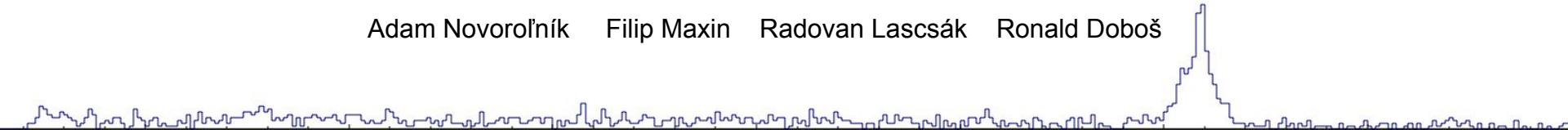
How

FAIR

is scientific data?

Supervised by Tibor Šimko

Adam Novoroľník Filip Maxin Radovan Lascsák Ronald Doboš



Is my data FAIR?

Findable

- indexed and searchable

Accessible

- retrievable by identifiers
- Standard protocols

Interoperable

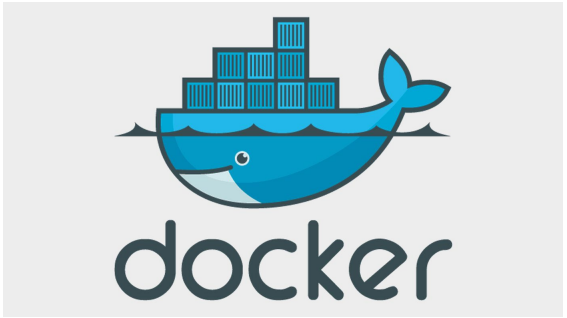
- common vocabularies

Reusable

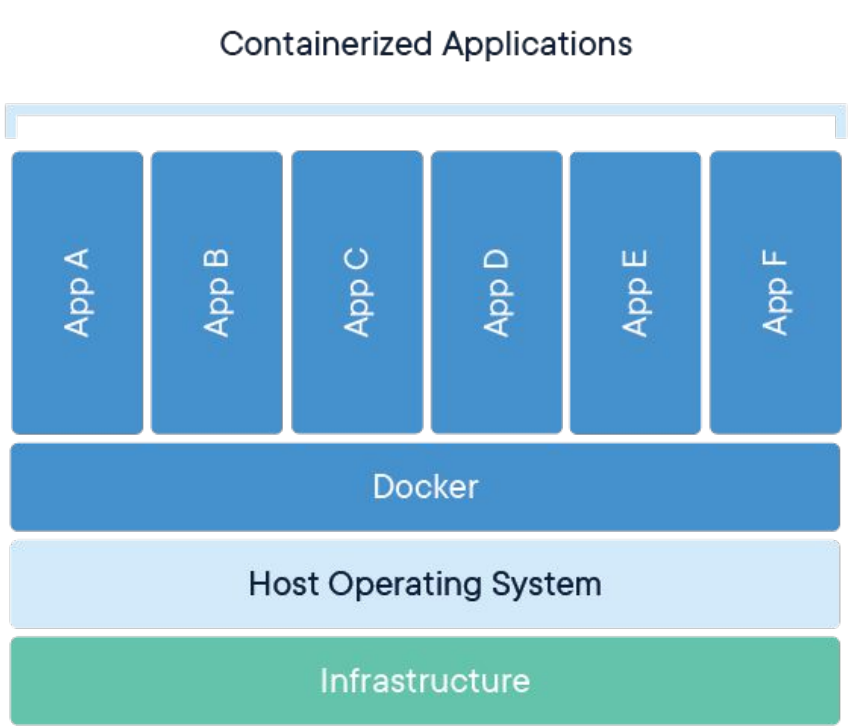
- clear licensing

