



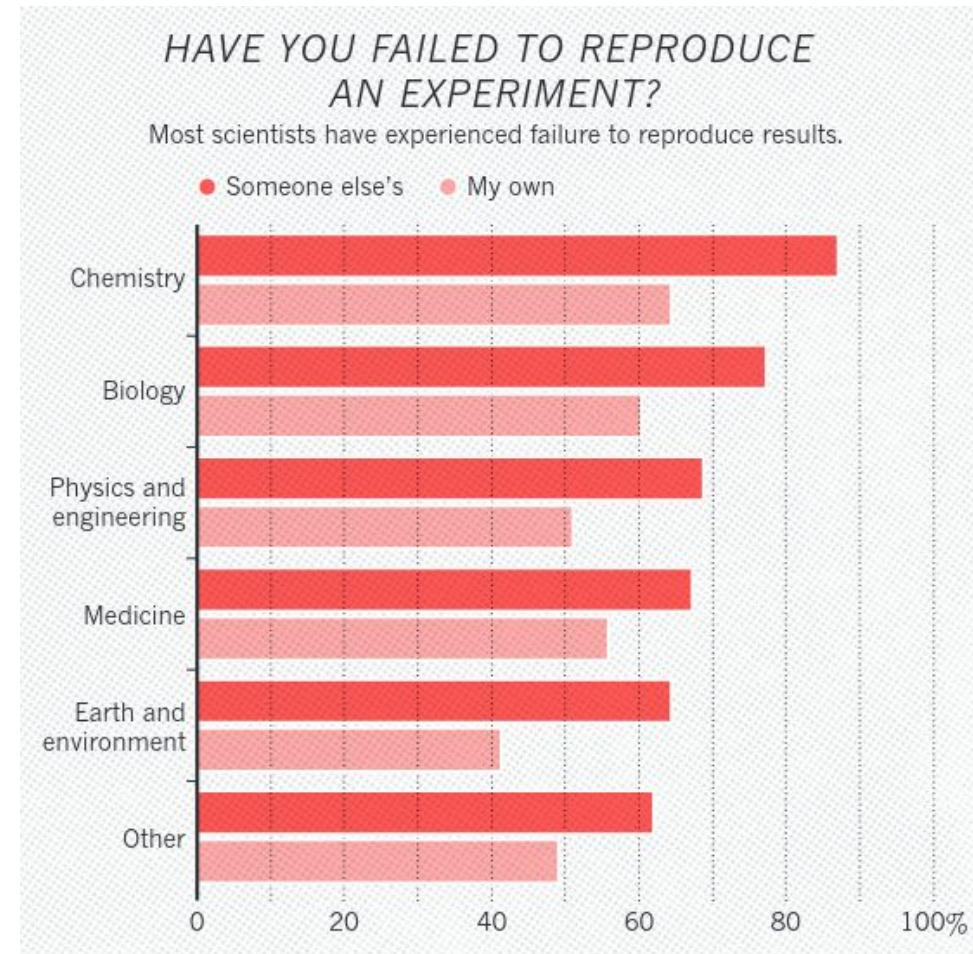
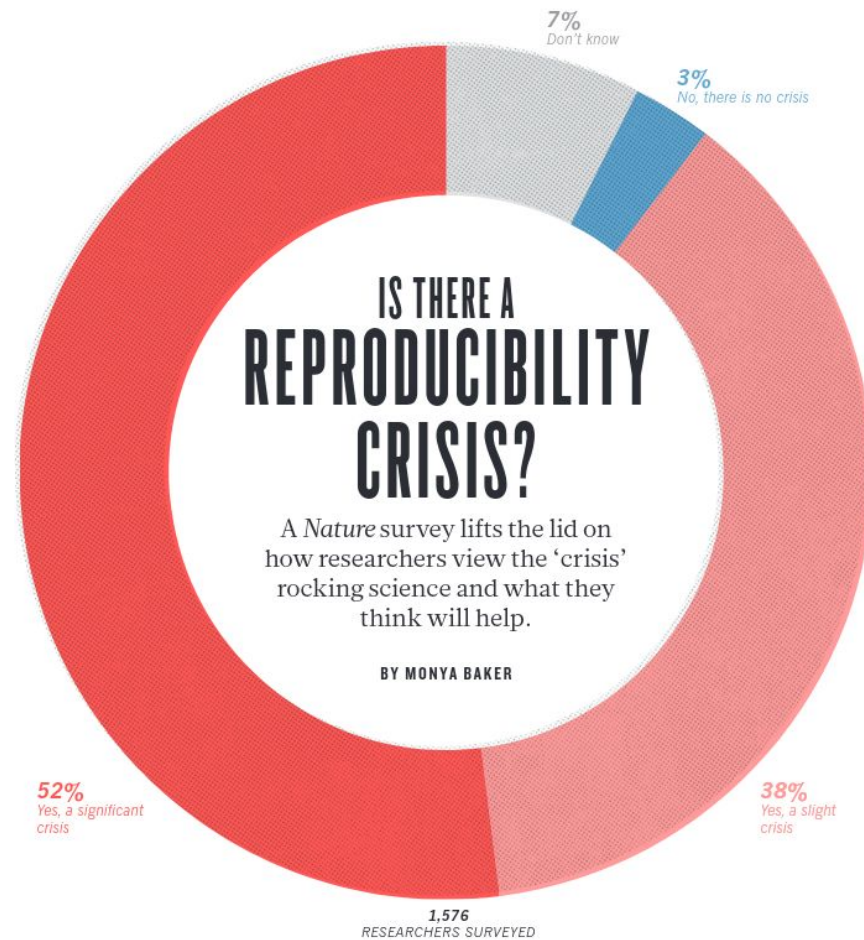
# **Abstraction of user storage mechanisms for heterogeneous REANA scientific pipelines**

