



# GeneROOT

*Scalable and Collaborative Research Platforms*

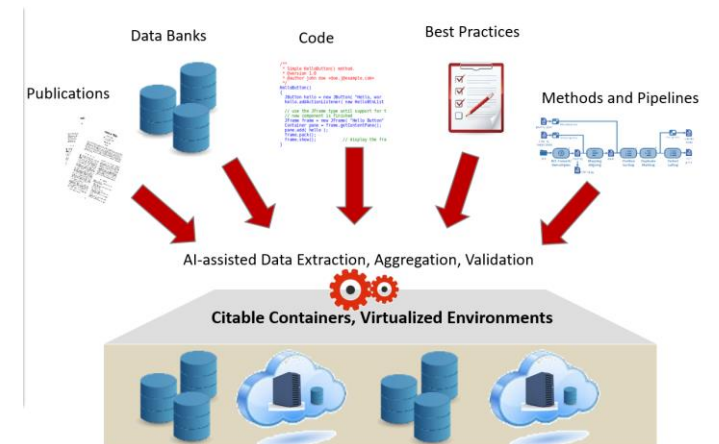
CERN Openlab Open Day - Taghi Aliyev

21/09/2017

# Idea behind the platform

## *And Final Goals*

- Large-scale collaborative research platform
- Main focus on ease-of-use, reproducibility of research
- Use of Machine Learning for Narrative interfaces
  - Information Retrieval (publications, knowledge repos)
  - Natural Language Processing (Chatbots)
- Provide and host in-house solutions and projects



# Current Status

- Iterative generation of the platform infrastructure based on available CERN technologies
  - Why a platform?
  - Technologies currently investigated: REANA, Zenodo, Openshift/Magnum, CVMFS
- Collecting and creating initial use cases
  - Two projects currently: Biomarker detection and Genomics in ROOT
- Prototype based on initial pilot studies

# Projects

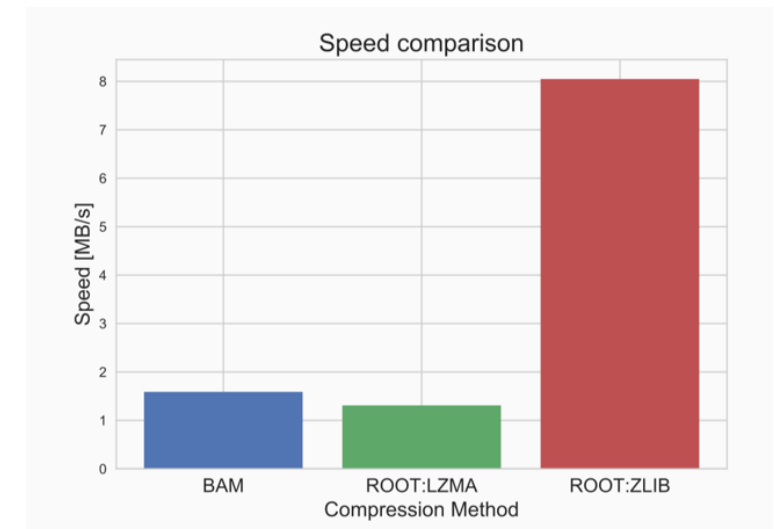
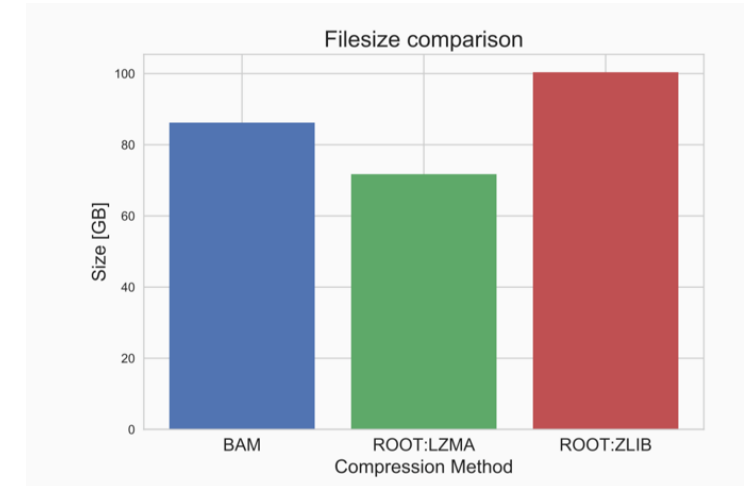
*Initial Projects for initial pipelines/workflows for the prototype stage*

- Biomarker detection
  - Collaboration with Maastricht University in Netherlands
  - Focused on Machine Learning applications for better modeling of the Cardiovascular diseases and earlier prediction of the disease symptoms
- Genomics in ROOT
  - Use of CERN infrastructure for storing and analyzing human genome sequences
  - TwinsUK Data set
  - Two directions:
    1. Development of a new data format based on ROOT for smarter storage of data
    2. Benchmarking of different Structural Variation Discovery tools and integration/interfacing into ROOT

# Genomics in ROOT

## *Storage and Analysis. First Results*

- Using ROOT Ttree to store genomics data
- Initial results are promising
  - Faster read speed
  - Less storage space required
- ROOT Ttree definition is adjustable





# Thank You

*taghi.aliyev@cern.ch*

Twitter: @TaghiAliyev