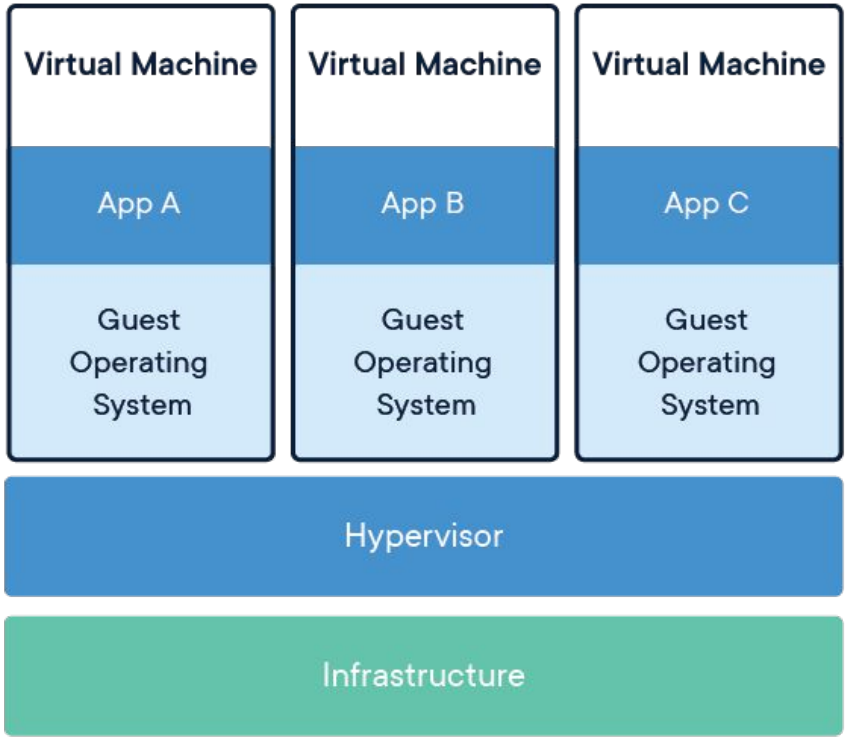
Necessary tools

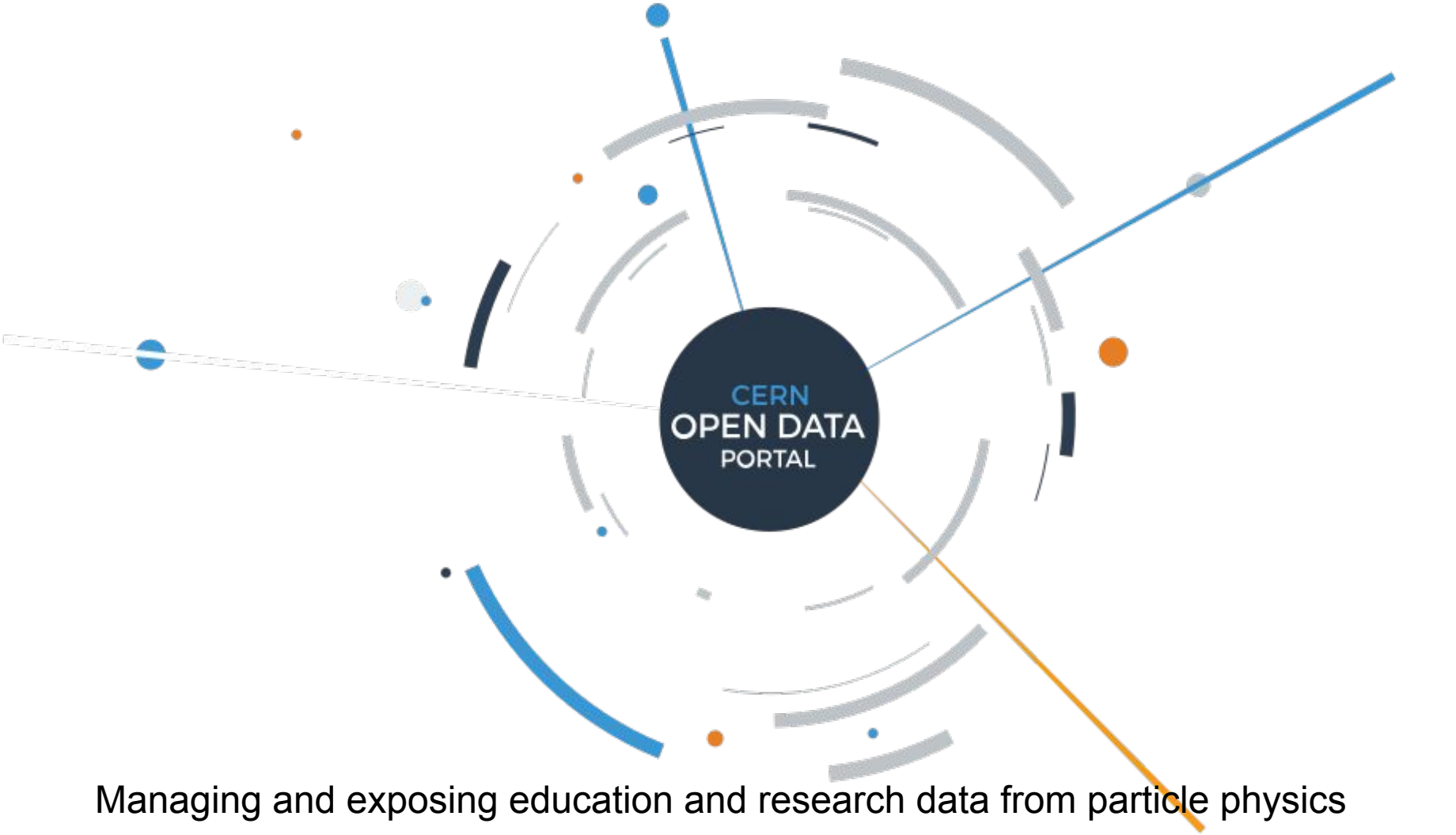
- Shell
- Git
- GitHub
- Docker
- Python



Docker


Containerized Applications
App A
App B
App C
App D
App E
App F
Docker
Host Operating System
Infrastructure


Virtual Machine
Virtual Machine
Virtual Machine
App A
App B
App C
Guest Operating System
Guest Operating System
Guest Operating System
Hypervisor
Infrastructure