*Maria Camila Diaz Sanchez*

Supervisors: Tibor Šimko, Audrius Mečionis

December 1st, 2021

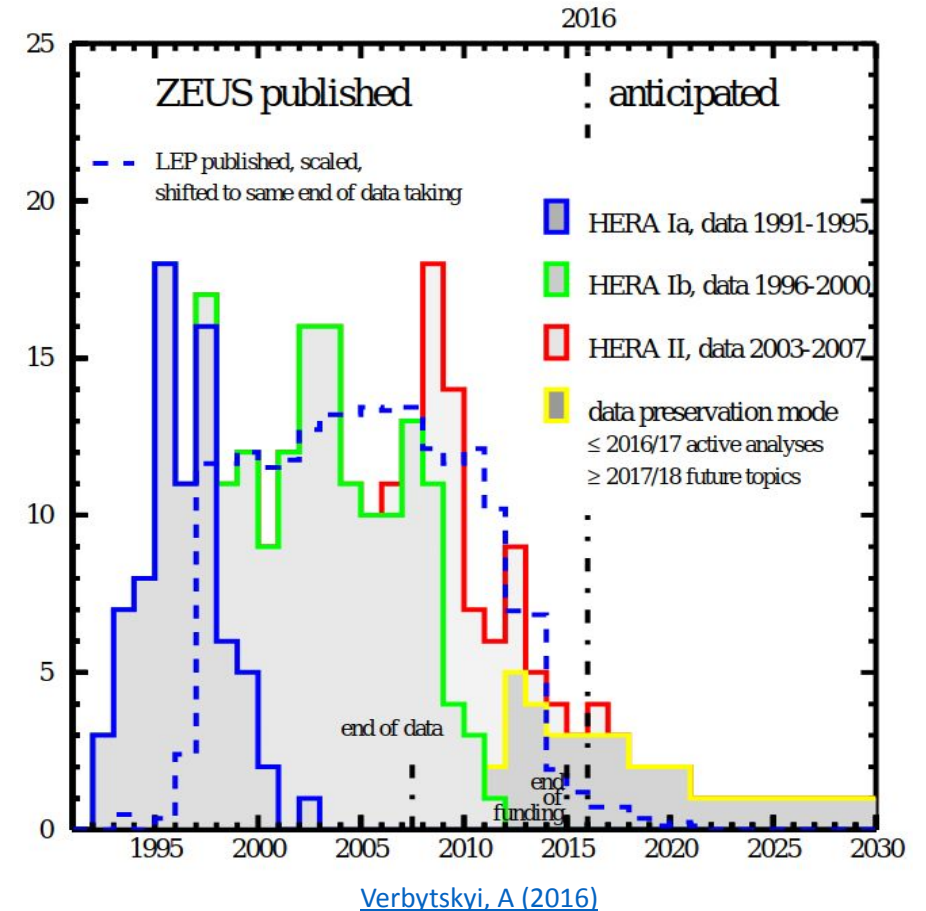
# Reproducibility Crisis



[Nature: 1,500 scientists lift the lid on reproducibility](#)

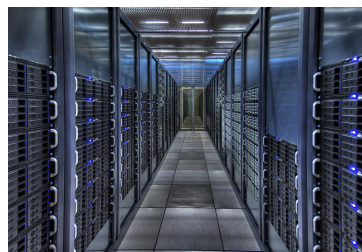
# HEP

The complexity and cost of HEP research makes it difficult to reproduce the experiments.



