

Published on *CERN openlab* (<http://openlab-archive-phases-iv-v.web.cern.ch>)

[Home](#) > Improving Reproducibility of Data Science Experiments

---

## Improving Reproducibility of Data Science Experiments <sup>[1]</sup>

**Date published:**

Tuesday, 2 February, 2016

**Document type:**

Conference paper

**Author(s):**

T. Likhomanenko

A. Rogozhnikov

A. Baranov

E. Khairullin

A. Ustyuzhanin

Data analysis in fundamental sciences nowadays is an essential process that pushes frontiers of our knowledge and leads to new discoveries. At the same time we can see that complexity of those analyses increases fast due to a) enormous volumes of datasets being analyzed, b) variety of techniques and algorithms one have to check inside a single analysis, c) distributed nature of research teams that requires special communication media for knowledge and information exchange between individual researchers. There is a lot of resemblance between techniques and problems arising in the areas of industrial information retrieval and particle physics. To address those problems we propose Reproducible Experiment Platform (REP), a software infrastructure to support collaborative ecosystem for computational science. It is a Python based solution for research teams that allows running computational experiments on shared datasets, obtaining repeatable results, and consistent comparisons of the obtained results. Several analysis using Key features of REP are illustrated on several practical cases that were performed at LHCb experiment at CERN. Keywords: machine learning, reproducibility, computation infrastructure, analysis preservation

**Event published at:**

AutoML 2015 workshop @ ICML 2015, Lille, France

[AutoML 2015 workshop @ ICML 2015](#) <sup>[2]</sup>

**Technical document file:**

 [Likhomanenko\\_REP\\_paper.pdf](#) <sup>[3]</sup>

- [Visit Us](#)
- [RSS Feeds](#)

- [Contact us](#)

DISCLAIMER: This Web page contains pointers to material related to the management of CERN openlab in the Information Technology Department at the European Organization for Nuclear Research (CERN). Their use and distribution are regulated by the [CERN copyright notice](#).



---

**Source URL:** [http://openlab-archive-phases-iv-v.web.cern.ch/publications/technical\\_documents/improving-reproducibility-data-science-experiments](http://openlab-archive-phases-iv-v.web.cern.ch/publications/technical_documents/improving-reproducibility-data-science-experiments)

#### **Links**

- [1] [http://openlab-archive-phases-iv-v.web.cern.ch/publications/technical\\_documents/improving-reproducibility-data-science-experiments](http://openlab-archive-phases-iv-v.web.cern.ch/publications/technical_documents/improving-reproducibility-data-science-experiments)
- [2] <https://indico.lal.in2p3.fr/event/2914/>
- [3] [http://openlab-archive-phases-iv-v.web.cern.ch/sites/openlab-archive-phases-iv-v.web.cern.ch/files/technical\\_documents/Likhomanenko\\_REP\\_paper.pdf](http://openlab-archive-phases-iv-v.web.cern.ch/sites/openlab-archive-phases-iv-v.web.cern.ch/files/technical_documents/Likhomanenko_REP_paper.pdf)