



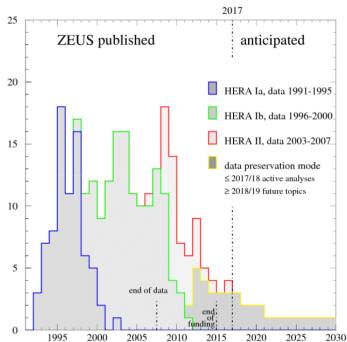
# REANA: reproducible research data analysis platform

Tibor Šimko

@tiborsimko

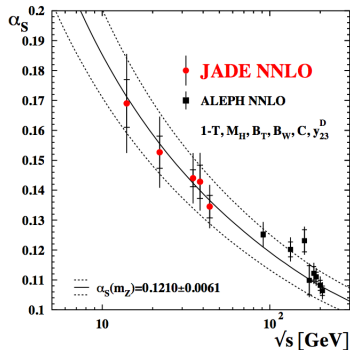
EP-SFT Group Meeting · CERN · 4 June 2018

# Long-term value of data!



Achim Geiser <https://indico.cern.ch/event/588219>

Collaborations publish papers even  $\sim 15$  years after data taking ends.



DPHEP <https://arxiv.org/abs/1205.4667>

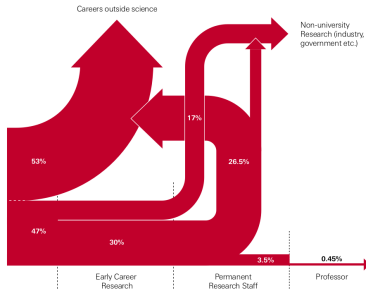
JADE data (1979–1986) still unique even  $\sim 35$  years later.

# Long-term value of knowledge?



## CMS collaboration

Experimental physics done by groups of  $\sim 3000$  physicists.



## Career after PhD

THE ROYAL SOCIETY

High turnover of young researchers.

# LHCb members of 2010 in 2017



Patrick Koppenburg  
@PKoppenburg

Following

Particle Physics author lists change with time. Here that of the first @LHCbExperiment paper in 2010. Violet: still in LHCb. Blue: left LHCb

