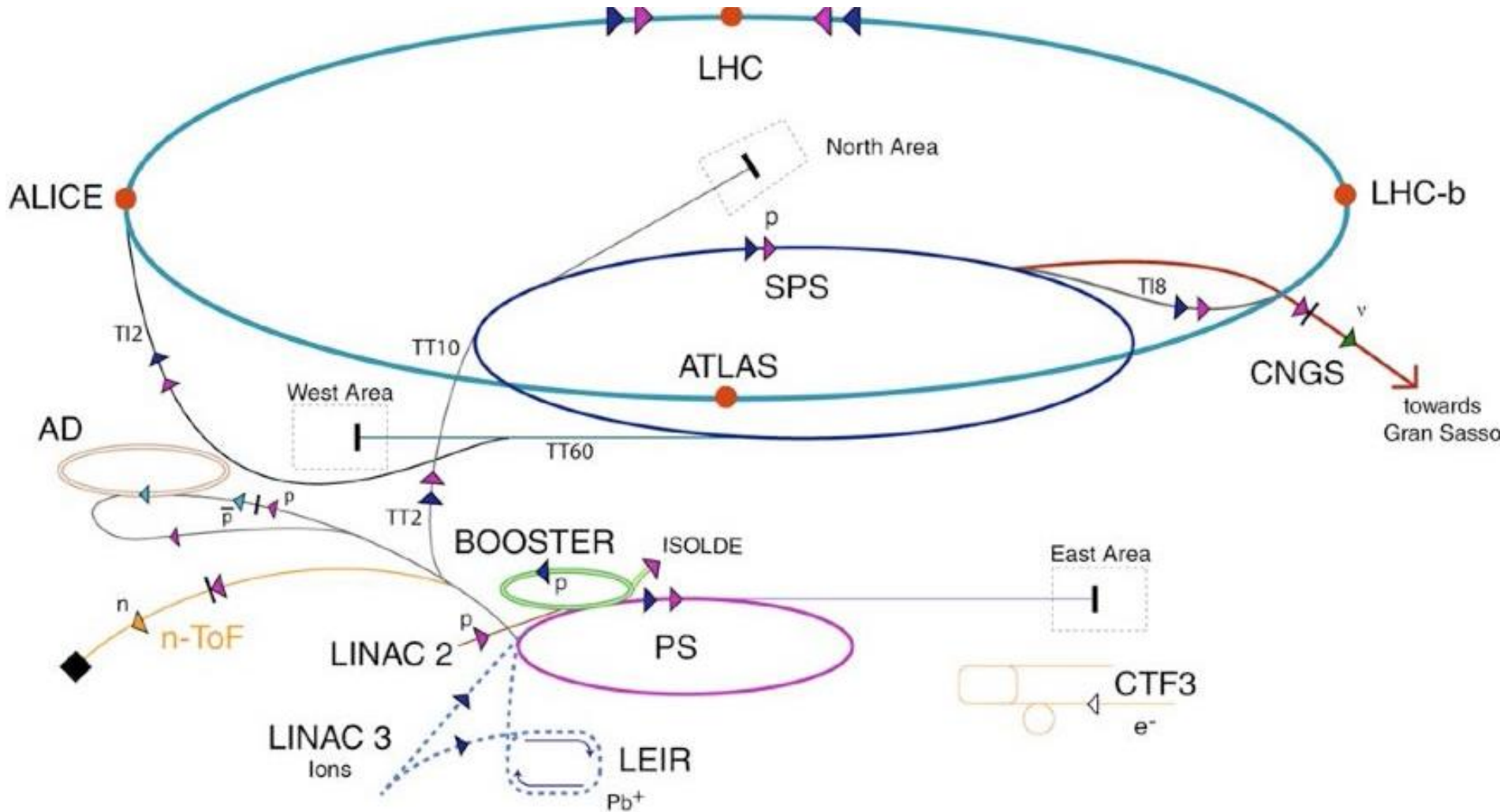


Accelerating Science and Innovation

Energy Frontier

LHC

CERN Accelerators Today



The First Accelerator Arrives...



Decommissioned Accelerators

- The Intersecting Storage Rings (ISR)
 - World's First Hadron Collider
 - Proton-proton and Proton-anti-proton
 - De-commissioned prior to LEP
 - **Re-used as a Storage Tunnel (tapes etc!)**
 - Paved the way to future hadron colliders: Z / W / Higgs!
- Large Electron-Positron Collider (LEP)
 - Sited in the same tunnel as LHC
 - **Use Case(s) for Data Preservation and Open Access**



What is CERN? (Summary)

- **Far more than just a lab / single accelerator**
- 21(+) member states – many many experiments and Programmes
- Collaboration is key to all activities (accelerators, detectors, experiments)
 - Including Medical Physics (detectors, accelerators)
- Many opinions, diverse policies, ...
- Part of a GLOBAL COMMUNITY with worldwide strategies
 - European Strategy for Particle Physics, US P5 Report, Japan
- **Who OWNS the data? Who is RESPONSIBLE? For WHAT? For HOW LONG? And THEN?**



Use Cases – LHC (and LEP)

1. Preserve data, **software, and know-how** in the collaborations
2. Share data and associated **software** with larger scientific community – **O(PB) in 2020?**
3. Open access to reduced data sets to general public – **O(TB) ?**
4. Bit preservation (**100PB today, 1EB ~2025, 10EB ~2035 – ALREADY “FILTERED”**)
 - Policies:
<http://opendata.cern.ch/collection/data-policies>



<http://opendata.cern.ch/collection/data-policies>

ATLAS Data Access Policy

This document contains the policy document regarding the access to ATLAS data by non-ATLAS members which was endorsed by the ATLAS Collaboration Board in June 2014.