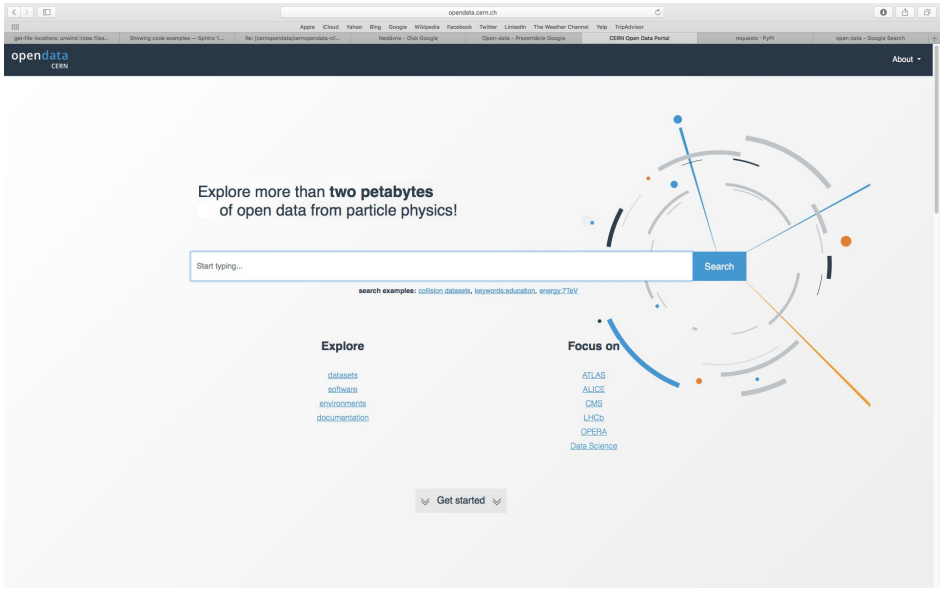


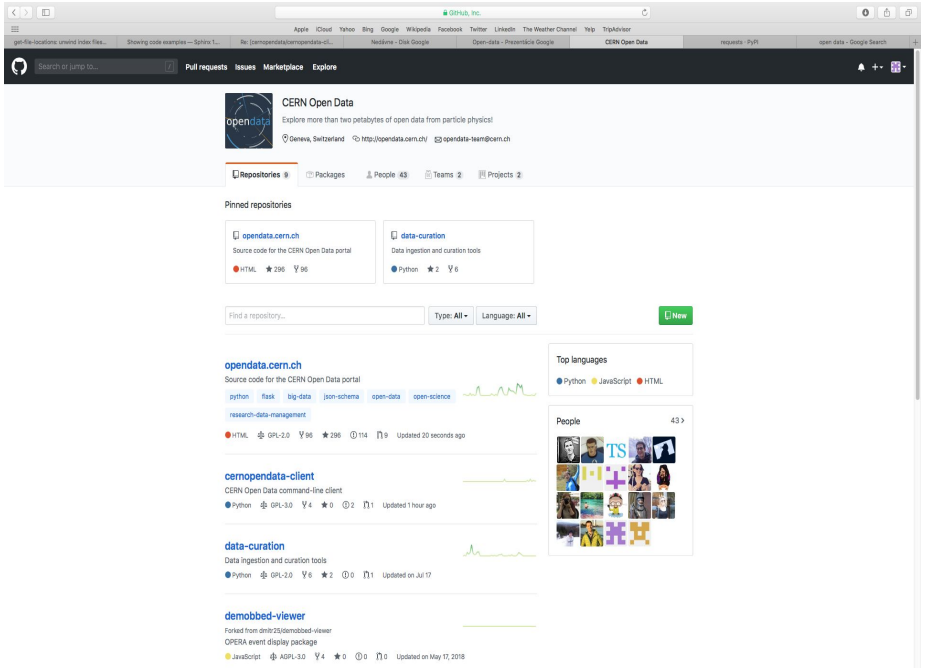
Managing and exposing education and research data from particle physics

Why we use open data?

- Education
- Accessible information
- FAIR science

How do open data sites look like?

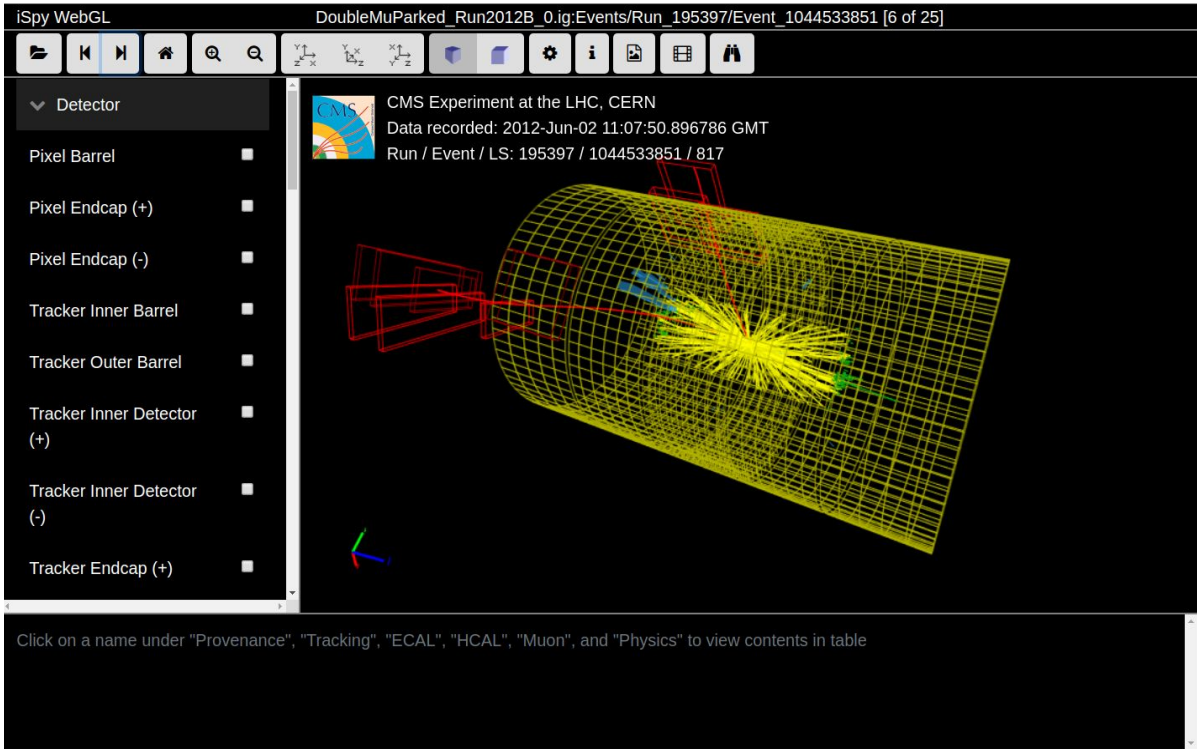




What was our job?

