Abstraction of user storage mechanisms for heterogeneous REANA scientific pipelines

Maria Camila Diaz Sanchez

Supervisors: Tibor Šimko, Audrius Mečionis

September 6th, 2021
Reproducibility Crisis

A Nature survey lifts the lid on how researchers view the ‘crisis’ rocking science and what they think will help.

BY MONYA BAKER

Nature: 1,500 scientists lift the lid on reproducibility
The complexity and cost of HEP research makes it difficult to reproduce the experiments.

With the emergence of new theories, scientists analyse and reuse the data many years later.

Verbytskyi, A (2016)
Can you easily run your analyses under new computing environments?

Non-reproducible research is problematic for advancing science.

Some technologies are heading to make this changes less problematic.
Reproducible research data analysis platform

Flexible
Run many computational workflow engines.

Scalable
Support for remote compute clouds.

Reusable
Containerise once, reuse elsewhere. Cloud-native.

Free
Free Software, MIT licence. Made with ❤ at CERN.
My task

Where are all the files? Where is the workspace?

Currently: Predefined folder in shared storage

How to generalize it, to plug other type of storage?
Abstract the workspace from the workflow

Refactoring the code to allow different storage strategies
Users can choose the workspace of their workflows

Administrators can decide which storage directories to allow
Other directories can now be plugged.
A natural extension could be different storage types, such as object storage
EXTRA RESOURCES

- www.reanahub.io
- docs.reanahub.io
- @reanahub
- info@reanahub.io
- @reanahub

Contact reana team!
QUESTIONS?

macamiladiazsan@gmail.com
@marcdiazsan