LHCb Collaboration: R. Aaij, C. Abellan Beteta, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, M. Agari, Z. Ajlouni, J. Albrecht, F. Alessio, M. Alexander, M. Allson, P. Alvarez Cartelle, A.A. Alves Jr, S. Amato, Y. Amhis, J. Amoraa, J. Anderson, R. Antunes Nobrega, R. Appleby, O. Aquines Gutierrez, A. Arefeyev, L. Arrabito, M. Artuso, E. Aslanides, G. Aurier, S. Bachmann, Y. Bagaturia, D.S. Bailey, V. Balagura, V. Baldini, G. Barber, C. Barham, R.J. Barlow, S. Barsuk, S. Basildare, A. Bates, C. Bauer, Th. Bauer, A. Bay, I. Bediaga, T. Bellunato, K. Belous, I. Belyaev, M. Benayoun, G. Bencivini, R. Bernert, P.P. Bernhard, M.-O. Bettler, M. van Beuzekom, J.H. Bibby, S. Bifani, A. Bizzeti, P.M. Bjornstad, T. Blake, F. Blanc, C. Blanks, J. Blouw, S. Blusk, A. Bobrov, V. Bocci, B. Bochin, E. Bonaccorsi, A. Bondar, N. Bondar, W. Bonifazi, S. Borghi, A. Borgia, E. Bos, T.J.V. Bowcock, C. Bozzi, T. Brambach, J. van den Brand, L. Brarda, J. Bressieux, S. Brissbane, M. Britsch, N.H. Brook, H. Brown, S. Brusa, A. Buechler-Germann, A. Bursche, J. Buytaert, S. Cadeddu, J.M. Caicedo Carvajal, O. Callot, M. Calvi, M. Calvo Gomez, A. Camboni, W. Cameron, L. Camilleri, P. Campana, A. Carbone, G. Carbone, R. Cardinale, A. Cardini, J. Carroll, L. Carson, K. Carvalho Akiba, G. Casse, M. Cattanéo, B. Chada, M. Charles, Ph. Charpentier, J. Cheng, N. Chiapolini, A. Chlopik, J. Christiansen, P. Ciambone, X. Cid Vidal, P.J. Clark, P.E.L. Clarke, M. Clemencia, H.V. Cliff, J. Clozier, C. Coca, V. Coco, J. Cogan, P. Collins, A. Comerma-Montells, F. Constantin, G. Conti, A. Contu, P. Cooke, M. Coombes, B. Corajod, G. Corfi, G.A. Cowan, R. Currie, B. D'Agnone, C. D'Ambrosio, I. D'Antonio, W. Da Silva, E. Danev, P. David, I. De Boer, D. De Capua, M. De Cian, F. De Lorenzi, J.M. De Miranda, L. De Paula, P. De Simone, D. Decamp, G. Decroise, H. Degaudenzi, M. Delgado, M. Delissenroth, L. Del Buono, C.J. Densham, C. Deplano, O. Deschamps, F. Deffori, J. Dickens, H. Dijkstra, M. Dima, S. Donlay, P. Dorman, D. Dossett, A. Dovbnya, R. Dumps, F. Dupret, L. Dwyer, R. Dzhelezyn, C. Eames, S. Easo, U. Egede, V. Egorychev, S. Eidelman, D. van Eijk, F. Eisele, S. Eisenhardt, L. Eklund, D.G. d'Enterria, D. Esperante Pereira, L. Esteve, E. Fanchini, C. Färber, G. Farrell, C. Farinelli, S. Farry, V. Favre, G. Felici, V. Fernandez Albor, M. Ferro-Luzzi, S. Filippov, C. Fitzpatrick, W. Fliegell, F. Fontanelli, C. Forti, R. Forty, C. Fournier, B. Franek, M. Frank, C. Frei, M. Brosini, J.L. Furgueleiro Pazos, S. Furcas, A. Gallias Torreira, D. Galli, M. Gandelman, P. Gandini, Y. Gao, J.-C. Garnier, L. Garido, D. Gascon, C. Gaspar, A. Gaspar De Valenzuela Cue, J. Gassner, N. Gauvin, P. Gavillet, M. Gersabeck, T. Gershon, J. Ghez, V. Gibson, Yu. Giiitsky, V.V. Gligorov, C. Göbel, D. Golubkov, A. Golubvin, A. Gomes, G. Gong, H. Gong, H. Gordon, M. Grabalosa Gándara, V. Gracco, R. Gracliani Diaz, L.A. Granado Cardoso, E. Graugés, G. Graziani, A. Grecu, S. Gregson, G. Guener, B. Gué, E. Gushchyn, Yu. Guz, Z. Guzik, T. Gys, G. Haefliger, S.C. Haines, T. Hampson, S. Hansmann-Menzemer, R. Harji, N. Harrow, P.F. Harrison, J. He, K. Hennessy, P. Henrard, J.A. Hemandora Morata, E. van Heerwijnen, A. Heiche, E. Hicks, H.J. Hillie, W. Hofmann, K. Holubey, P. Hopche, W. Hulsbergen, P. Hunt, T. Huse, R.S. Huston, D. Hutchcroft, F. Iacoangeli, V. Iakovenko, C. Iglesias Escudero, C. Ilgner, J. Imong, R. Jacobsson, M. Jahjah Khechik, O. Jamet, E. Jans, F. Jansen, P. Jaton, B. Jean-Marie, M. John, D. Johnson, C.R. Jones, B. Jost, F. Kapusta, T.M. Karbach, A. Kashchuk, S. Katvasz, J. Kaveney, U. Kerzel, T. Ketel, A. Keune, S. Khallil, B. Khanji, Y.M. Kim, M. Knoch, S. Koblit, A. Kocny, P. Koppenburg, M. Korolev, A. Kozlikin, L. Kravchuk, R. Kristic, G. Krocker, P. Krokovy, F. Kruse, K. Kruszec, M. Kucharczyk, I. Kudryashov, S. Kukulski, R. Kumar, T. Kvaratskheliya, Y.N. Lathi, D. Lacarera, A. Lai, R.W. Lambert, G. Lanfranchi, C. Langenbruch, T. Latham, R. Le Gac, J.-P. Lees, R. Lefevre, A. Leflat, J. Lefrançois, F. Lehner, M. Lenzi, O. Leroy, T. Lesiak, L. Li, Y.Y. Li, Y.Y. Li, L.L. Li, G. Li, J. Libby, M. Livan, R. Lindner, S. Lindsey, C. Linn, B. Liu, G. Liu, S. Ljochner, H. Lopes, E. Lopez Anadon, N. Lopez-March, P. Lowerre, J. Luisler, B. Mcharek, F. Macheof, I.V. Machikhiliani, F. MacIac, O. Maev, J. Magni, A. Mair, S. Maide, R.M.D. Marun, R. Marica, G. Mancini, L.M. Mangatvalay, U. Marconi, R. Märki, J. Marks, G. Martelliotti, A. Martini, L. Martin, D. Martin, D. Martinez Santos, A. Massafferri, Z. Mathe, C. Matteucci, V. Maurice, B. Maynard, A. Mazuren, G. McGregor, R. McHugh, C. Moisan, A. Moke, J. Morkel, M. Morin, R. Menni, F.C.D. Melica, S. Migloranni, M.-N. Minard, G. Moine, S. Monteil, D. Moran, J. Morand, J.V. Morris, J. Moscicki, J. Mountain, I. Mous, F. Muheim, R. Mueser, F. Murat, B. Muryn, M. Mury, J. Myhring, S.M. Naik, T. Nakada, R. Nandakumar, J. Nardulli, A. Nasrati, M. Nedos, M. Needham, N. Neufeld, P. Neustroev, M. Nicci, L. Nicolas, S. Nies, N. Nikitin, A. Noor, A. Obiakovska-Mucha, V. Obraztsov, S. Oggero, O. Okhtemirov, R. Oldeman, M. Okada, A. Ostankov, J. Palacios, M. Palutan, J. Panman, A. Papadimitriou, A. Papanestis, M. Papageorgiou, C. Parkes, S.J. Parkinson, G. Passalunghi, G.D. Patel, M. Patel, S.K. Paterson, G.N. Patrick, C. Patrignani, E. Pausa, C. Pausa (ChloéZaida), C. Pavesi (Nicoretta), A. Pazos Alvarez, A. Pellegrino, G. Penso, M. Pepe Altarelli, S. Perazzini, D.L. Perego, E. Perez Trigo, A. Pérez-Calero Yzquierdo, F. Perini, G. Pessina, A. Petrella, A. Petrucci, E. Picatoste Ochoa, B. Pie Valis, D. Piedrograsi, B. Pietrzyk, D. Pinci, S. Playfer, M. Pio Cassauas, M. Polzella, G. Polok, A. Poluektov, E. Polycarpo, D. Popov, B. Popovic, S. Poss, C. Potterat, A. Powell, S. Pozzi, T. du Pree, V. Pugatch, A. Puig Navaro, W. Qian, J.H. Rademacker, B. Rakotomiramanana, I. Raniuk, G. Raven, S. Redford, W. Reese, A.C. dos Reis, S. Ricciardi, J. Riera, K. Rinnert, D.A. Roa Romero, D. Robber, E. Rodrigues, F. Rodrigues, C. Rodriguez Cobo, P. Rodriguez Perez, G.J. Rogers, V. Romanovsky, E. Rondan Sanabria, M. Rosello, G. Rosopaje, J. Rouniev, I. Roy, T. Ruf, H. Ruiz, C. Rummel, T. Rusinov, G. Sabatino, J.L. Saborido Silva, N. Sagidova, P. Sailer, B. Sallita, T. Sakhelashvili, C. Salzmann, A. Sanabria Varela, M. Sannino, R. Santacesaria, R. Santinelli, E. Santovetti, M. Sapunov, A. Sarici, C. Satriano, A. Satta, T. Savijide, M. Savin, P. Schacht, M. Schiller, S. Schleich, M. Schmelling, B. Schmidt, O. Schneider, T. Schneider, A. Schopper, M.-H. Schune, R. Schwemmer, A. Scubba, M. Seco, A. Semennik, K. Senderovska, N. Serra, J. Serrano, B. Shoar, M. Shapkin, I. Shatalov, P. Shatalov, Y. Shchegolev, Y. Shchegolev, T. Shears, L. Shekhitman, V. Shevchenko, A. Shires, S. Sigurdsson, E. Simioni, H.P. Skottowe, T. Skwarnicki, N. Smaie, A. Smith, A.C. Smith, N.A. Smith, K. Sobczak, F.J.P. Soler, A. Solomin, P. Somogy, F. Soomro, B. Souza De Paula, B. Spaan, A. Sparks, E. Spradlenkov, P. Spradlin, A. Srednicki, F. Stagni, S. Stahl, S. Steiner, O. Steinrück, O. Steynakins, S. Stoica, S. Stone, B. Storaci, U. Straumann, N. Styles, M. Szczekowski, P. Szczepka, T. Szumlanski, S. T'Jampens, E. Tarlowski, E. Tenedores, H. Terrier, F. Teubert, C. Thomas, E. Thomas, J. van Tilburg, V. Tisserand, M. Tobin, S. Topp-Joergensen, S.M.T. Tran, S. Traynor, U. Trunk, A. Tsaregordtsev, N. Tuning, A. Ukleja, O. Ullaland, U. Uwer, V. Vagnoni, G. Valentí, A. Van Lysebeeten, R. Vandenbogaert, P. Vazquez Regueiro, S. Vecchi, J.J. Veitch, M. Veltri, K. Vervink, B. Viad, L. Videau, A. Vieira, X. Vilasis-Cardona, J. Visniakov, A. Vollhardt, D. Volynskiy, D. Voong, A. Vorobyev, An. Vorobyev, H. Voss, K. Wacker, S. Wabemuhji, G. Wang, D.R. Ward, A.D. Webber, D. Webster, D. Wiedner, L. Wiggers, G. Wilkinson, M.P. Williams, M. Williams, F.F. Wilson, J. Wishahi, M. Witek, W. Witzeling, M.L. Woodward, S.A. Wotton, K. Wylie, Y. Xie, F. Xing, Z. Yang, G. Yvelis Smit, R. Young, O. Yushchenko, M. Zeng, L. Zhang, Y. Zhang, A. Zhelezov, E. Zverev (collapsable list)