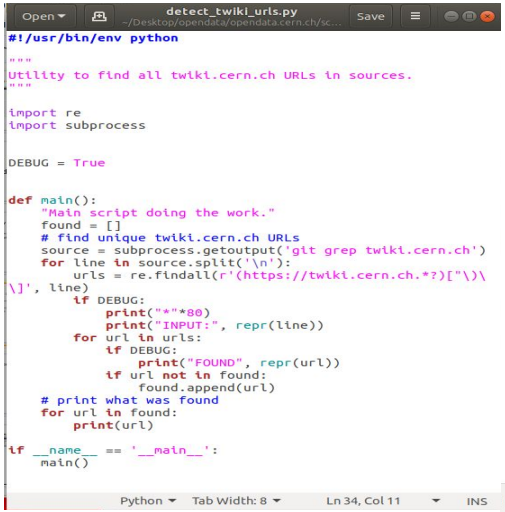
Event Display

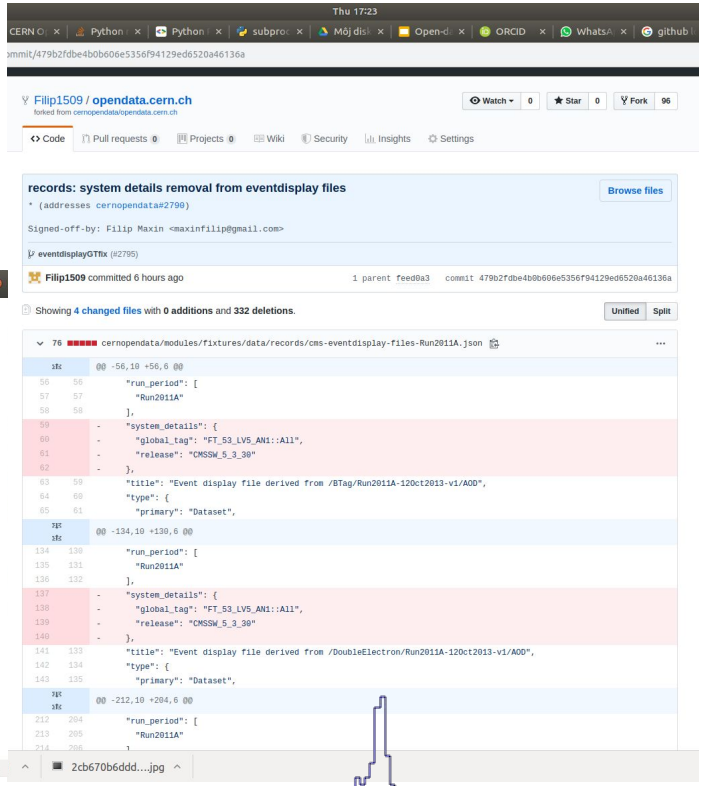
- Interactive event simulations available in opendata



Global tag fix and Twiki URL finder

- Bugfixing: replacement/removal of wrong tags
- New features: script searching for twiki.cern.ch urls





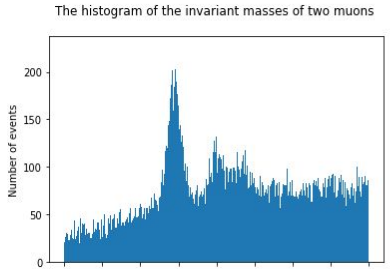
Jupyter notebooks and invariant mass

- Dynamic environment to run and understand code in scientific context

3) Plotting the histogram

Now we can create and plot the histogram of the values of the invariant masses that we got. The histogram shows that in how many events the invariant mass of the muon pair has been in the certain value range. Note that we will use total 500 bins in the histogram, so you won't spot the separate bins because there are so many of them.

In [3]:
```
# Plot the histogram with the function hist() of the matplotlib.pyplot module:
# (http://matplotlib.org/api/pyplot_api.html?highlight=matplotlib.pyplot.hist#matplotlib.pyplot.hist).
# 'Bins' determines the number of the bins used.
plt.hist(invariant_mass, bins=500)

# Name the axises and give the title.
plt.xlabel('Invariant mass [GeV]')
plt.ylabel('Number of events')
plt.title('The histogram of the invariant masses of two muons \n') # \n creates a new line for making the title loo

# Show the plot.
plt.show()
```



Opendata-client

- Used in shell

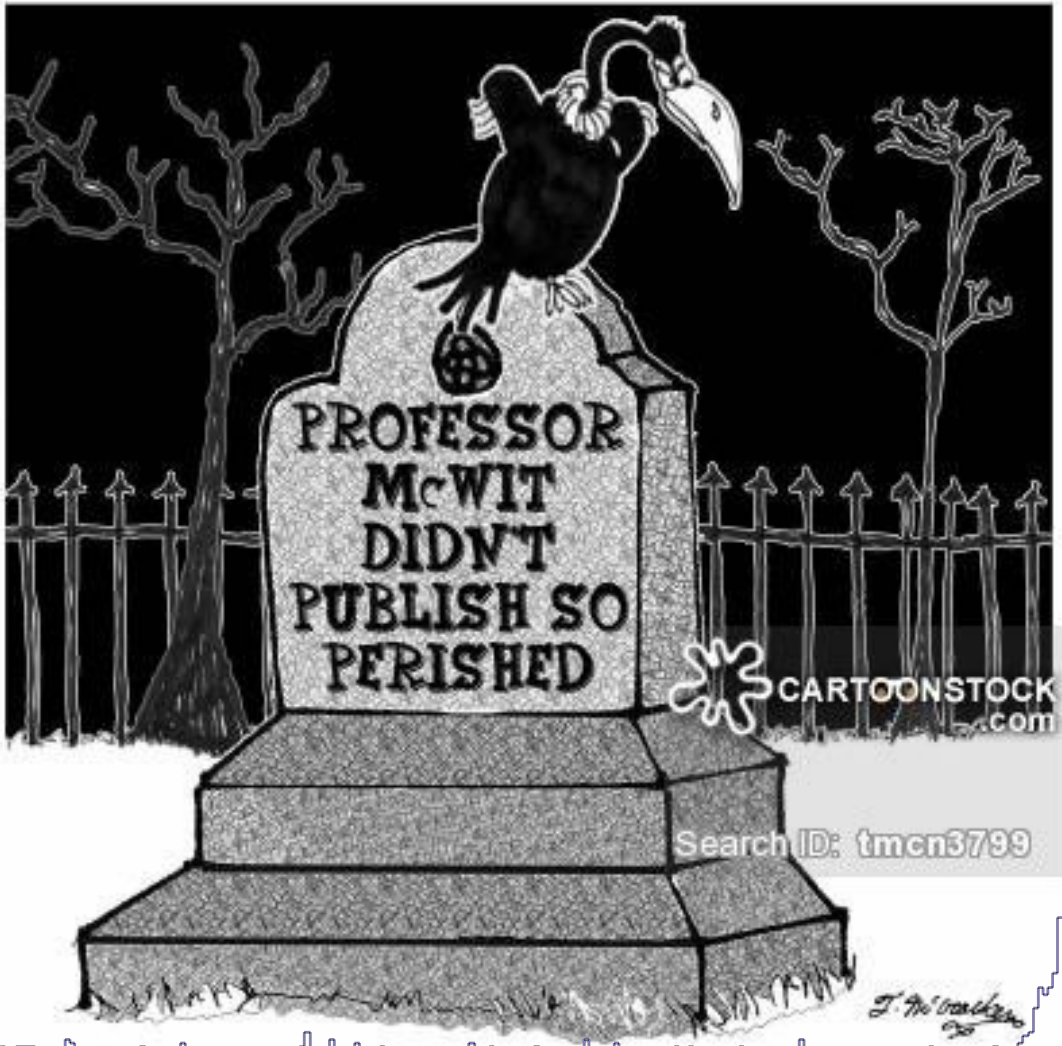
- What can we use it for?
  - cernopendata-client → get-file-locations (list of files belonging to a dataset)
    → get-record (records content by its recid, doi or title..)