[Verbytskyi, A \(2016\)](#)

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



```
1 import pandas as pd
2 pd.__version__
```



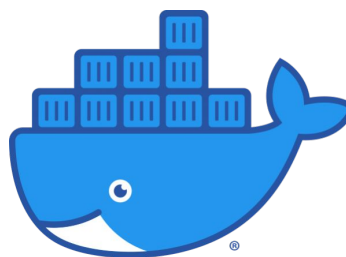
```
1 import pandas as pd
2 pd.__version__
```

Out[1]: '0.24.2'

Out[1]: '1.0.1'



Can you easily run your  
analyses under new  
computing environments?



Non-reproducible research  
is problematic for  
advancing science

**Some technologies are heading to  
make these changes less problematic**



# reana

## Reproducible research data analysis platform

### Flexible

Run many computational workflow engines.



### Scalable

Support for remote compute clouds.



### Reusable

Containerise once, reuse elsewhere. Cloud-native.

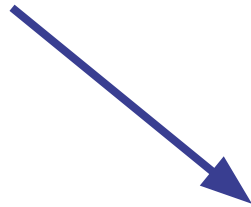


### Free

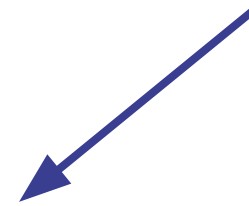
Free Software. MIT licence. Made with ❤ at CERN.



