

BABAR Data Management Plan

June, 17th, 2013

1. BaBar data

BABAR experiment collected data from May 1999 (first data used for physics research was collected on October 22nd, 1999) until April 7th, 2008 at the PEP-II asymmetric $e+e-$ storage ring at the SLAC National Accelerator Laboratory.

Most of the data were collected at an energy of 10.58GeV corresponding to the $Y(4S)$ resonance, and 40MeV for the continuum and background studies. BABAR also collected data samples at the energies corresponding to the $Y(2S)$ and $Y(3S)$ resonances. The $Y(3S)$ dataset is, as of today, the largest ever collected.

The raw data were collected in XTC format (eXtended Tagged Container) and, after going through an appropriate processing chain, is made available to the BABAR users in ROOT (<http://root.cern.ch/>) format.

The amount of data, both experimental data (including XTC files) and simulated data, currently in use is about 2.3 Pbytes stored at SLAC computing division HPSS Storage Tek tape system on about 2400 T10000B tapes. The most used data is available on disk.

A migration to a newer generation of tapes (T10000C or T10000D) is foreseen in one or two years from now.

The primary goals of BABAR research program include B meson and D meson physics, CP and T violation, charm and tau physics, as well as searches for physics beyond the Standard Model.

To date BABAR has published 512 papers (on PRL, PRD, PRD-RC, and NIM) and 13 more are submitted.

2. Metadata

BABAR data, code releases, and tools for physics analysis are accurately documented in html format (<http://www.slac.stanford.edu/BFROOT/>) and in a password protected wiki web site (<https://bbr-wiki.slac.stanford.edu/>). Code, tools, and internal documents relative to all BABAR physics analyses are stored and developed using a dedicated cvs (Concurrent Versions System, <http://cvs.nongnu.org/>) repository.

Metadata for the experimental data, simulated data, code management, and internal documents are stored in several tens of tables in Oracle and MySQL databases. Tools and web interfaces for accessing and managing information are provided for the collaborators.

Public information is found at <http://www-public.slac.stanford.edu/babar/>.

3. Data availability

The BABAR data is only available to BABAR collaborators, both members and associates. Even though the Collaboration is discussing its future Governance, no position is taken with respect to making the data publicly available. On the other end the BABAR collaboration welcomes new associates that wish to have access to the BABAR data and have a valid use case.

The Computing Center of the French “*Institut National de Physique Nucléaire et de Physique des Particules*” (IN2P3) of CNRS (<http://www.in2p3.fr/>), a major Tier-A for BABAR computing, has agreed to host a full copy of the BABAR dataset until at least 2018. The data transfer started during summer 2012 and will be completed by fall 2013.

Subsamples of BABAR dataset are hosted in few Tier-A sites that support BABAR physics analysis. At the moment these are CC-IN2P3 in Lyon, France; CNAF in Bologna, Italy; University of Victoria, Canada; GridKa in Karlsruhe, Germany. With the exception of CC-IN2P3, their support is based on a year by year agreement depending on funding.

4. Long Term Data Access and Data Preservation

In order to preserve the capability of performing new analyses in the future, BABAR computing developed virtual machines designed to support the BABAR analysis environment and a dedicated cluster provides most of the resources needed by the Collaboration on the long term.

The Long Term Data Access project is introduced in the following paper (NIM reference).

BABAR computing group is also a member of the DPHEP (<http://www.dphep.org/>), an ICFA study group for data preservation and long term analysis in high energy physics.