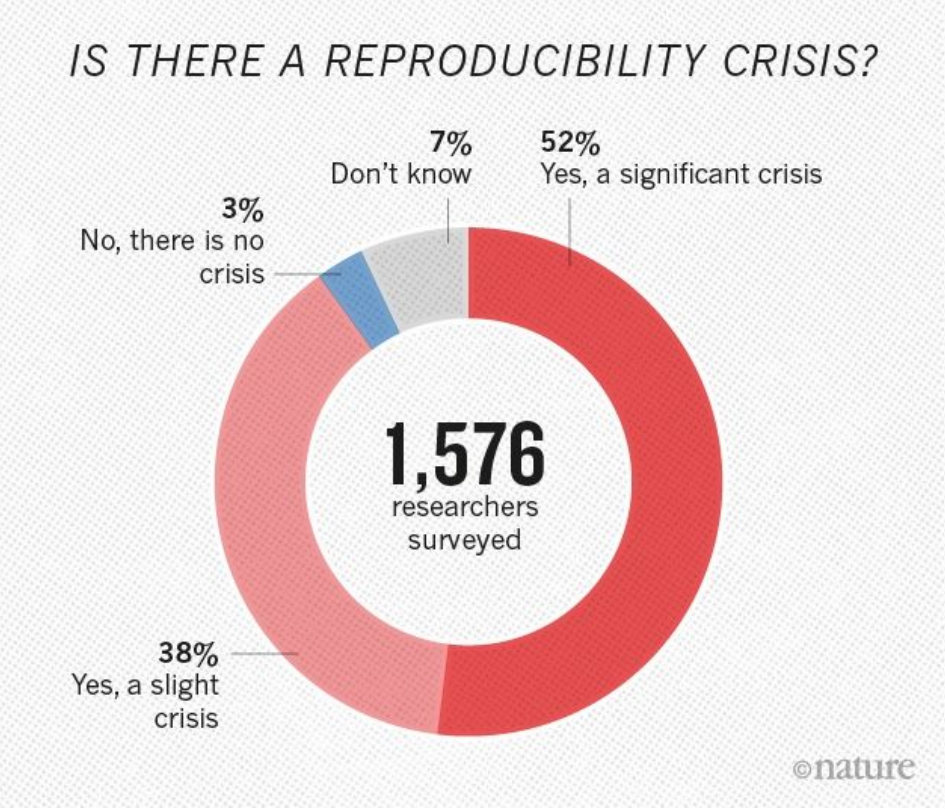
- Were there any issues?