(Submitted on 18 Aug 2010 (v1), last revised 18 Sep 2010 (this version, v2))



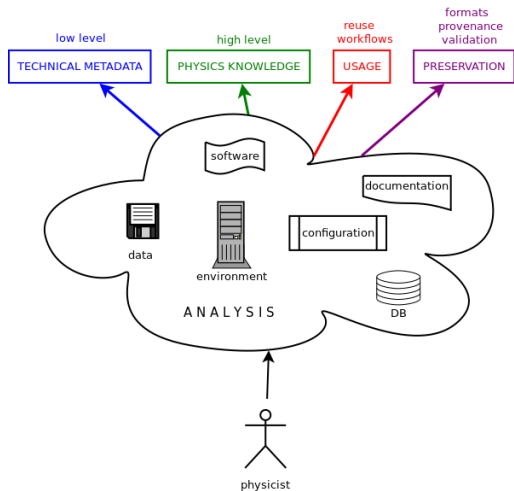
<https://twitter.com/PKoppenburg/status/918097093616652293>

# CERN Analysis Preservation

- A platform for **preserving knowledge** and **assets** of an individual physics analysis.
- Capturing the elements needed to **understand** and **reuse** an analysis even several years later:
  - ✓ data
  - ✓ software
  - ✓ environment
  - ✓ workflow
  - ✓ context
  - ✓ documentation
- Applying standard **collaboration access restrictions**