# Input Data



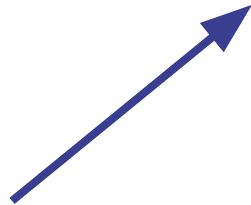
# Code



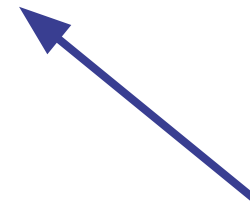
# reana



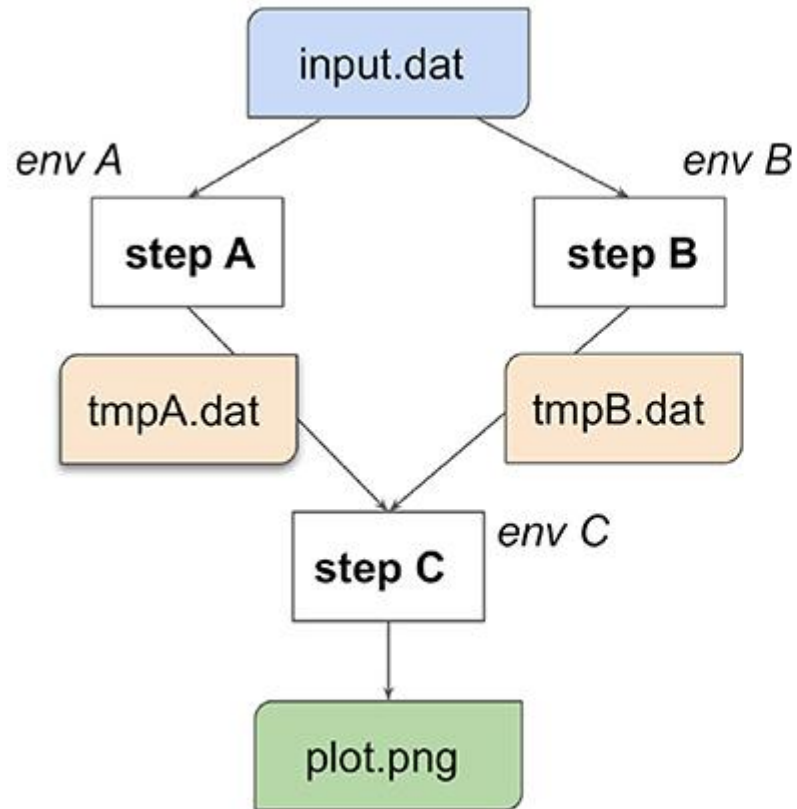
# Environment



# Workflow

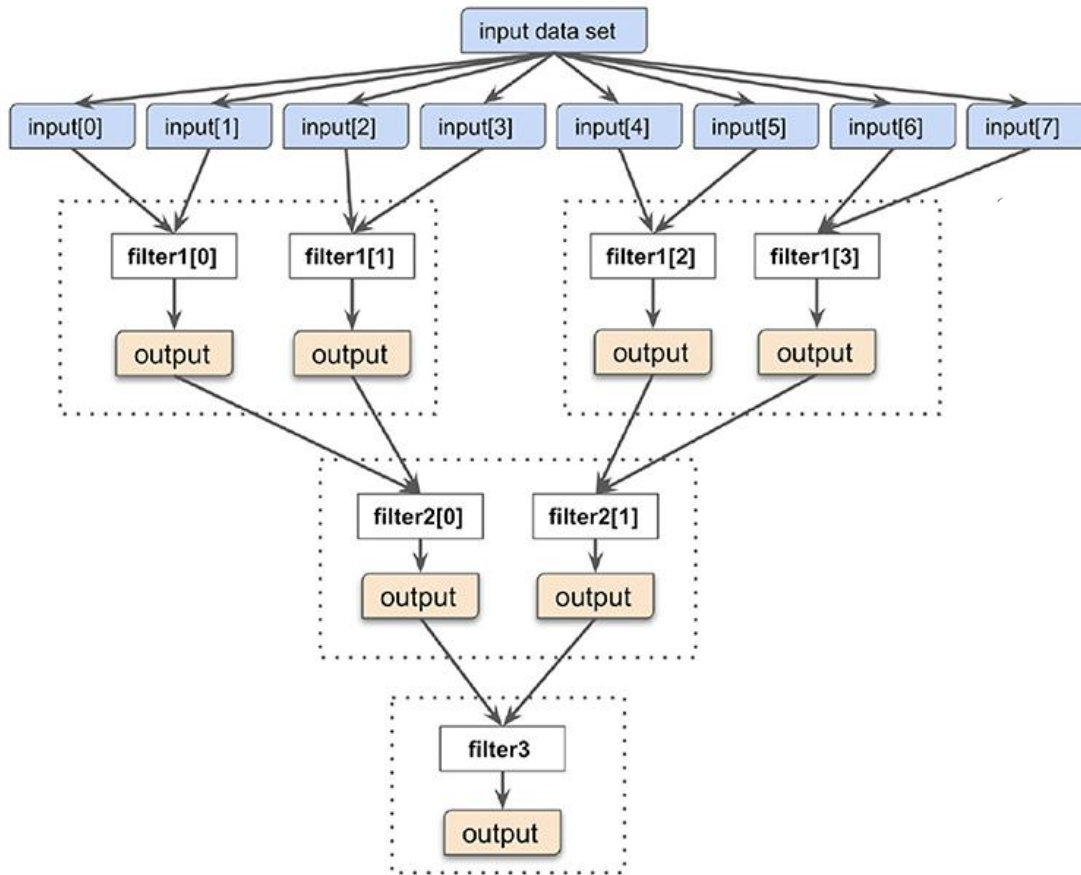


# What is a declarative workflow?



The focus is on expressing **“what”** step should do, separating from and without focusing on **“how”**\*