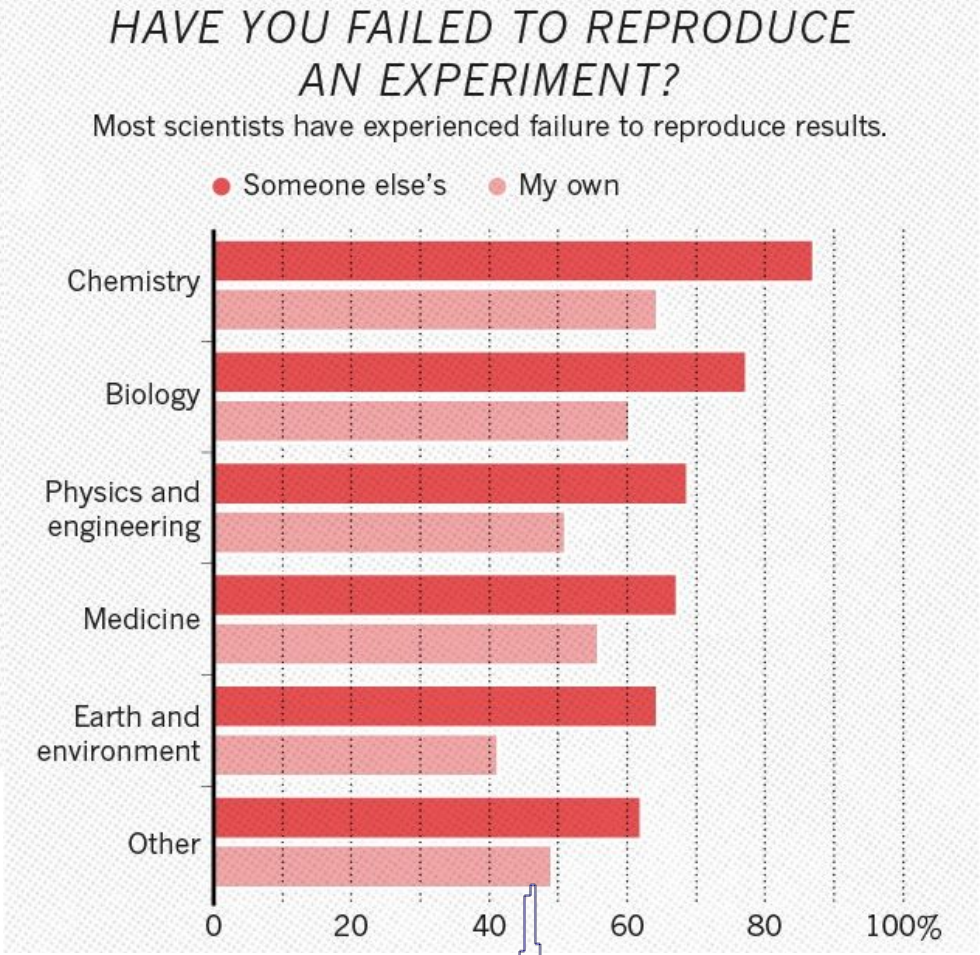
```
(base) rokas-macbook:cernopendata-client anovorol$ git branch
  11-git-fix
  11-gt-fix
  9-unwind-index-files
  docs-authors
* master
(base) rokas-macbook:cernopendata-client anovorol$ git checkout 9-unwind-index-files
Switched to branch '9-unwind-index-files'
(base) rokas-macbook:cernopendata-client anovorol$ git status
On branch 9-unwind-index-files
Untracked files:
  (use "git add <file>..." to include in what will be committed)

        .DS_Store

nothing added to commit but untracked files present (use "git add" to track)
(base) rokas-macbook:cernopendata-client anovorol$ cernopendata-client --help
Usage: cernopendata-client [OPTIONS] COMMAND [ARGS]...

Options:
  --help  Show this message and exit.

Commands:
  get-file-locations  Get a list of files belonging to a dataset.
  get-record          Get records content by its recid, doi or title...
(base) rokas-macbook:cernopendata-client anovorol$ cernopendata-client get-file-locations --helo
Usage: cernopendata-client get-file-locations [OPTIONS]
Try "cernopendata-client get-file-locations --help" for help.

Error: no such option: --helo
(base) rokas-macbook:cernopendata-client anovorol$ cernopendata-client get-file-locations --help
Usage: cernopendata-client get-file-locations [OPTIONS]

  Get a list of files belonging to a dataset.

Options:
  --recid INTEGER         Record ID
  --doi TEXT              Digital Object Identifier.
  --title TEXT            Record title
  --protocol [root|http]  Protocol to be used in links.
  --expand / --no-expand  Expand file indexes? [default=yes]
  --help                  Show this message and exit.
(base) rokas-macbook:cernopendata-client anovorol$ cernopendata-client get-file-locations --recid 1 --no-expand
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0000_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0000_file_index.txt
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0001_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0001_file_index.txt
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0002_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0002_file_index.txt
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0003_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0003_file_index.txt
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0004_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0004_file_index.txt
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0005_file_index.json
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/file-indexes/CMS_Run2010B_BTau_AOD_Apr21ReReco-v1_0005_file_index.txt
(base) rokas-macbook:cernopendata-client anovorol$
```

```
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E267BCB0-9571-E011-AD39-003048670B64.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E2CC8B41-D071-E011-9E39-0025B3E05D46.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E2FD8E5A-9C71-E011-991A-00E081791801.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E2FFED6B-8E71-E011-A97B-002590200868.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E452B761-8E71-E011-8475-003048D4DBFC.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E466F3DF-E271-E011-ABE3-00E081791881.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E4964201-8E71-E011-8776-003048670BAA.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E4A1D547-9F71-E011-A1AC-00E08179184D.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E4F62652-8571-E011-A65E-003048D47A14.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E63473E8-9971-E011-81AC-003048635E3A.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E66006D8-9571-E011-9E40-0025B3E06508.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E66F8283-9171-E011-A60F-002590200A98.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/E6EFFAC9-9871-E011-AE2D-002481E14FCA.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EA60A39C-C171-E011-A4DF-0025B3E06448.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EA90DE1A-EC71-E011-9283-00E081791875.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EA97C535-8971-E011-B6A5-002590200A54.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EABFF6BA-C871-E011-9C52-001A64787078.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EACAD52B-9671-E011-8E0B-002590200B0C.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EC11028C-8971-E011-BBD3-002590200B74.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EC2A0B3E-8D71-E011-A6DA-0025B3E0639C.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EC3DC30D-A771-E011-B268-002590200B78.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EC73B8CC-C671-E011-8728-003048D45FEC.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EC8979F9-8B71-E011-93E2-00E08179182F.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EE431361-B071-E011-88E5-003048D4773A.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EE64929C-8471-E011-8DDB-002481E14E2C.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EE6921FC-EA71-E011-9749-003048670B4A.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EE748B89-B671-E011-AE90-00E08179184B.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EEAA7C07-9471-E011-85E6-003048D46062.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/EED042AB-8671-E011-87DA-0030486709B4.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F0110DE6-9371-E011-B65F-003048D45F8A.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F05BC0E4-9C71-E011-B742-002481E154CE.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F240959C-9271-E011-9B03-003048D47718.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F245A5F7-8B71-E011-A160-002481E150FC.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F280FB63-AC71-E011-BEE0-003048D45FF0.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F28AC7E6-8F71-E011-B9B2-00E0817918B1.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F2C031A6-8A71-E011-A374-003048D47A0C.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F444BC82-8E71-E011-BFC0-003048D45F62.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F473F61C-EA71-E011-B452-0025902009B0.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F4DB1939-8471-E011-A7E4-003048D47792.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F61DDA17-BA71-E011-B26F-0025B3E06522.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F6641C8E-C971-E011-B366-003048D47A40.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F87BF221-8871-E011-9809-001A64789DC8.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F8AD76E9-EB71-E011-8EDB-003048673F24.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F8AF5B43-9771-E011-B859-003048D45F9C.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/F8FFA0DE-8E71-E011-A4E2-00E08178C0AF.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FAA85E49-8471-E011-9DFF-00E08179189F.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FAC46640-B971-E011-A6ED-00E08178C11F.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FC811BCA-9371-E011-95B9-00E08178C01F.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FCB68E6E-8871-E011-B684-002481E14D46.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FE0BF564-9371-E011-9B95-003048673FE6.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FE2F1468-8E71-E011-A76A-0025B3E066A4.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FE3F8388-E471-E011-9377-00E08179189B.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FE579291-C871-E011-8315-0025902008C8.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FEAD9FB9-C971-E011-A592-002590200898.root
root://eospublic.cern.ch//eos/opendata/cms/Run2010B/BTau/AOD/Apr21ReReco-v1/0005/FEE31718-C771-E011-9E69-00E08179179B.root
(base) rokas-macbook:cernopendata-client anovorol$
```