*Developed by CERN IT and CERN SIS in close collaboration  
with LHC experiments*

# Pillar 1: Describe

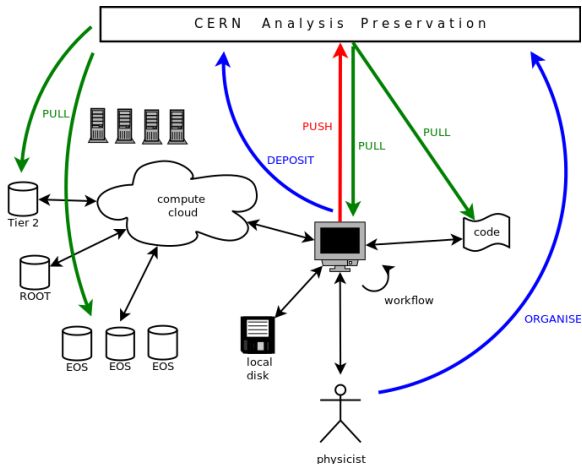


INVENIO

- JSON Schema
- W3C DCAT
- domain-specific fields

Structuring knowledge behind research data analysis.

# Pillar 2: Capture

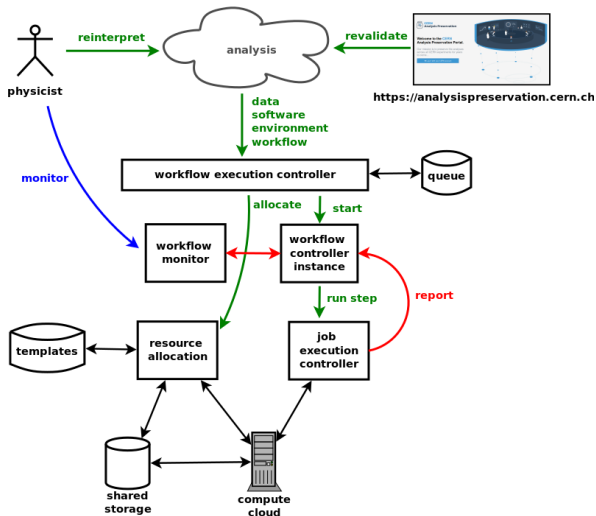


INVENIO

- datasets:  
local storage,  
cloud storage
- software:  
Git, SVN
- information:  
DBs, TWiki,  
SharePoint
- protocols:  
HTTP, XRootD

Taking consistent snapshot of analysis assets at a certain time.

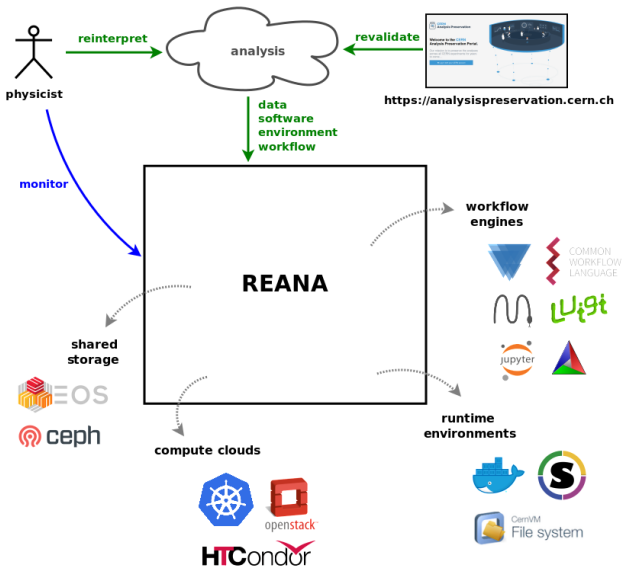
# Pillar 3: Reuse



Instantiating preserved analysis on the cloud.



# REANA = RE usable ANA lyseses



# Four questions

## 1 Input data

What is your input data?

- input files
- input parameters

## 3 Compute environment

What is your environment?

- operating system
- database calls

## 2 Analysis code

Which code analyses it?

- software frameworks
- user code

## 4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

# Simple example

```
Region,1500,1600,1700,1750,1800,1850,1900,1950,1999,2008,2010,2012,2050,2150
World,100,100,100,100,100,100,100,100,100,100,100,100,100,100
Africa,18.8,19.7,15.5,13.4,10.9,8.8,8.1,8.8,12.8,14.5,14.8,15.2,19.8,23.7
Asia,53.1,58.4,63.9,63.5,64.9,64.1,57.4,55.6,60.8,60.4,60.4,60.3,59.1,57.1
Europe,18.3,19.1,18.3,20.6,20.8,21.9,24.7,21.7,12.2,10.9,10.7,10.5,7.5,3
Latin America and the Caribbean,8.5,1.7,1.5,2.2,2.5,3,4,5,6,6,8,5,8,6,8,6,9,1,9,4
Northern America,0.7,0.5,0.3,0.3,0.7,2.1,5.6,8,5,1,5,5,5,4,4,4,1
Oceania,0.7,0.5,0.4,0.3,0.2,0.2,0.4,0.5,0.5,0.5,0.5,0.5,0.5,0.5
```

