



Making the Most of Research Data

Preserving Research Data to Enable Long-term Investigation and Discovery

Project Name: High Energy Physics (HEP) Legacy Data Project

Project Lead: University of Victoria, British Columbia

CANARIE contribution: \$578,000

Participants:

- National Research Council, Ottawa, ON

What is the HEP Legacy Data Project?

This project builds an environment for the preservation of data and software from the BaBar high-energy physics (HEP) project. BaBar is a project at the SLAC National Accelerator lab at Stanford University that is trying to find out why the universe is made of matter and not anti-matter. It stopped recording electron and positron collisions on schedule in 2008, but the analysis of data will continue for many years.

The goal of this legacy data project is to provide a technical solution that will enable the analysis of the BaBar data for ten or more years.

In order to preserve both the data and the software environment from BaBar, this project is focusing on three aspects:

- Understanding how to run complex HEP applications in a virtual computing environment;
- Development of science cloud facilities in Canada, and use of commercial cloud resources such as the Amazon EC2 cloud; and
- Investigation of how to manage the virtual computing environment, the scientific data and the use of multiple clouds seamlessly as a single entity, regardless of their geographic location.

The project uses the CANARIE high-speed network to transmit data flows and connect data, tools and researchers in Canada and the US.

Value to Research and to Canada:

- Canada invests millions of dollars in data obtained from research in a wide range of fields; ensuring that the data and the software environment can continue to be mined for knowledge leverages this investment
- The platform developed in the project will allow researchers to exploit research data for years to come

Did you know?

The use of distributed cloud computing resources from academic organizations and commercial providers gives researchers on-demand access to large amounts of computing power and data storage.

It can be used for the archiving and preservation of research data and for new and ongoing research projects.



The SLAC National Accelerator Laboratory of the U.S. Department of Energy



An engineer works inside a synchrotron.