\*Until is needed, declared and used



**Can be orchestrated**

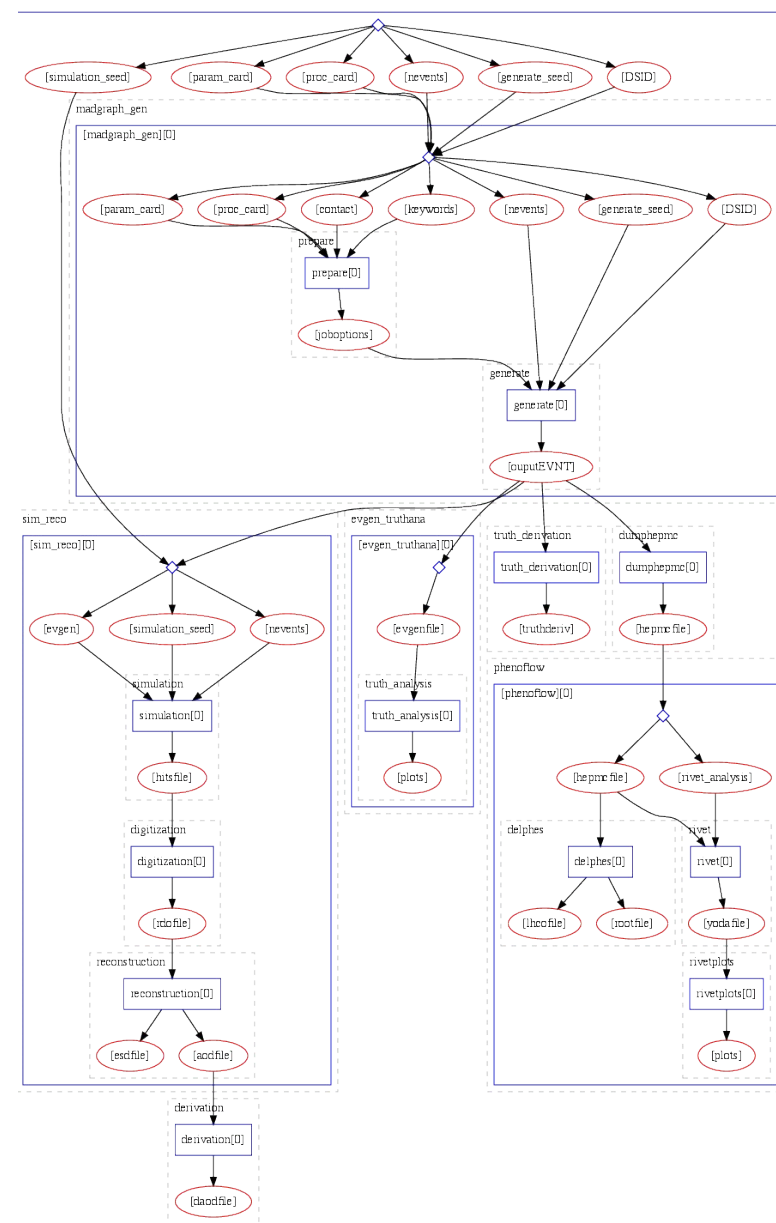
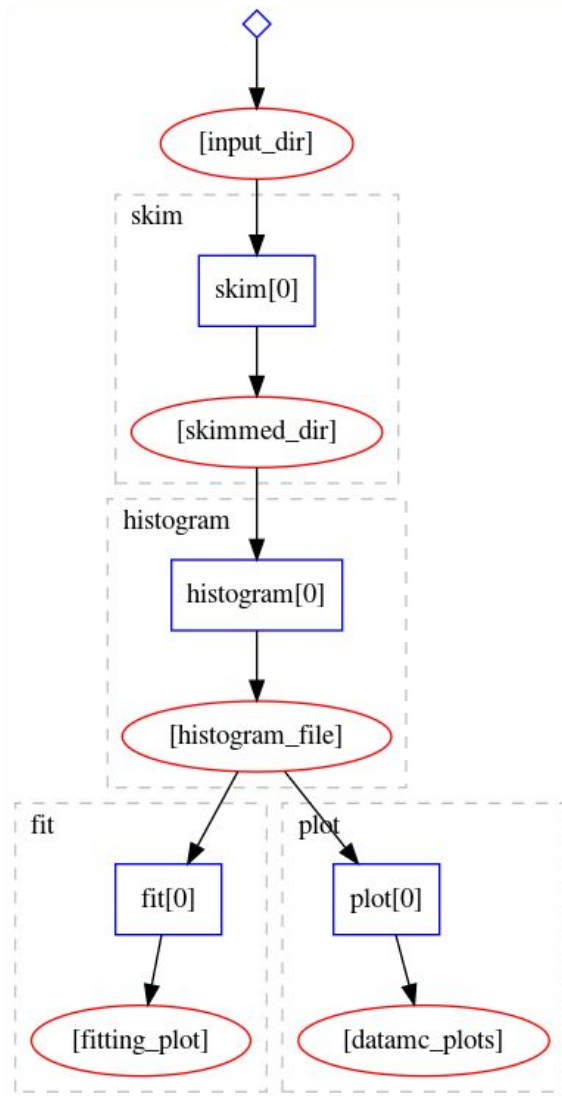