## 1 input: CSV file

```
FROM centos:7
RUN yum install -y epel-release
RUN yum install -y \
    gcc \
    python-devel \
    python-pip
RUN pip install ipython==5.0.0 jupyter==1.0.0
ADD world_population_analysis.ipynb /code/
ADD World_historical_and_predicted_populations_in_percentage.csv /code/
WORKDIR /code
CMD ["jupyter", "nbconvert", "world_population_analysis.ipynb"]
```

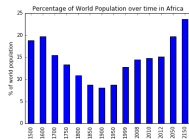
## 3 environment: CentOS7, IP5

### Regional Analysis

We'll start with a histogram depicting the evolution of a specific region's portion of the world population, in percentage.


```
In [6]: def histogram_by_region(region):
        local_pop=pop[['Region', str(region)]].groupby('Region').sum()
        plot=local_pop.plot(kind='bar', legend=None, title='Percentage of World Population over time in '+ str(region))
        plot.set_ylabel('% of world population')
        plot.set_xlabel('')

In [7]: histogram_by_region('Africa')
```

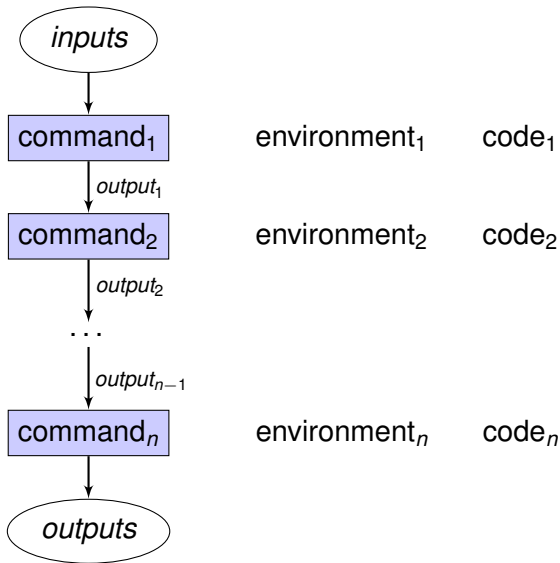


## 2 code: Jupyter notebook

## 4 workflow: papermill ...

 <https://github.com/reanahub/reana-demo-worldpopulation>

# Simple workflows



# Example: ROOT RooFit

workflow:

type: serial

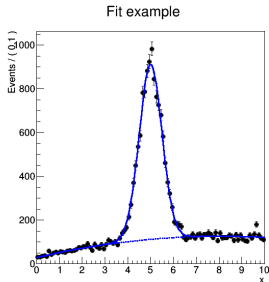
steps:


- environment: reanahub/reana-env-root6

commands:

- root -b -q './code/gendata.C(20000,"data.root")'

- root -b -q './code/fitdata.C("data.root","plot.png")'



 <https://github.com/reanahub/reana-demo-root6-roofit/>

# User perspective: reana.yaml

- analysis structure described via `reana.yaml`

```
version: 0.2.0
code:
  files:
    - code/mycode.py
inputs:
  files:
    - inputs/mydata.csv
parameters:
  myparameter: myvalue
environments:
  - type: docker
    image: johndoe/mypython:1.0
workflow:
  type: cwl
  file: workflow/myworkflow.cwl
outputs:
  files:
    - outputs/myplot.png
```

# User perspective: CLI client

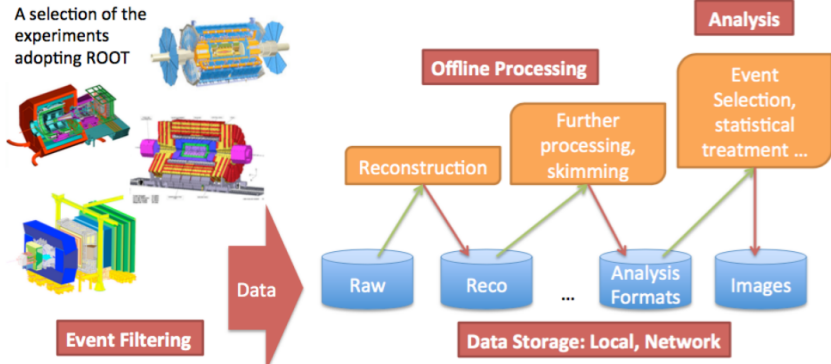
## ■ select REANA cloud

```
$ export REANA_SERVER_URL=https://reana.cern.ch/
```

## ■ run your workflows

```
$ # install reana-client
$ mkvirtualenv reana-client -p /usr/bin/python2.7
$ pip install reana-client
$ reana-client ping
$ # create new workflow
$ export REANA_WORKON=$(reana-client workflow create)
$ # upload runtime code and inputs
$ reana-client code upload ./code/*
$ reana-client inputs upload ./inputs/*
$ # start workflow and check progress
$ reana-client workflow start
$ reana-client workflow status
$ # download outputs
$ reana-client outputs list
$ reana-client outputs download myplot.png
```