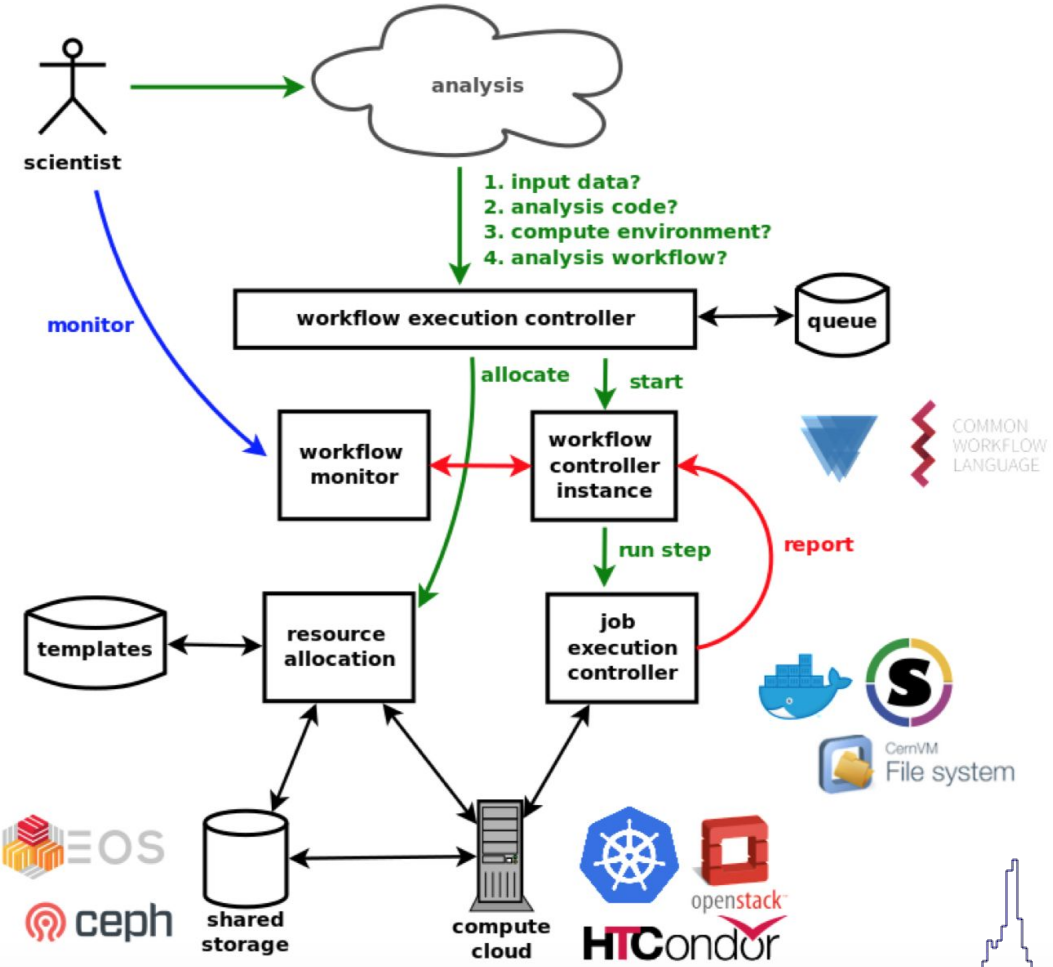
REusable ANAlyses

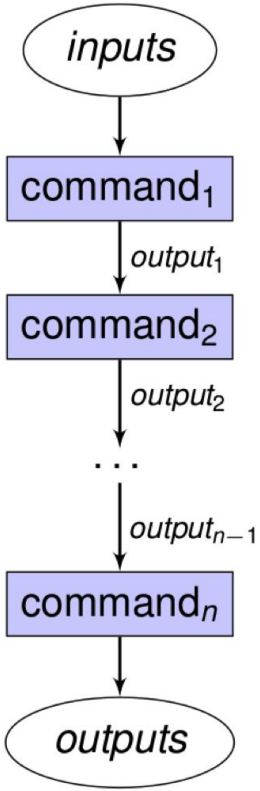

PROFESSOR
McWIT
DIDN'T
PUBLISH SO
PERISHED
CARTOONSTOCK
.com
Search ID: tmcn3799

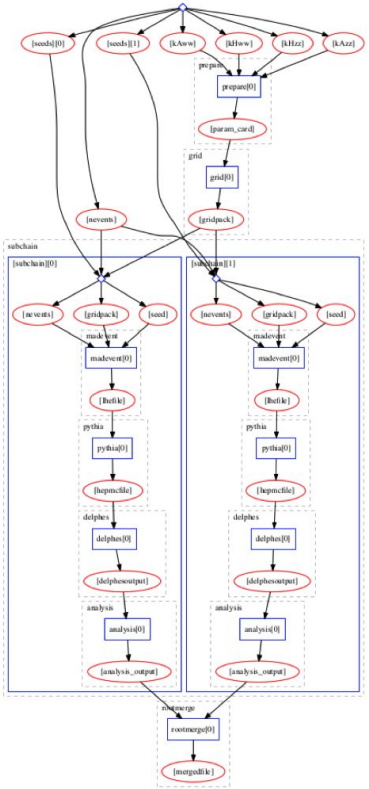


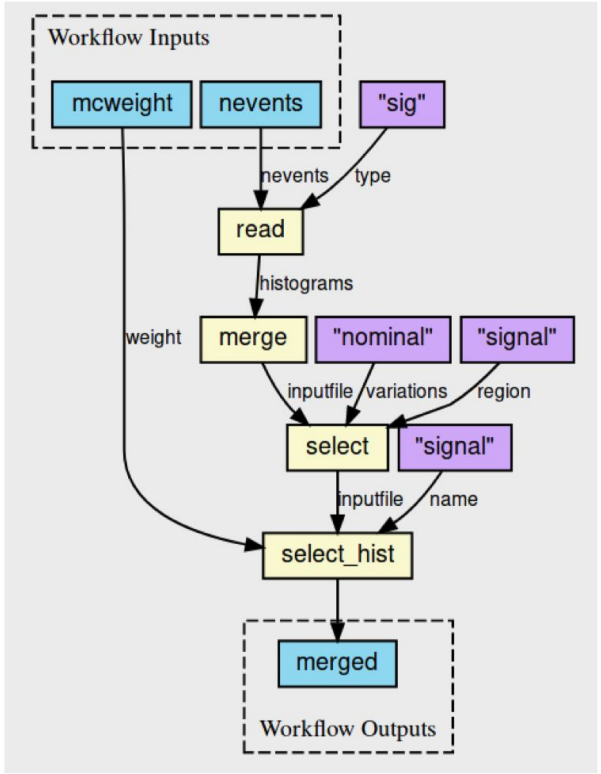


https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970


analysis
scientist
1. input data?
2. analysis code?
3. compute environment?
4. analysis workflow?
monitor
workflow execution controller
queue
allocate
start
workflow monitor
workflow controller instance
COMMON WORKFLOW LANGUAGE
run step
report
templates
resource allocation
job execution controller
CernVM File system
EOS
ceph
shared storage
compute cloud
openstack
HTCondor







Serial

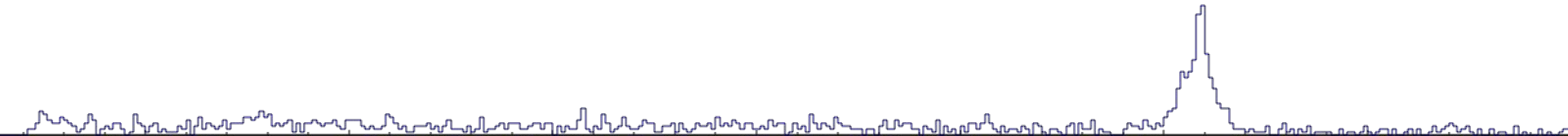
Yadage

CWL

Our contribution - REANA example

1. Take data from 

2. Use existing code to produce dimuon spectrum.
3. Containerize the code using docker
4. Produce a REANA example on GitHub.



