EIROforum

EIROforum Serving European Science

Large Scale Data Projects

Jamie.Shiers@cern.ch

On behalf of the EIROforum IT WG

http://www.eiroforum.org/

CERN, EUROfusion, EMBL, ESA, ESO, ESRF, European XFEL, ILL

Questions (& Answers)

1. What are the opportunities wrt data which you see in your work that are relevant to the RDA?
2. Are any of the first WG results have relevance for your work?
3. Which RDA groups are working towards solutions that are relevant for your work?

4. **Are there results that you urgently need that you would work towards with others in the RDA?**

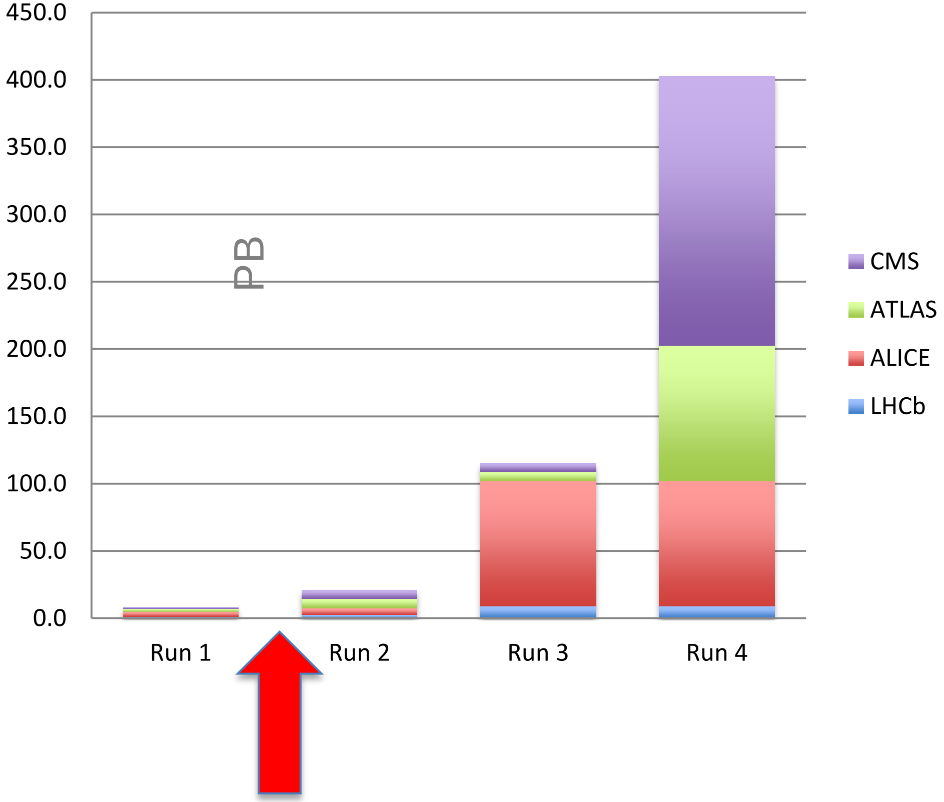
5. What are the most efficient + cost-effective solutions in your environment and how might they be relevant to the RDA?
6. Is cross-border (countries, disciplines, etc.) data exchange and re-use relevant and an issue in your environment?

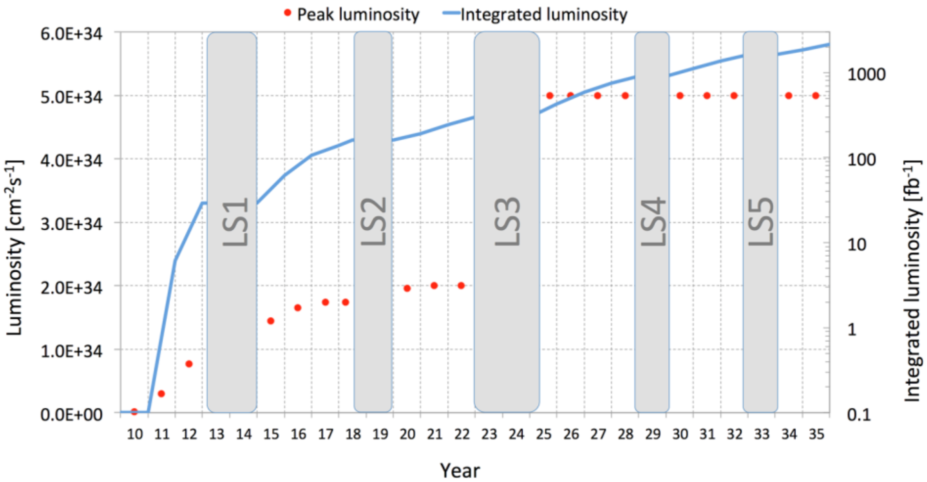
Project Characteristics

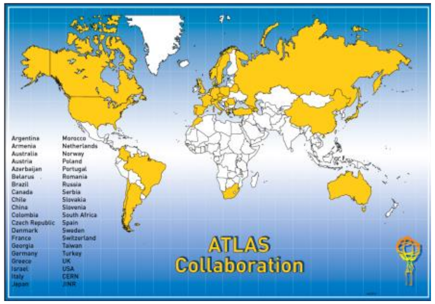
- **Extremely large** data volumes + future growth (**EB & beyond**);
- The **distributed** nature of the (international) user communities;
- **Significant computational requirements to process the data;**
- High network bandwidth / low latency requirements to adequately distribute, collect and / or remotely access the data;
- **Long (and increasing) project lifetimes ("data preservation");**
- **The strong desire for "common" solutions = cost effective services + an enabler for data sharing and re-use;**

- **The need – in at least some cases – for highly available services to match the above requirements.**

Data: Outlook for HL-LHC @ CERN







ATLAS consists of ~2900 persons from 37 countries, more than 170 institutes / universities

- ➢ **0.5 EB / year (2025 – 2035) is probably an under estimate!**

Data Preservation plans

- The ALICE collaboration is committed to develop a long term program for Data Preservation to serve the triple purpose of

    i. preserving data, **software** and **know-how** inside the Collaboration,

    ii. **sharing** data and associated software and documentation with the larger scientific community, and

    iii. **give access to reduced data sets and associated software and documentation to the general public for educational and outreach activities.**

2020 Vision for LT DP in HEP

- ***Long-term – e.g. FCC timescales**: **disruptive change***
  - By 2020, all **archived data** – e.g. that described in DPHEP Blueprint, including LHC data – easily **findable**, fully **usable** by **designated communities** with clear (Open) access policies and possibilities to annotate further
  - Best practices, tools and services well run-in, fully documented and sustainable; built in common with **other disciplines**, based on standards
  - **DPHEP portal**, through which data / tools accessed
    - **"HEP FAIRport": Findable, Accessible, Interoperable, Re-usable**

- **Agree with Funding Agencies clear targets & metrics**

Vision & its Implementation Significantly Assisted Through RDA

The Challenge(s)

1. **Reproducibility of results – over long periods of time and changing e-infrastructures**

2. **Data Sharing – even with long-ish embargo periods – can translate to significant demands**

3. **From Open Access to Open Data to Open Knowledge**