

Workshop Goals

- 1. Provide an update on the changing (or changed) landscape, e.g. FAIR data management (plans), reproducibility, sustainability of data repositories, an update on the status of OAIS and related "standards" and so forth.**
 - Describe the "new world order" wrt "data stewardship / curation / preservation"
- 2. Status reports of the services / developments in the area of LTDP and their outlook**
 - These are now largely production services
- 3. Perform a site-experiment round-table to capture the current situation HEP-wide**



What does DPHEP do?

- DPHEP has become **a Collaboration** with signatures from the main HEP laboratories and some funding agencies **worldwide**.
- It has established a "**2020 vision**", whereby:
 - All archived data – e.g. that described in DPHEP Blueprint, including LHC data – should be easily **findable** and fully usable by the **designated communities** with clear (Open) access policies and possibilities to annotate further;
 - Best practices, tools and services should be well run-in, **fully documented** and **sustainable**; built in common with other disciplines, based on standards;
 - There should be a DPHEP **portal**, through which data / tools accessed;
 - Clear **targets & metrics** to measure the above should be agreed between Funding Agencies, Service Providers and the Experiments.

CERN Services for LTDP

- 1.State-of-the art "**bit preservation**", implementing practices that conform to the ISO 16363 standard
- 2."**Software preservation**" - a key challenge in HEP where the software stacks are both large and complex (and dynamic)
- 3.Analysis **capture and preservation**, corresponding to a set of agreed Use Cases
- 4.Access to **data behind physics publications** - the HEPData portal
- 5.An **Open Data portal** for released subsets of the (currently) LHC data
- 6.A **DPHEP portal** that links also to data preservation efforts at other HEP institutes worldwide.



iPRES 2016

13th International Conference
on Digital Preservation //

Bern // October 3–6, 2016



How do we measure progress / success?

- **Practice:** through Open Data releases
 - Can the data really be (re-)used by the Designated Community(ies)?
 - What are the support costs?
 - Is this sustainable?
- **Theory:** by applying state of the art "preservation principles"
 - Measured through ISO 16363 (self-) certification and associated policies and strategies
 - Participation in relevant working & interest groups

One, without the other, is probably not enough. The two together should provide a pretty robust measurement...

ISO 16363 Certification

- Discussed already at 2015 PoW and several WLCG OB meetings
- Proposed approach:
 - An **Operational Circular** that describes the organisation's commitment to the preservation of scientific data & general principles (draft exists);
 - **Data Management Plans** by project where needed to refine embargo periods, designated communities etc.
 - A **Preservation Strategic Plan** covering a 3-5 year period
 - **DPHEP Blueprint (2012) and Status Report (2015) can be considered the first in such a series**
- This should cover the "holes" we have wrt section 3 of ISO 16363
- Needs to be done in close collaboration with experiments and other LTDP service providers: **start with a Workshop in 2017**
- **March 13 – 15: <https://indico.cern.ch/event/588219/>**

Assessing ISO 16363 “compliance”

Assessment	Meaning
0	Non-compliant or not started: The repository has not yet addressed the requirement or has not started the review of the requirement.
1	Slightly compliant: The repository has something in place, but has a lot of work to do in addressing the requirement.
2	Half compliant: The repository has partially addressed the requirement and (<i>but?</i>) has significant work remaining to fully address the requirement.
3	Mostly compliant: The repository can demonstrate that it has mostly addressed the requirement and is working on full compliance.
4	Fully compliant: The repository can demonstrate that it has comprehensively addressed the requirement.

Organisational Infrastructure (3)

(25 sub-metrics)

3.1	Governance & Organisational Viability	Mission Statement, Preservation Policy, Implementation plan(s) etc.
3.2	Organisational Structure & Staffing	Duties, staffing, professional development etc. [APT etc.]
3.3	Procedural accountability & preservation policy framework	Designated communities, knowledge bases, policies & reviews, change management, transparency & accountability etc.
3.4	Financial sustainability	Business planning processes, financial practices and procedures etc.
3.5	Contracts, licenses & liabilities, "rights"	For the digital materials preserved...

Infrastructure & Security

Risk Management (5)

(24 sub-metrics)

5.1 Technical Infrastructure Risk Management



Technology watches, h/w & s/w changes, detection of bit corruption or loss, reporting, security updates, storage media refreshing, change management, critical processes, handling of multiple data copies etc.

5.2 Security Risk Management



Security risks (data, systems, personnel, physical plant), disaster preparedness and recovery plans ...
ISO 27000 etc.