Invariant mass

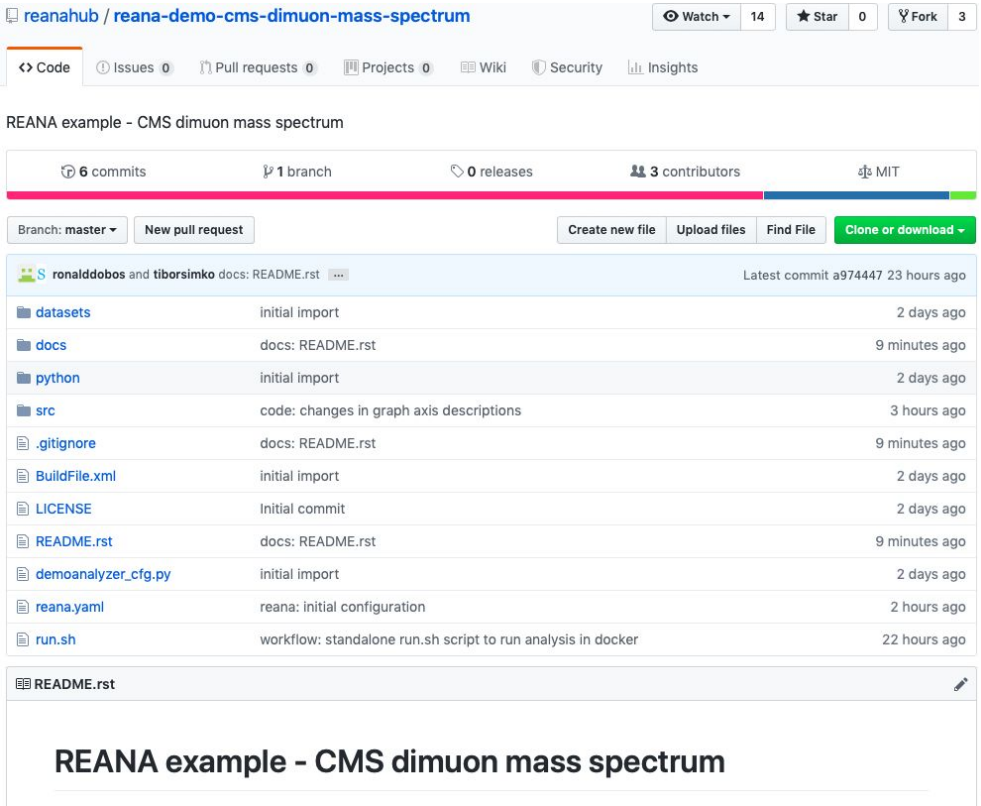
$$X \equiv \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \quad X' = \Lambda X, \quad ds^2 \equiv dX \cdot dX = c^2(t)^2 - dx^2 - dy^2 - dz^2 \text{ is invariant}$$

$$ds = cd\tau, \quad d\tau = \frac{ds}{c} \text{ is invariant}$$

$$U \equiv \frac{dX}{d\tau} = \frac{dt}{d\tau}\begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} = \gamma \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}, \quad U \cdot U = c^2$$

$$P \equiv mU = \begin{pmatrix} m\gamma c \\ m\gamma \mathbf{u} \end{pmatrix} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}, \quad P \cdot P = m^2c^2 = (E/c)^2 - \mathbf{p}^2$$

Final result



reanahub / reana-demo-cms-dimuon-mass-spectrum

Watch 14 | Star 0 | Fork 3

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Security | Insights

REANA example - CMS dimuon mass spectrum

6 commits | 1 branch | 0 releases | 3 contributors | MIT

Branch: master | New pull request | Create new file | Upload files | Find File | Clone or download

ronalddobos and tiborsimko docs: README.rst — Latest commit a974447 23 hours ago

datasets	initial import	2 days ago
docs	docs: README.rst	9 minutes ago
python	initial import	2 days ago
src	code: changes in graph axis descriptions	3 hours ago
.gitignore	docs: README.rst	9 minutes ago
BuildFile.xml	initial import	2 days ago
LICENSE	Initial commit	2 days ago
README.rst	docs: README.rst	9 minutes ago
demoanalyzer_cfg.py	initial import	2 days ago
reana.yaml	reana: initial configuration	2 hours ago
run.sh	workflow: standalone run.sh script to run analysis in docker	22 hours ago

README.rst

REANA example - CMS dimuon mass spectrum

We start by creating a reana.yaml file describing the above analysis structure with its inputs, code, runtime environment, computational workflow steps and expected outputs:

We can now install the REANA command-line client, run the analysis and download the resulting ROOT file containing plots:

```
$ # create new virtual environment
$ virtualenv ~/.virtualenvs/myreana
$ source ~/.virtualenvs/myreana/bin/activate
$ # install REANA client
$ pip install reana-client
$ # connect to some REANA cloud instance
$ export REANA_SERVER_URL=https://reana.cern.ch/
$ export REANA_ACCESS_TOKEN=XXXXXXX
$ # create new workflow
$ reana-client create -n my-analysis
$ export REANA_WORKON=my-analysis
$ # upload input code and data to the workspace
$ reana-client upload
$ # start computational workflow
$ reana-client start
$ # ... should be finished in about 1 minute
$ reana-client status
$ # download output root file with generated plots
$ reana-client download
```

Please see the REANA-Client documentation for more detailed explanation of typical `reana-client` usage scenarios.

Contributors

The list of contributors in alphabetical order:

- Radovan Lascsak
- Ronald Dobos
- Tibor Simko

Thank you for your attention

Yes I know what REANA means

R - *I*
E - *have*
A - *no*
N - *idea*
A – *send help*

Yes I'm HAPPY

H
A
P
P
Y Sad

2happy4meirl