Belle II Data Management Plan

1 Purpose of this Document

The primary focus of this document is to detail the Belle II policy ensuring that Belle II data will be reliably maintained in a form accessible to Belle II members.

2 Policies for Different Belle II Data Types

Access to Belle II data can be considered at each level of increasing complexity described below. Belle II data comprises the collected raw experimental and simulated data, the derived data products stored and catalogued in the Belle II Distributed Data Management system, the calibration data, and all metadata and documentation required to reproduce the derived production and obtain physics results[1]. These policies defined below also pertain to event generators, simulation and reconstruction software, physics and detector simulation, and data from detector R&D groups.

Belle II is committed to preserving all raw data from collisions to allow for reprocessing and analysis for the active lifetime of the collaboration. Access to Belle II data can be considered at each level of increasing complexity described below.

2.1 Level-1. Published Results

Only Level-1 data are considered Digital Research Data, according to the US DOE[2]. The Belle II Collaboration intends to maintain the information necessary to perform and review physics analysis with Belle II data for the active lifetime of the Belle II collaboration. Belle II internal documentation (see Level-3 below) relating to physics results are maintained in a professionally operated document management system. Scientific output is published in journals or submitted to repositories such as the arXiv and are assumed to be preserved by the journal or repository operators in addition to being archived by the Belle II Collaboration. Data associated with journal publications including tables, data from plots, etc. are made available in supplementary materials.

2.2 Level-2. Outreach and Education

The Belle II collaboration recognizes the vital role of outreach and education. Outreach and educational activities are encouraged and supported, and selected data are made available both for use within the collaboration and for use by third parties. Data provided for these activities are not intended for physics publications. While Belle II is committed to supporting these activities, no long term commitment is made for preserving the datasets and materials for outreach and educational purposes.

2.3 Level-3. Unpublished Physics Data

Unpublished physics data include preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. The associated materials are maintained in the Belle II document management system. Access is restricted to collaboration members as the data may have publication potential.

---

[1]Belle II physics data do not contain any personally identifiable information.
[2]https://science.energy.gov/funding-opportunities/digital-data-management

2.4 Level-4. Unpublished Detector Data

Belle II will produce enormous amounts of data in various formats. These data are intended for physics and other publications by collaboration members and access is restricted to same.

2.4.1 Non-reproducible Detector Data

The preservation policy for raw data is intended to reduce the risk of loss due to technical failures or disasters like fire, flood, or earthquake. At least two replicas of Belle II raw data will be stored on accessible archival media. One complete replica will be maintained at KEK and another at one or more computing sites, in accordance with the Computing memorandum of understanding (MOU).

Other non-reproducible data including calibration data, metadata, and documentation will be stored in professionally engineered and backed up databases or file systems within the Belle II grid and/or at BNL and DESY.

2.4.2 Derived Data

Derived data and simulated data are, in principle, fully reproducible at any time provided the appropriate software, calibration data, metadata, CPU architecture or emulations thereof, and documentation are available.

2.5 Level-5. Non-reproducible Data

It is generally not possible to reconstruct data from beam test campaigns and detector R&D efforts without the associated metadata, such as logbooks and environmental conditions data. Belle II is committed to preserving this data for collaboration members for the lifetime of the collaboration. However, the collaboration makes no provisions to make such data available to outside members, beyond the publications and associated tables from those campaigns.

2.6 Level-6. Software

The Belle II software is publicly accessible[3] and is maintained in git repositories hosted by DESY.

---

[3]url:https://github.com/belle2/basf2, doi:10.5281/zenodo.5574115