SiD Data Mangement Plan v1.0

September 6, 2016

Draft plan under consideration by SiD Executive Committee and SiD Institute Board

Introduction

SiD has fully supported the principle of open access in its data management policy. The main objective is to preserve the data reliably for SiD members, and to make the data available in a usable way to people external to the SiD consortium.

The collaboration's need to preserve data for its own use shares some requirements with making them open access. To support open access to data additional resources will be required to develop and support the infrastructure to make the data available.

Policies for Different Data Levels

Access to SiD data can be considered at each level of increasing complexity described below, with associated conditions, see Ref. [1]. These policies pertain to event generators, simulation and reconstruction software, physics and detector simulation, as well as data from detector R&D groups.

Level-1. Published Results

All scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by collaborating institutions.

Data associated with journal publications (plots, tables, etc.) is made available upon request. Software versions used to produce published results are listed in the publication.

Level-2. Outreach and Education

SiD recognizes the vital role of outreach and education, and participates in and encourages outreach and education activities, and makes selected data available for them. Data provided for this role is not intended for publication of physics papers, but may be used for publications in computer science, or education, after consultation with SiD.

Level-3. Unpublished Detector and Physics Data

SiD recognizes the potential value of making its unpublished detector and physics data available. SiD is committed to preserving all the information required to support the use of level-3 data throughout the lifetime of the consortium, which also means that the option to make it open access will be preserved. This publication would require additional infrastructure to host the data in a publicly accessible way.

SiD is willing to engage with any external project that might wish to develop such infrastructure. This would likely require significant additional resources beyond those required for SiD internal needs, and any publication of level-3 data is contingent on SiD being able to identify these resources.

If and when the required infrastructure for making the data accessible by non-SiD members becomes available, the Institute Board or its delegates would consider a staged and delayed release of the level-3 data, subject to the required additional resources being provided. There would be an embargo period on each dataset to allow the members of the consortium a reasonable time to perform analyses. Given the complexity of particle physics analyses, this embargo period will be measured in years, not months.

By far the most practical means of conducting a specific new analysis that requires level-3 data is in partnership with the consortium.

Level-4. Non-reproducible Data

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

Level-5. Software

Software developed by the SiD consortium is published on publicly accessible servers and made available under open source licenses.

References

[1] Status Report of the DPHEP Study Group: Towards a Global Effort for Sustainable Data Preservation in High Energy Physics, arXiv:1205.4667 [hep-ex]