

LHCb External Data Access Policy

LHCb Public Note

Issue: 1
Revision: 1

Reference: LHCb-PUB-2013-003
Created: 22th April 2013
Last modified: 22th April 2013

Prepared By: LHCb
P.Clarke/Editor



Abstract

This document contains the LHCb Data Access Policy. This was adopted at the Collaboration Board meeting of 27th Feb 2013.

Data Access Policy for LHCb

1. Data preservation is fundamentally important for the collaboration itself, regardless of any external requirements. This is to enable collaboration members to access data for many years after it was taken and requires a consistent set of the data, associated software, metadata and conditions and documentation to be preserved. LHCb will seek to develop such a data preservation capability as soon as practical. We will need to identify additional resources for this.
2. LHCb supports the principle of open access. In principle we can envisage providing some such open access based upon the work needed internally for data preservation (point 1 above).
3. LHCb is extremely resource limited at present. Therefore whilst this policy expresses a spirit of intent, we cannot commit to implementation of any capability on any specific timescale. Specifically in respect of open access we will not be able to undertake any significant development to support this without injection of additional resources.
4. Overall the collaboration expects to follow the guidelines being developed by CERN and the LHC experiments jointly on these matters, after appropriate approval by the LHCb Collaboration Board.
5. Open access to its data by people outside the collaboration can be considered at four levels of increasing complexity, listed below, with associated conditions [note: these “levels” 1-4 are those arising in the DPHEP model, and are often referred to as such by all the experiments]. In this first iteration, this policy only pertains to collision physics data (i.e. that sent offline and destined for physics analysis).
6. This policy is adopted by LHCb in good faith according to the spirit of the principles. The collaboration reserves the right to review the policy at any time in the light of experience including, but not limited to, the policy being found to be inadequate in the light of actual requests or any other unintended consequences arising.

Level-1. Published results

All scientific output is published in journals, with preliminary results made available in Conference Reports. All are Open Access, without restriction on use beyond the standard conditions agreed by CERN.

Data associated to the publications will also be made available: tables and data from plots (e.g. including likelihood profiles). The Editorial Board (EB) will recommend the appropriate forms of data and repositories to which these should be submitted.

Level-2. Outreach and Education

LHCb already participates in outreach activities and will continue to do so. This includes event displays of selected events, Ntuple or similar level data for illustrating the calculation of invariant mass distributions, lifetimes, CP asymmetries, etc. Such activity is strongly encouraged.

The data are provided for educational purposes only, and are not considered suitable for publication. Only a limited fraction of the complete LHCb data-set may be used.

Level-3. Reconstructed data

Subject to the resources being identified, LHCb will endeavor to provide open access to some reconstructed (DST) level data on disk at CERN. This will be dependent upon the data preservation work specified in bullet point 1.

In general data will be retained for the sole use of the collaboration for a period commensurate with the very large investment in effort needed to record, reconstruct and analyse those data. After this period some portion of the data will then be made available externally, with this proportion rising with time. The CB will keep such periods and proportions under review and may reconsider whether they should be varied in the light of experience.

In the first instance access will be granted to portions of the DST data five years after data is taken. The portion of the data which LHCb would normally make available is 50% after 5 years, rising to 100% after 10 years. All requests will be considered by the CB and the period and proportion may be varied for specific requests.

For the data that have already been taken, the five years will be counted from the date of ratification of this policy.

The associated software will also be available as open source. Such documentation as exists will also be made available, however LHCb will be unable to guarantee to provide any additional assistance or further documentation.

No review of such publications will normally be undertaken by the collaboration. Any publication that results from data analysis by non-members of the collaboration will require a suitable acknowledgement and disclaimer to be included: acknowledgement that the data was collected by LHCb, and disclaimer that no responsibility for the results is taken by the collaboration. A suitable disclaimer is:

This paper is based on data obtained by the LHCb experiment, but is analyzed independently, and has not been reviewed by the LHCb collaboration.

The use of a suitable license (such as one of the Creative Commons licenses) and reference to a DOI may be adopted.

Members of the LHCb collaboration will not be permitted to sign papers presenting physics results (papers based on LHCb data from any members of the collaboration should be signed by the collaboration as a whole). For papers which do not present physics results, the CB will consider requests on a case-by-case basis, such as for papers presenting novel methods tested upon our data.

Level-4. Raw data

It is practically impossible to make the full raw data-set from scientific endeavours of the scale of high-energy physics easily usable in a meaningful way outside of the collaboration. This is due to the complexity of the data and software, the required knowledge of the detector itself and the methods of reconstruction, the extensive computing resources required and access issues for the enormous volume of data stored on tape. It should be noted that, for these reasons, direct access to the raw data is not even permitted to individuals within the collaboration, and that instead the production of reconstructed data (as discussed in 3. above) is performed centrally. Therefore LHCb will not devote any resources to developing any service to access to the full raw data-set for non-members of the collaboration. Access to representative smaller samples of raw data might be considered if well motivated.

This position is purely pragmatic, and is not driven by any wish to deny access to data, but simply that it is exceedingly unlikely that such data could be useful outside of the collaboration, and that the cost of actively supporting this would be very substantial.

Compliance at CERN level

The policy above embodies that which LHCb regards as a common sense response to the spirit of open data access. It describes that which LHCb will endeavour to do pro-actively to make data available. Notwithstanding this policy, LHCb will naturally not hinder compliance with any requirements from member states which are accepted and agreed by CERN as applying to LHC data.