

NA62 Data Preservation and Open Access Policy

The NA62 Collaboration

Abstract

This document describes the principles governing NA62 data preservation and open access policy. It has been discussed and endorsed by the Steering Committee held on March 27, 2015.

1 Introduction

The NA62 Collaboration at CERN will collect in the coming years the largest sample of kaon decays ever recorded in the world by such dedicated experiments. These data will be unique and are the result of a large investment in human and financial resources by the international community.

The NA62 Collaboration fully supports the principle of open access publication of future results and is also committed to long term data preservation for later re-use both by itself and by a wider community. Educational benefits have also been considered. In all aspects, the NA62 data management practices and policies will mainly follow the LHC experiments scheme [1], adapted to the particular NA62 environment when necessary.

This policy is adopted by NA62 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 future requests or any other unintended consequences arising. Overall the Collaboration expects to follow the guidelines being developed by CERN on these matters, after appropriate approval by the NA62 Steering Committee (SC).

2 Data Preservation

Data preservation is fundamentally important for the Collaboration itself, regardless of any external requirements. This is to enable Collaboration members to access data for years after they were taken and requires a consistent set of the data, associated software, meta-data and conditions and documentation to be preserved. NA62 has some data preservation capability in place, benefiting from the vast expertise available in the larger HEP community [2]. The longer term preservation of associated items will need to be secured and may need specific funding in terms of manpower and financial resources.

3 Open Access

NA62 supports the principle of Open Access as this will, in the long term, allow the maximum realization of the data's scientific potential. However, the level of support that NA62 will be able to provide to external users depends on the resources available to the collaboration.

As with any High Energy Physics experiment data, NA62 data take many forms starting from raw experimental data through reconstructed data and datasets of higher abstraction generated

by analysis work-flows, and finally to data represented in scientific publications. Four levels of complexity, as defined by the DPHEP model [3], are considered below with associated conditions.

3.1 Published results (Level-1)

All scientific output is to be 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. Journals from publishers that are signatories to the SCOAP³ [4] project are prime examples of those fulfilling the open access conditions.

Data associated to the publications, such as tables and numerical input to figures, may also be made available. The NA62 Editorial Board (EB) will recommend the appropriate forms of data and repositories to which these should be submitted.

3.2 Outreach and Education (Level-2)

NA62 does not yet have the capability to develop its own outreach activities but participates in the events organized by CERN, such as Open days and Master classes. Within their national communities, members of NA62 contribute to scientific events that are reported in the press and on radio and television, as well as hosting school events and participating in a wide variety of science festivals and exhibitions. Each year, the NA62 experimental area EHN3 welcomes more than 200 visitors.

The NA62 collaboration is committed to contribute to educational programs, either through the CERN Summer student program or through specific national programs sending undergraduate students to CERN for training periods up to 12 months. NA62 strongly encourages the Collaboration members to supervise such trainees and give them a chance to share everyday research work either hands-on in the Lab or in offline data analysis.

NA62 welcomes many Masters and PhD students during the preparation of their degree to be obtained in their national Universities and includes them as authors on relevant publications as part of the NA62 collaboration. The corresponding degree manuscripts are promoted in a dedicated repository (<http://na48.web.cern.ch/NA48/Welcome/thesis/>).

3.3 Reconstructed data (Level-3)

In general, such 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 analyze those data. As NA62 is going to collect a first round of data from 2015 to 2018, it is not possible to consider making this dataset available to external users before the primary goal of the experiment -a measurement of the ultra-rare decay $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ with $\sim 10\%$ precision- has been reached. After this period some portion of the data could be made available externally, with a time scale better defined after some experience with the data taking and data analysis. The management bodies (SC and EB) will keep such details under review and may consider whether they should be varied in the light of experience.

The associated software could also be made available as open source and the corresponding documentation as part of the NA62 code repository. However NA62 will be unable to guarantee the provision of any additional assistance or further documentation.

No review of publications by external users will normally be undertaken by the Collaboration. Any publication that results from data analysis by non-members of the Collaboration will require a suitable acknowledgment and disclaimer to be included: acknowledgment that the

data was collected by NA62, and disclaimer that no responsibility for the results is taken by the Collaboration. A suitable disclaimer could be:

"This paper is based on data obtained by the NA62 experiment but analyzed independently, and has not been reviewed or endorsed by the NA62 Collaboration."

Publications based on the analysis of NA62 physics data will be signed by the collaboration as a whole with eligible authorship determined by the SC and EB. Neither individual members of the NA62 collaboration nor sub-groups of the NA62 collaboration have the right to author publications presenting physics results based on the analysis of NA62 data. For papers which do not present physics results, the SC and EB bodies will consider requests on a case-by-case basis, such as for papers presenting novel methods tested upon NA62 data.

3.4 Raw data (Level-4)

Because of the complexity of the detector, the data and software, together with access and resource implications, it will not be possible to make sizeable sets of raw data available in any useful way outside the NA62 collaboration. In particular, NA62 are unable to devote any resources towards developing a service to enable such access to raw data for non-members of the NA62 collaboration. However, access to smaller, representative samples of raw data will be considered by the EB and SC if the request is well motivated and if resources permit.

4 Compliance at CERN level

The above policy embodies a pragmatic response to the spirit of open data access. Subject to available resources, NA62 will, however, comply with any requirements from member states which are accepted and agreed by CERN as applying to data produced at CERN.

References

- [1] <http://opendata.cern.ch/about> for general statements and access to specific ATLAS, ALICE, CMS and LHCb policy documents:
<http://opendata.cern.ch/record/413/files/ATLAS-Data-Policy.pdf>,
<http://opendata.cern.ch/record/412/files/ALICE-Data-Policy.pdf>,
<http://opendata.cern.ch/record/411/files/CMS-Data-Policy.pdf>,
<http://opendata.cern.ch/record/410/files/LHCb-Data-Policy.pdf>.
- [2] P. Valente, D. Protopopescu and D. Britton, *A Computing Model for NA62*, CERN NA62-14-03.
- [3] ICFA Study Group on Data Preservation and Long Term Analysis in High Energy Physics, DPHEP, <http://dphep.org>. In particular C. Diaconu et al., *Data Preservation in High-Energy Physics*, <http://arxiv.org/abs/0912.0255>.
- [4] SCOAP³, Sponsoring Consortium for Open Access Publishing in Particle Physics, <http://www.scoap3.org>