Sustainability of e-Infrastructures

Jamie.Shiers@cern.ch

DPHEP Project Manager

Former WLCG Service (Challenge) Coordinator

Former CERN Program Librarian

Former VAX Operations Manager

Context

- **Consider concrete, non-trivial Use Cases, e.g.**

    Reproducibility of Data / Results

+ the "Data Fabric" needed to support them

- This involves many sub-services with differing levels of complexity, requirements on reliability / availability etc

➢ **Don't imagine that is trivial**
➢ **Don't expect that people will have deployment / operations experience**
➢ *Don't ignore the vast experience already out there!*

Communities

- **Involve communities as early on as possible!**

  - Understand their Use Cases, their priorities, their **language**(s)
  - Set realistic expectations!
  - No software is bug free, no services are 100% reliable, no support is infallible…

  - **This requires a lot of leg-work during startup phase – which may be long (WLCG: 5-6 years?)**

- But with simple, clear metrics & procedures even extremely complex e-infrastructures can be built!
- And **grow** with **decreasing** cost of ownership…

The challenge of keeping the ATLAS run 1 data alive

1) It must be possible to reprocess the RAW data with the desired conditions and the new software version and the AOD must be made available to users

2) There must be software available to read and analyse the data AODs

3) It must be possible to simulate newly generated MC events with the geometry corresponding to the data

4) It must be possible to digitize the MC events with the appropriate software to emulate the readout, pileup, beam conditions etc. corresponding to the data

5) It must be possible to reconstruct the MC events in the same way as the data were reconstructed and write MC AODs

6) It must be possible to determine the trigger efficiency for physics analysis

7) It must be possible to retrieve any metadata required for physics analyses, e.g. the LHC beam conditions, ATLAS data taking and data quality conditions etc.

Example of "Physicist language"

Conclusions

- I know I haven't provided a "silver bullet" – there isn't one

- But we are waiting / willing / wanting to share our experience, contribute to an eventual IG, help build, run, promote, maintain relevant eResearch tools / "Data Fabric(s)", participate in User Events / Fora etc.

- "Existing experience" documented in published papers, plus milestones / deliverables / reports from worldwide e-infrastructure projects
  - Maybe a "catalog" of these & others would be a useful starting point?

- This includes not only CERN but also "our partners" in existing and future e-Infrastructures…