# HEP data analysis pipelines



D. Krücker *et al* <https://indico.desy.de/indico/event/18343>



# Example: CMS open data

The screenshot shows the Open Data Portal interface for a specific dataset. At the top, there is a search bar and the 'open data CERN' logo. The main content area is titled 'MinimumBias primary dataset sample in RAW format from RunA of 2011 (from /MinimumBias/Run2011A-v1/RAW)'. Below the title, there is a citation and a list of tags including 'Dataset', 'Collision', 'CMS', 'Collision energy 7TeV', and 'Accelerator CERN/LHC'. The 'Description' section states it is a sample from the MinimumBias primary dataset in RAW format from RunA of 2011, with a run period of 160957. The 'Notes' section mentions that the dataset contains one run from 2011 RunA and provides a link to a list of validated runs. The 'Characteristics' section lists 1913190 events and 72 files totaling 248.8 GB. The 'System Details' section includes the global tag 'FT\_53\_LV5\_AN1' and the recommended release for reconstruction 'CMSSW\_5\_3\_32'. The 'How were these data selected?' section explains that events were selected based on the presence of low-energy particles (soft-QCD events).

open data  
CERN

Search

About

## MinimumBias primary dataset sample in RAW format from RunA of 2011 (from /MinimumBias/Run2011A-v1/RAW)

/MinimumBias/Run2011A-v1/RAW, CMS collaboration

Cite as: CMS collaboration (2017). MinimumBias primary dataset sample in RAW format from RunA of 2011 (from /MinimumBias/Run2011A-v1/RAW). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.I8HN.DF32

Dataset Collision CMS Collision energy 7TeV Accelerator CERN/LHC

### Description

A sample from MinimumBias primary dataset in RAW format from RunA of 2011. Run period 160957.

### Notes

This dataset contains one run from 2011 RunA. The list of validated lumi sections, which must be applied to all analyses, can be found in

[CMS list of validated runs Cert\\_160404-180252\\_7TeV\\_ReRecoNov08\\_Collisions11\\_JSON.txt](#)

### Characteristics

Dataset: **1913190** events **72** files **248.8 GB** in total

### System Details

Global tag: FT\_53\_LV5\_AN1

Recommended release for reconstruction: CMSSW\_5\_3\_32

### How were these data selected?

Events stored in this primary dataset were selected because of the presence of low-energy particles (soft-QCD events).

<http://opendata.cern.ch/record/35>

# Example: CMS AOD production

The screenshot shows a web page from the OpenData CERN portal. At the top, there is a dark navigation bar with the 'opendata CERN' logo on the left, a search bar in the center, and an 'About' link on the right. The main content area has a white background and is framed by a light gray border. The title of the record is 'Validation code for reprocessing AOD from 2011 MinimumBias RAW sample', with the author 'Lassila-Perini, Kati' listed below it. A citation is provided: 'Cite as: Lassila-Perini, Kati; (2017). Validation code for reprocessing AOD from 2011 MinimumBias RAW sample. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.99j.BGV4'. Below the citation are several category tags: 'Software', 'Validation', 'CMS', and 'Accelerator CERN-LHC'. The 'Description' section explains that the code reproduces AOD format from a specific record and contains a configuration file for running on an OpenData VM. The 'Use with' section indicates that the code should be used with a specific dataset: '/MinimumBias/Run2011A-v1/RAW'. The 'Characteristics' section shows that the dataset consists of 1 file with a total size of 14.6 kB. The 'System Details' section provides information on the CMS Open Data VM environment, software release (CMSSW\_5\_3\_32), and the VM image used for 2011 and 2012 CMS open data.

opendata  
CERN

Search

About

## Validation code for reprocessing AOD from 2011 MinimumBias RAW sample

Lassila-Perini, Kati

Cite as: Lassila-Perini, Kati; (2017). Validation code for reprocessing AOD from 2011 MinimumBias RAW sample. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.99j.BGV4

Software Validation CMS Accelerator CERN-LHC

### Description

This code for validation reproduces AOD format in [this record](#) from the [2011 MinimumBias RAW sample](#). It only contains the configuration file to be run on [2011 OpenData VM](#) using CMSSW\_5\_3\_32. No compilation is required. The configuration file has been slightly modified from the [original one](#) to take into account the OpenData computing environment (global tag, input file, commenting out unnecessary steps). Note that the code in this category is not meant to be a pedagogical example but is a validation tool.

### Use with

Use this with the following dataset:

[/MinimumBias/Run2011A-v1/RAW](#)

### Characteristics

Dataset: 1 files **14.6 kB** in total

### System Details

Use this code with the CMS Open Data VM environment for 2011 open data  
Software release: CMSSW\_5\_3\_32  
[CMS VM image, for 2011 and 2012 CMS open data](#)

<http://opendata.cern.ch/record/463>

# Example: CMS jet tuple production

opendata CERN Search About

## Example code for production of flat jet tuple using 2011 data

Zenaiev, Oleksandr; Haapalehto, Matias;

Cite as: Zenaiev, Oleksandr; Haapalehto, Matias; (2017). Example code for production of flat jet tuple using 2011 data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.J193.12X1

Software Analysis CMS Accelerator CERN/LHC

### Description

This code is a CMSSW module producing flat tuples from 2011 Jet data, set up at Research level, i.e. it requires university-student-level programming experience. Minimal acquaintance with Linux and the ROOT analysis package (<https://root.cern.ch/>) as well as a basic text editor is needed.

### Use with

Use this with 2011 Jets primary dataset and QCD MC dataset (see detailed instructions in the readme file and code).

