IAEA
International Atomic Energy Agency

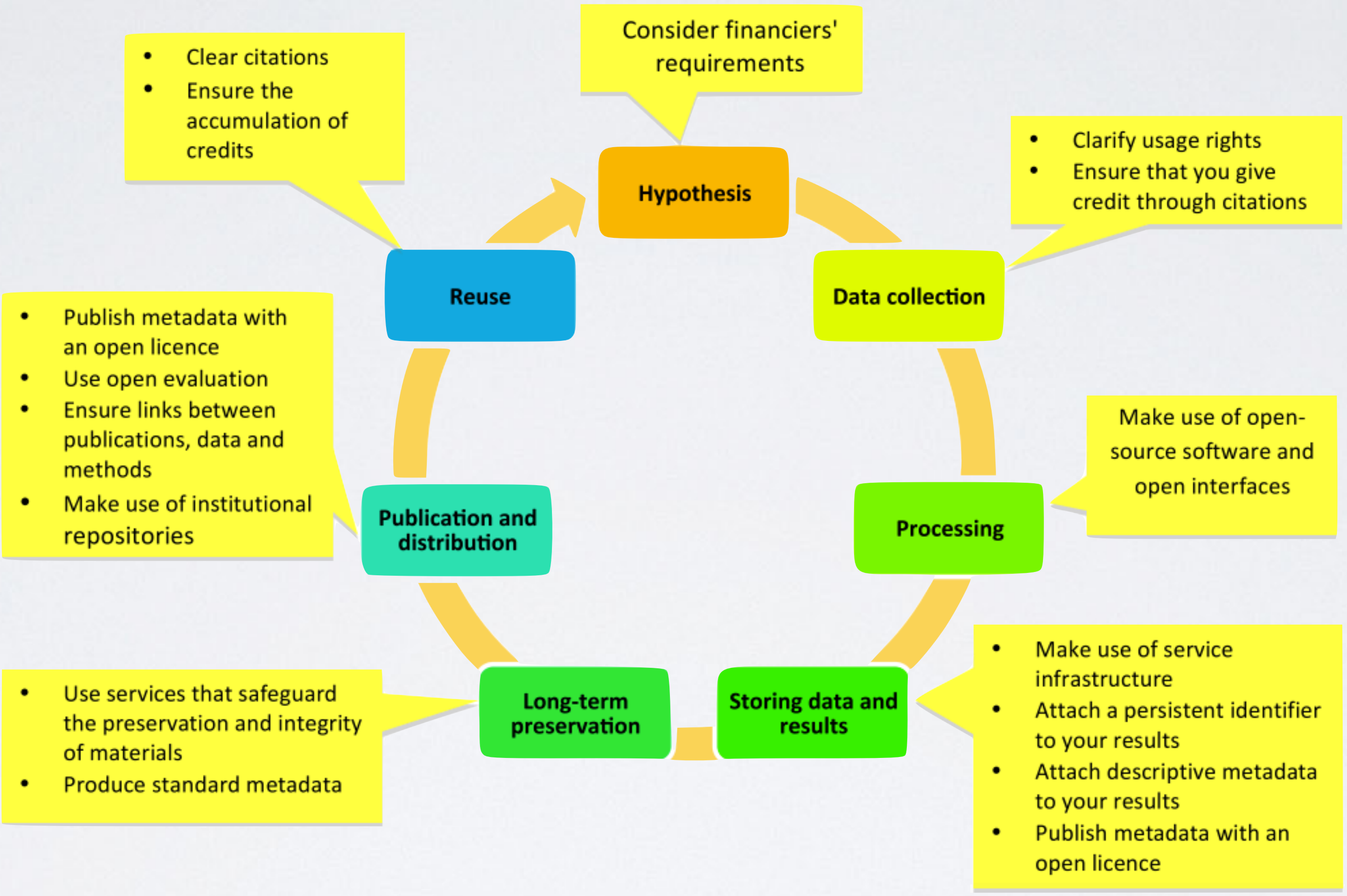
TOOLS & SERVICES FOR OPEN SCIENCE

Ludmila Marian
Scientific Data Manager, IAEA

WHAT IS NOT OPEN SCIENCE?