Collection Data-Policies DOI 10.7483/OPENDATA.ATLAS.T9YR.Y7MZ

ALICE data preservation strategy

This document contains the ALICE data preservation strategy and policy.

Collection Data-Policies DOI 10.7483/OPENDATA.ALICE.54NE.X2EA

CMS data preservation, re-use and open access policy

This document describes the CMS collaboration's policy on long-term data preservation, re-use and open access. The policy has been approved by the CMS Collaboration Board in March 2012.

Collection Data-Policies DOI 10.7483/OPENDATA.CMS.UDBF.JKR9

LHCb External Data Access Policy

This document contains the LHCb Data Access Policy. This was adopted at the Collaboration Board meeting on 27th Feb 2013.

Collection Data-Policies DOI 10.7483/OPENDATA.LHCb.HKJW.TWSZ Author Clarke, Peter

<http://opendata.cern.ch/collection/data-policies>

ATLAS Data Access Policy

This document c
endorsed by the

Collection Data-f

Gold Open Access for Publications

Open Access to Specific Data Samples for Outreach

ALICE data

This document c

Collection Data-f

Open Access to (some) Reconstructed data

CMS data p

This document c
policy has been

Collection Data-f

Raw data closed even to collaboration (today)

→ LEP data O(100TB), resources now “trivial”

LHCb Extern

This document c
2013.

Collection Data-Policies

DOI 10.7483/OPENDATA.LHCb.HKJW.TWSZ

Author Clarke, Peter

Data Formats, “Knowledge” etc?

Back to LEP (1989 – 2000)

- Of the 4 experiments, the data for two has been copied externally for LTDP
 - ALEPH: INFN.IT; OPAL: MPI.DE
- In many cases, CERN as an institute is not part of the original collaboration (even though a copy of the bits is kept)
- **IMHO, this is “just fine” – but a policy for all experiments? [All HEP labs??]**



A Possible Strategy...

- As host laboratory, it is expected that (from now on?) a copy of all data acquired by CERN experiments and targeted for long-term preservation be stored in the CERN [certified] digital repository [**Long-term funding?**]
- One or more copies of the above data are maintained outside, at or spread over institutes that form part of the collaboration. [Also certified?]
- **In order to ensure sufficient reliability and adherence to “best practices”, ... such repositories follow agreed guidelines / standards (“certification”)**
- **Feasible for LHC experiments (WLCG), far from clear for the MANY other collaborations**

Open Questions

- Sharing data requires well adapted infrastructure – potential conflicts with “key users” – **additional infrastructure & costs?**
- Open Access is only one step – typically **SIGNIFICANT KNOWLEDGE** is required to usefully process data – regardless of format
 - Software, detector conditions, environment etc.
- How can we work on **COMMON SOLUTIONS** / implementation?





International Collaboration for Data Preservation and Long Term Analysis in High Energy Physics

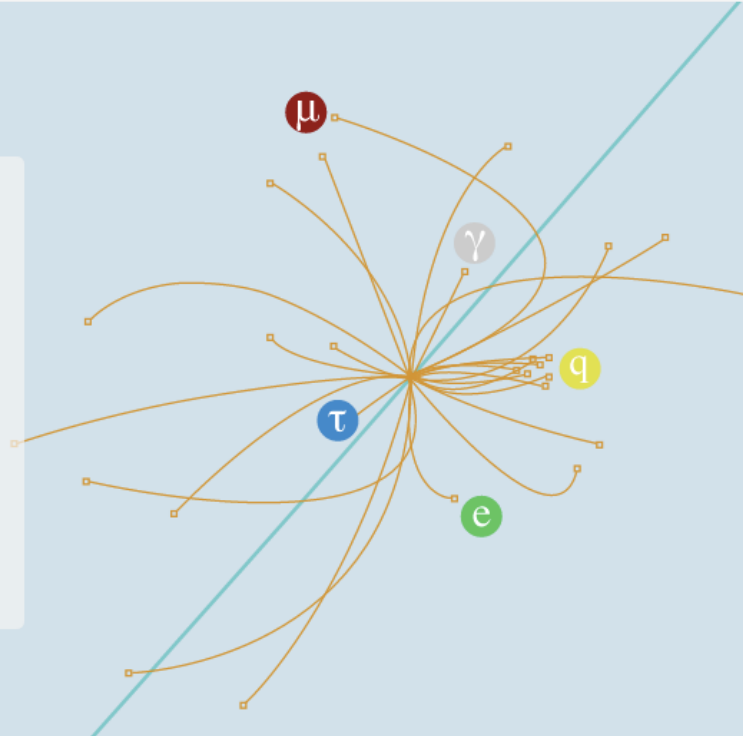
opendata
CERN

ABOUT SEARCH EDUCATION RESEARCH

Education

Visualise events, check reconstructed data, run tools or build your own!

Start learning



Research

Get the genuine working environments, virtual machines and datasets to start your research

Start analysing