[/jet/Run2011A-12Oct2013-v1/AOD](#)

[/QCD\\_Pt-80to120\\_Tune22\\_7TeV\\_pythia6/Summer11LegDR-PU\\_513\\_START53\\_LV6-v1/AODSIM](#)

### Notes

The content of the resulting root file is described in readme

### Characteristics

Dataset: 1 files 17.0 kB in total

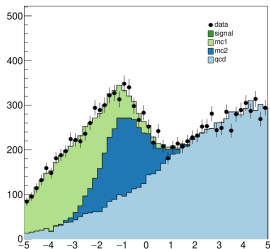
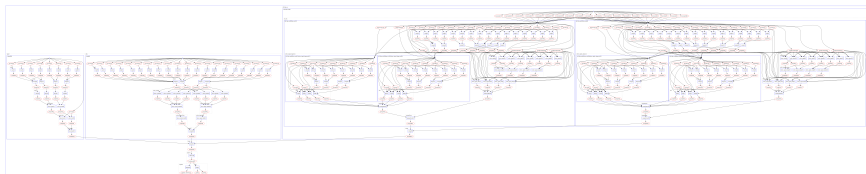
### System Details


Use this code with the CMS Open Data VM environment  
Software release: CMSSW\_5\_3\_32

<http://opendata.cern.ch/record/5104>



# Example: BSM searches

- full emulation of typical complex analysis workflows

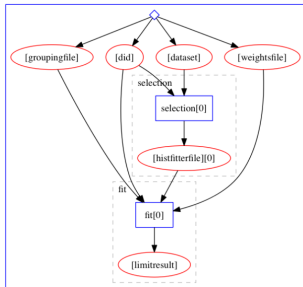


 <https://github.com/reanahub/reana-demo-bsm-search/>

# Yadage declarative workflows

- developed by  dianahep  recast
- several ATLAS internal use case examples
  - reinterpreting MonoH(bb) exotics searches
  - adding new systematics on displaced objects
  - ...

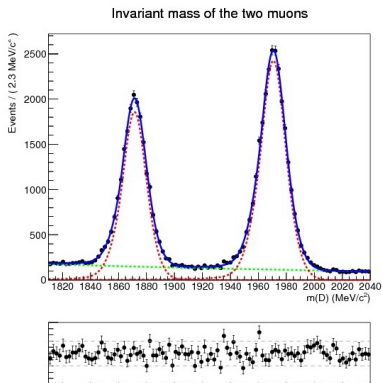
```
stages:  
- name: selection  
  dependencies: ['init']  
  scheduler:  
    scheduler_type: singlestep-stage  
    parameters:  
      dataset: {stages: init, output: dataset, unwrap: true}  
      submitdir: '{workdir}/submitdir'  
      outputprefix: '{workdir}/histfitter.root'  
      did: {stages: init, output: did, unwrap: true}  
      step: {$ref: 'selscript.yml#'}  
- name: fit  
  dependencies: ['selection']  
  scheduler:  
    scheduler_type: singlestep-stage  
    parameters:  
      bkgtree: 'root://eosuser.cern.ch///eos/project/r/recast/Bkg_2.4.15-2-0_merged.root'  
      datatree: 'root://eosuser.cern.ch///eos/project/r/recast/Data_2.4.15-2-0.root'  
      outputjson: '{workdir}/fitoutput.json'  
      selectionoutput: {stages: selection, output: histfitterfile, unwrap: true}  
      weightsfile: {stages: init, output: weightsfile, unwrap: true}  
      did: {stages: init, output: did, unwrap: true}  
      step: {$ref: 'fitscript.yml#'}
```




Lukas Heinrich <http://github.com/yadage>

# Example: LHCb rare charm decay

- search for  $D_{(s)}^+ \rightarrow \pi^+ \mu^+ \mu^-$  and  $D_{(s)}^+ \rightarrow \pi^- \mu^+ \mu^+$  decays  
Phys. Lett. B 724 (2013) 203-212




 <https://github.com/reanahub/reana-demo-lhcb-d2pimumu/>

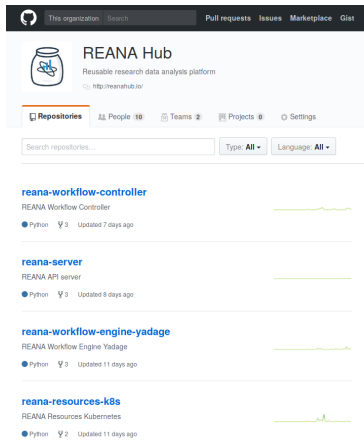
# Example: ALICE LEGO train

- ALICE LEGO train validation
- large images (~8 GB) including all dependencies (CVMFS)

```
$ root -b -q -x lego_train.C
...
I-AliPhysicsSelection::Initialize: Initializing for run 117222
I-AliPhysicsSelection::Initialize: Using Standard OADB
Physics selection cut settings:
  FMDLowCut          0.200000
  FMDHitCut          0.500000
...
==== Logical file name: /inputs/___alice__data__2011__LHC11h_2__000170387/AliESDs.root =====
...
=Analysis train= init time:          0.858033[sec]
                   I/O & data mng.: 16.6985 [sec]
                   task execution: 20.0355 [sec]
                   total time:      CPU=24.4 [sec] REAL=37.592[sec]
=Analysis train= Terminate time: 0.933296[sec]
Real time 0:00:40, CP time 27.390
===== lego_train.C finished with exit code: 0 =====
```

 <https://github.com/reanahub/reana-demo-alice-lego-train-validation/>

# REANA technology



REANA @ GitHub

## ■ micro-services



## ■ REST API



Bravado

## ■ services



RabbitMQ



## ■ deployments





# Developer perspective

## ■ set up REANA cluster

```
$ # install kubectl 1.9.1 and minikube 0.23.0
$ sudo dpkg -i kubectl*.deb minikube*.deb
$ minikube start --kubernetes-version="v1.6.4"
$ # install reana-cluster utility
$ mkvirtualenv reana-cluster
$ pip install reana-cluster
$ # deploy new cluster and check progress
$ reana-cluster init
$ reana-cluster status
$ # set environment variables for reana-client
$ eval $(reana-cluster env)
```

## ■ REANA client $\longleftrightarrow$ REANA cluster



### Researchers

Find out how you can use REANA to describe, run, preserve and reuse your analyses.