**Each step is isolated and independent**



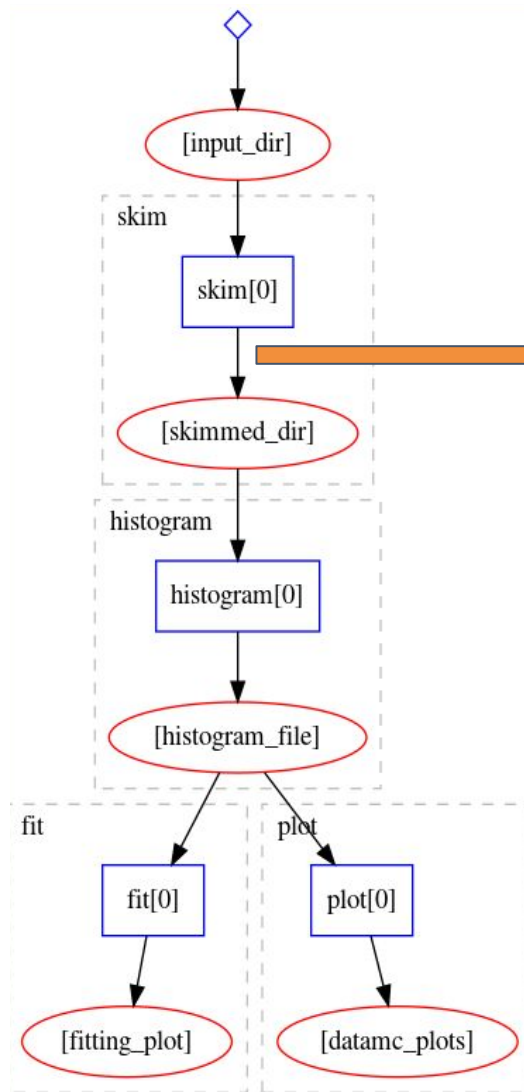
**Can be parallelized and scalable**





[Hildreth, M; et al \(2018\)](#)

# 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

models: extract the workspace concept from within the Workflow model #94

Edit

New issue

Closed

audrium opened this issue on Oct 9, 2020 · 0 comments



audrium commented on Oct 9, 2020

Member



Addressing the following line in #89:

(2) opening interactive sessions without necessarily having any workflow yet.

There is now no notion of `workspace` for `InteractiveSession`'s, however, the `InteractiveSession`'s implementation depends on it. So far, we have been [depending on the workflow model to retrieve it](#), which was ok because the interactive session was also within the workflow model.

As we are evolving to this new idea of separate sessions, I see a need to [extract the workspace concept from within the Workflow model](#) to a new entity. This idea gains more importance once we start thinking about using different workspaces than CephFS.

Originally posted by @diegodelemos in #90 (comment)

Assignees

marcdiazsan



Labels

None yet



Projects

Workspace-Abstraction  
Done ▾



Milestone



## Refactoring the code to allow different storage strategies

```
models: modify workspace_path to store full path #391
Changes from all commits ▾ File filter ▾ Conversations ▾ Jump to ▾ ⚙ ▾

7 docs/openapi.json
221 221 },
222 + {
223 +   "description": "A root path under which the workflow workspaces are stored.",
224 +   "in": "query",
225 +   "name": "workspace_root_path",
226 +   "required": false,
227 +   "type": "string"
228 + },
229 {
230   "description": "JSON object including workflow parameters and workflow specification in JSON format ('yadageschemas.load()' output) with new
231   "in": "body",
232 }

4 reana_workflow_controller/k8s.py
@@ -268,9 +268,7 @@ def build_interactive_jupyter_deployment_k8s_objects(
268 268     command_args = [
269 269         "start-notebook.sh",
270 270         "--notebookApp.base_url='{base_url}'".format(base_url=access_path),
271 271         "--notebook-dir='{workflow_workspace}'".format(
272 272             workflow_workspace=os.path.join(SHARED_VOLUME_PATH, workspace)
273 273         ),
274 274         "--notebook-dir='{workflow_workspace}'".format(workflow_workspace=workspace),
275 275     ]
276 276     if access_token:
277 277         command_args.append(
```

## Input Data



Other directories can now  
be plugged

## Code



**reana**



# EXTRA RESOURCES



[www.reanahub.io](http://www.reanahub.io)



[docs.reanahub.io](http://docs.reanahub.io)



[@reanahub](https://github.com/reanahub)



[info@reanahub.io](mailto:info@reanahub.io)



[@reanahub](https://twitter.com/reanahub)

Contact **reana** team!





# QUESTIONS?

*marcdiazsan@unal.edu.co*  
*@marcdiazsan*