

# Table of Contents

<b>Analysis preservation and reproducibility.....</b>	<b>1</b>
Current Projects and Developments.....	1
LHCb specific practices.....	1
Analysis code repositories.....	1
Ntuple Storage.....	2
Working Group Production.....	3
Automated analysis workflow.....	3
Runtime environments.....	3
Tools.....	3
Example analyses.....	4
Documents.....	4

# Analysis preservation and reproducibility

THIS TWIKI IS RETIRED AND REPLACED BY  
<https://lhcb-dpa.web.cern.ch/lhcb-dpa/wp6/index.html>

## Current Projects and Developments

The following projects are currently open for contributions (please contact Sebastian Neubert or Adam Morris):

- Build a bridge between the LHCb publications database [https://lhcbproject.web.cern.ch/Publications/LHCbProjectPublic/Summary\\_all.html](https://lhcbproject.web.cern.ch/Publications/LHCbProjectPublic/Summary_all.html) and the CERN analysis preservation portal. Goal: submit a CAP entry for each entry in the publications database. (Technical Student Project?)
- NTuple-Wizzard: tool to create analysis production option files for basic ntuple productions. (Summerstudent Project, Dillon Fitzgerald, Chris Burr, Adam Morris)
- Run 1 Open data release
- Semantic description of analyses, in particular a decay tree and selections (Explorational project)

### CERN Central Analysis Preservation

The central analysis preservation services CAP and REANA offer full archiving of analyses, including input data, analysis code and documentation. Meta information will be fully searchable and the REANA project will enable rerunning and reinterpretation of analyses.

There is a command line tool available to interface to CAP. A tutorial for this cap-client is also available.

## LHCb specific practices

\*As of December 2017 LHCb has adopted a minimal set of mandatory AP practices as part of the review process. See the presentation <https://indico.cern.ch/event/672229/#75-analysis-preservation>\*

There are 4 domains in which an analysis can be made more reproducible in order to be able to fully exploit the services offered by CAP:

- Analysis code repositories
- Ntuple storage
- Analysis automation
- Runtime environment preservation

Details on this architecture can be found in the internal note LHCb-INT-2017-021 "LHCb Analysis Preservation Roadmap".

In each domain tools and examples have been developed.

## Analysis code repositories

Analysis code is kept in gitlab repositories at <https://gitlab.cern.ch>

The main repository of each analysis should be kept in the PHWG gitlab groups. This will ensure that the code is available to the collaboration even after individual analysts have left.

LHCb physics PHWG gitlab groups (access managed by WG egroups):

- B&Q [↗](#)
  - ◆ B&Q exotics [↗](#)
- RD [↗](#)
- B2OC [↗](#)
- B2CC [↗](#)
- Charm [↗](#)
- BnoC [↗](#)
- QEE [↗](#)
- SLB [↗](#)
- IFT [↗](#)

It is easy to transfer a project from a personal user into a gitlab group but it depends on the access to the groups being setup correctly. This is a responsibility of the conveners.

Typically the access rights are:

- WG Conveners (as per egroup) -- Owner
- WG members (as per egroup) -- Maintainer
- lhcb-general -- Developer

NOTE: Gitlab has a Guest role, but for us it is of no use as it does not allow for the code to be browsed or downloaded (see here [↗](#)).

In addition, please make sure that the project is marked as **private** (as opposed to "public", which makes it accessible to everyone, or "internal", which makes it accessible to anyone with a CERN account including our colleagues on ATLAS and CMS).

If you have questions on how to transfer your project, please contact your WG conveners.

Documentation is available here:

<https://docs.gitlab.com/ce/user/project/settings/index.html#transferring-an-existing-project-into-another-namespace>

**In short:** Settings > General > Advanced (expand) > Transfer project > Select a new namespace (type name of your working group) > Transfer

## Ntuple Storage

There is dedicated storage available on eos: `/eos/lhcb/wg`

The organisation of the working group directories is up to the working groups. Please contact your conveners.

There is a service account available to allow authentication of remote machines (such as gitlab ci runners) for data access.

account name: `lbanadat`

A keytab to be used for authentication is available in a preliminary location at `/eos/lhcb/wg/BandQ/Test/anadat.keytab`. A detailed example of how this is intended to be used is part of the analysis containerisation template [↗](#).

## Working Group Production

Using working group production has the advantage that the ntuples will automatically be available in a central place, accessible to the collaboration. The output of the working group production is managed in the LHCb bookkeeping.

WG productions can be automatically submitted using the WG/CharmWGProd repository on GitLab, see the README for details. Analysts from other working groups are welcome to use the Charm package.

## Automated analysis workflow

Scripting and automating an analysis is no new concept. Indeed anybody who had run (parts of) an analysis on a computing cluster has already done exactly this. In the context of APR practices, scripting and automation has the added benefit of capturing exactly how the analysis tools were executed. Obviously the more parts of the analysis are included in the automation the better this information is preserved. The analysis script(s) also provide an invaluable starting point for new people who want to learn about the analysis and reuse or improve the analysis tools.

Often a simple ROOT, python or even bash script will be all that is needed. Almost any analysis we are aware of has such scripts in one form or another.

In order to handle complex analysis flows, with several data preparation steps and many systematic checks, there are several utilities available. All these so called "workflow engines" provide a way to organize the component scripts of an analysis into a common workflow.

There now is a dedicated lesson on the Snakemake workflow engine [in the LHCb StarterKit](#). Additionally, the official documentation contains a in depth tutorial and lots of additional information.

The goal for full reproducibility is to have a full documentation of how the analysis scripts need to be executed to reproduce the results. An automated workflow can be thought of as a machine-readable documentation.

## Runtime environments

Preservation of the runtime environment of an analysis in Linux containers is working and is used by CAP. A lot of development is still going on in this area.

ROOT docker image LHCb docker image Customizing docker images

## Tools

Tools for analysis preservation and reproducibility are gathered under this gitlab group: <https://gitlab.cern.ch/lhcb-analysis-preservation/>

- Containerisation template [in the LHCb StarterKit](#) - a template allowing easy setup of a new gitlab repository with automatic containerisation and continuous integration, including safe data access on EOS
- A lot of information on gitlab continuous integration and the use of docker images on that service is available here: <https://cern.service-now.com/service-portal/article.do?n=KB0003690>
- Watermarking tool [in the LHCb StarterKit](#) - scripts to watermark plots with the version of the code that was used to produce them

## Example analyses

- [Lambdac  \$\Delta\$ ACP](#) [gitlab repository](#) (pulished)
- [Lb -> LcD0K amplitude anaysis](#) [gitlab repository](#) (in preparation)
- [Lb -> LcD0K branching fraction](#) [gitlab repository](#) (in WG review)
- [Jpsi polarisation](#) [gitlab repository](#) (candidate for validation measurement)

## Documents

- [Internal note LHCb-INT-2017-021](#) ["LHCb Analysis Preservation Roadmap"](#)
- [Presentation](#) [at LHCb Computing Wokshop Nov 2017](#)

-- SebastianNeubert - 2018-06-12

---

This topic: LHCb > AnalysisPreservationReproducibility

Topic revision: r18 - 2021-02-18 - SebastianNeubert



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.  
or Ideas, requests, problems regarding TWiki? use [Discourse](#) or [Send feedback](#)