[User Guide](#)



### Administrators

Install and manage the REANA reusable analysis platform on your own compute cloud.

[Administrator Guide](#)



### Developers

Understand REANA source code, adapt it to your needs, contribute changes back.

[Developer Guide](#)

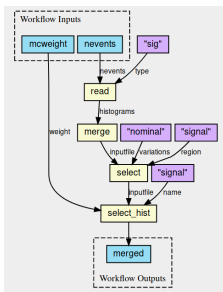
# Physics meets Life sciences

## ■ Common Workflow Language

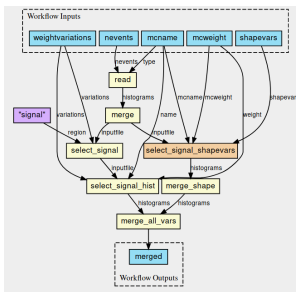
- originated in bioinformatics
- rich ecosystem: implementations, editors, tools



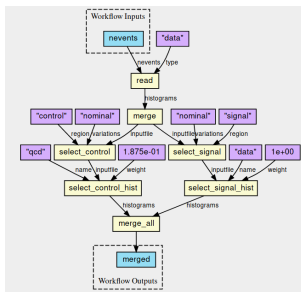
## ■ BSM search example using CWL workflows



sig

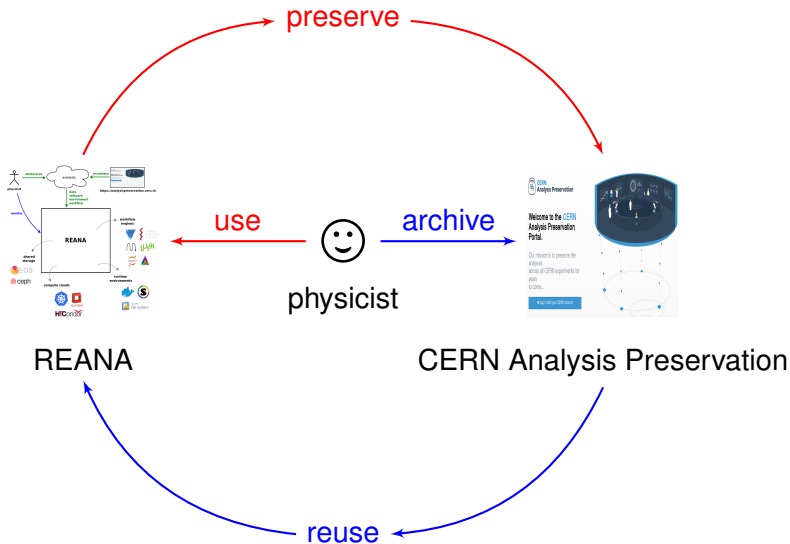


mc



data

# Reusability $\Rightarrow$ Preservation



# References

CERN  
Analysis Preservation

Welcome to the CERN  
Analysis Preservation  
Portal.

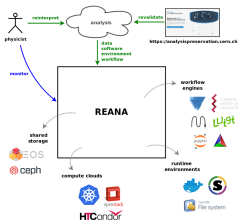
Our mission is to preserve the  
analyses across all CERN experiments for  
years to come.

More information on the website



## CERN Analysis Preservation

- <http://analysispreservation.cern.ch>
- <http://github.com/cernanalysispreservation>
- [analysis-preservation-support@cern.ch](mailto:analysis-preservation-support@cern.ch)



## REANA

- <http://www.reanahub.io>
- <http://github.com/reanahub>
- [reanahub](https://twitter.com/reanahub)
- [info@reanahub.io](mailto:info@reanahub.io)

**CERN IT** H. Hirvonsalo, D. Kousidis, D. Rodriguez, T. Šimko · **CERN SIS** S. Dallmeier-Tiessen, S. Feger, P. Fokianos, A. Lavasa, I. Tsanaktsidis, A. Trisovic, A. Trzcinska · **ALICE** D. Berzano, M. Gheata, C. Grigoras, Y. Schutz, M. Zimmermann · **ATLAS** J. Berlingen, K. Cranmer, L. Heinrich, L. Henkelmann, A. Sanchez Pineda, D. Rousseau, F. Socher · **CMS** A. Calderon, E. Carrera, A. Geiser, A. Huffman, C. Lange, K. Lassila-Perini, L. Lloret, T. McCauley, A. Rao, A. Rodriguez Marrero · **LHCb** S. Amerio, C. Burr, B. Couturier, S. Neubert, C. Parkes, A. Pearce, S. Roiser · **DASPOS** M. Hildreth · **DPHEP** J. Shiers