Data Preservation Update

Data Preservation, Curation & Stewardship

Jamie.Shiers@cern.ch

WLCG Overview Board (Extract)

November 28 2014

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

Certification of DRs

- Since last May, a training course on certification of digital repositories, offered by the authors of the standards, has appeared

- In addition, an H2020 proposal “CTRUST” has been submitted which aims to align the (already compatible) procedures
  - DANS, DIN, ISO, CO DATA + corresponding RDA IG (message here!)

- Certification of DRs expected to be one (of many) topics at the 2015 “PV” data preservation conference (Nov in Darmstadt)

➢ **Several people have asked about this course, and it could be “offered” at CERN in 2015 if sufficient interest.**
  - e.g. for people from Tier0 + Tier1 sites
  - The estimated cost for ~15 people is £10K + subsistence for the trainers (+ travel and subsistence for remote attendees)
  - (Covered centrally from WLCG budget)

Why Bother?

- ✚ Help align policies and practices across sites
- ✚ Improve reliability, eliminate duplication of effort, reduce “costs of curation”
  - Some of this is being done via HEPiX WG
- ✚ Help address the “Data Management Plan” issue required by Funding Agencies
- ✚ Increase “trust” with experiments wrt stewardship of the data
- ✚ Increase our attractiveness for future H2020 bids and / or to additional communities
  - What we offer is unique qualitatively and quantitatively – having a “Data Seal of Approval” would back this up

- ➢ **Following the course would not be a commitment to proceed with certification…**
- (But I think it would be a good idea)

- The Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB) plays a major role in training auditors and repository managers. There are three important ISO standards:
  - ISO 14721 (OAIS – a reference model for what is required for an archive to provide long-term preservation of digital information)
  - ISO 16363 (Audit and certification of trustworthy digital repositories – sets out comprehensive metrics for what an archive must do, based on OAIS)
  - ISO 16919 (soon to be published - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories – specifies the competencies and requirements on auditing bodies)

- These three standards form a closely-related family and an understanding of their principles and use will become increasingly important in establishing **an internationally recognised set of trustworthy digital repositories**.

- ***Details of the programme and registration information is available on the PTAB web site*** http://www.iso16363.org/

Dear colleagues,

In a recent Overview Board, Jamie Shiers presented the idea of offering training for certification of the WLCG data centres as long term digital repositories. In particular, as we discussed, it may be useful in the long term to have these certifications, especially if we think that our data centres can become the core of an infrastructure for long term scientific data archiving. Please encourage your Tier 1 managers to strongly consider participating in this offered course.

Best Regards,

Ian

See http://indico.cern.ch/event/336878/ for presentation(s) and draft minutes

Register for the course (15-19 June at CERN) here: https://indico.cern.ch/event/376809/