Canadian "Research Data" Activities

Canadian HEP projects

Research data management in Canada

Related activities

Summary



Institute of Particle Physics

The Canadian HEP projects include those at international laboratories (CERN, JPARC/Kamioka, KEK, SLAC)

and

national laboratories (TRIUMF and SNOLAB).

IPP Program

While the IPP broadly supports particle physics research in Canada, we maintain a set of "IPP projects" that constitute our core programme. These have been judged (by **IPP Council**) to satsify our **IPP project criteria**. Our research scientists spend the majority of their time working on IPP projects (see our **policies** for details). The current list of IPP projects is included below

Under Construction	Operating	Post Data-Taking
Belle II (KEK)	ATLAS (CERN)	BaBar (SLAC)
DEAP (SNOLAB)	EXO-WIPP	PiENu (TRIUMF)
SNO+ (SNOLAB)	IceCube	
SuperCDMS (SNOLAB)	PICO/PICASSO (SNOLAB)	
	T2K (JPARC/Kamioka)	
	Veritas	



Projects

- Off-shore projects: DP managed by local laboratory and/or collaboration
 - T2K, IceCube, Veritas
 - OPAL, BaBar, Bellell, ATLAS (discussed by others in this meeting)
- TRIUMF projects
 - Nuclear physics (rare-isotope science and offshore program)
 - Rare pion decay (PiENu)
- SNOLAB projects

Under construction: DEAP, SNO+, SuperCDMS

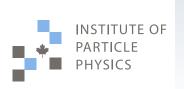
Operating: ExoWIPP, PICO/Picasso

– Completed: SNO



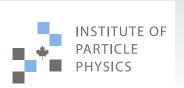
Canadian initiatives

- HEP
 - Cloud for HEP (ATLAS and Belle II)
 - Participated in the SLAC LTDA for BaBar
- Research Data Canada
 - Initiated to address data management issue for all research communities
 - (modestly) supported by our funding agencies and other organizations
 - Discussing issues with international organizations
 - Pilot project for data curation
 - Targeting small science and humanities research communities
- Funding agencies discussing policies for research data
- Compute Canada provides computation and storage resources
- CANARIE provides research network and international links



Summary

- International HEP projects in Canada are relying on laboratories and collaborations to lead DP activities
 - Provided help with virtualization of BaBar
- Smaller TRIUMF and SNOLAB experiments store their data on nationally supported computing facilities
 - Ensures backup and secure storage
- National initiatives
 - Research Data Canada looking at a variety of issues (data storage, archives, data preservation)
 - HEP funding body (NSERC) developing "data management" policies



Appendix of Project Information

International:

- ATLAS (CERN), BaBar (SLAC), Belle II (KEK), IceCube (FNAL)
- T2K, Veritas (see following slides)

SNOLAB:

- DEAP EXO-WIPP (SL6, ROOTv5, python, Geant4)
- PICO/PICASSO
- SNO+
- SuperCDMS
- SNO

TRIUMF:

- PiENu
- Rare-isotope science and other projects



T2K/SuperK (Canada)

First, the near detector. In accord with your expectation, long term archival has not received much attention.

Some salient points:

- 1. Raw data is in MIDAS format. Format is fairly simple, but need MIDAS or ND280 software to read it. We have at least three copies of the raw data files.
- 2. Processed data and MC files are in ROOT format. We usually keep this data for a while, but eventually remove old data for space reasons. High level analysis files have all been kept up to now.
- 3. A bigger problem will be ensuring that we can still run the calibration, reconstruction and analysis software in a couple years. We agree that virtualization would be helpful; we have not worked much on this yet.

The situation at Super-K is different in some respects:

- We haven't thought much about data preservation; since ICRR/Kamioka is the host lab, they are actively maintaining the raw data.
- 2. There is a serious issue with obsolete CERNLIB formats (hbook, ZBS) that is at the core of the Super-K framework, though limited parts of the code are now ROOT-based. There has been a process ongoing for a while to update the code to be at least gfortran-compatible so that it can run on modern compilers that we find on ComputeCanada.



DEAP (SNOLAB)

- DEAP uses a custom data structure using root v5.
- Our software runs under Scientific Linux 6.
- We use our version of rat, which is the "Reactor Analysis Toolkit", a root and geant4 based code with a long history in neutrino physics.
- It runs under standard SL6 compilers.
- We use python for scripting.
- Build is managed with scons.
- Software versioning is done with git.

Our data could absolutely be read in a virtual environment.



Veritas

In the VERITAS collaboration we give this topic some thought from time to time but it suffers from a lack of manpower.

Also the problem is not a big one since we anticipate the data set to be 500 TB at the end of data-taking and we'll reduce that to 50 TB with mild cuts for long-term storage.

We use ROOT for some of our analysis code but the data are stored in files that are in a VERITAS-specific format.



ICECUBE (Canada)

I note that for IceCube the long term storage is a responsibility of UW-Madison (and the NSF) for the project and the preservation activities are now being arranged via FermiLab.

We have our short-term data RAC on Compute-Canada for our analyses, and in terms of Canadian data storage this should be sufficient for our activities.