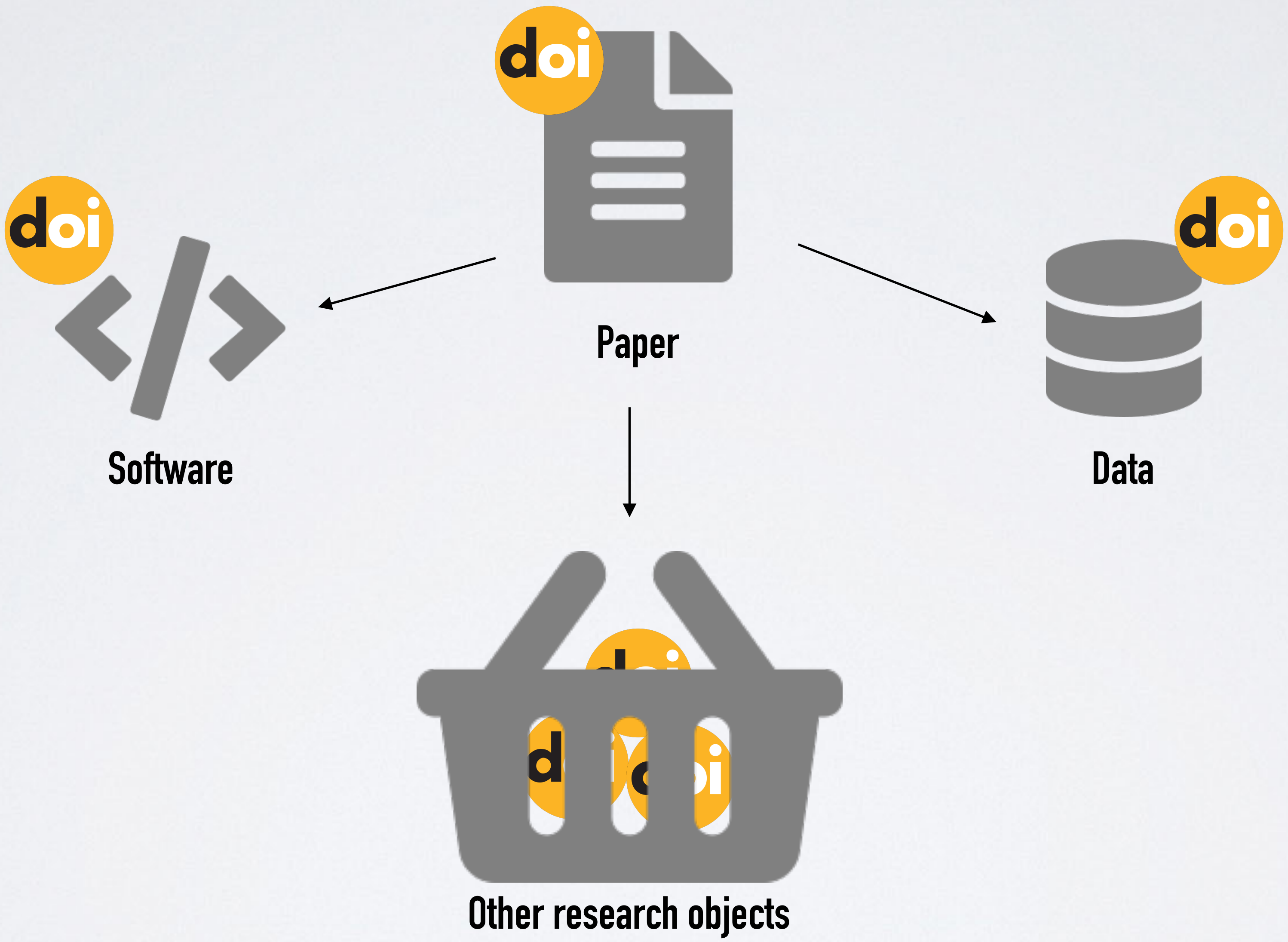
Published
Paper

OPEN THE RESEARCH PROCESS

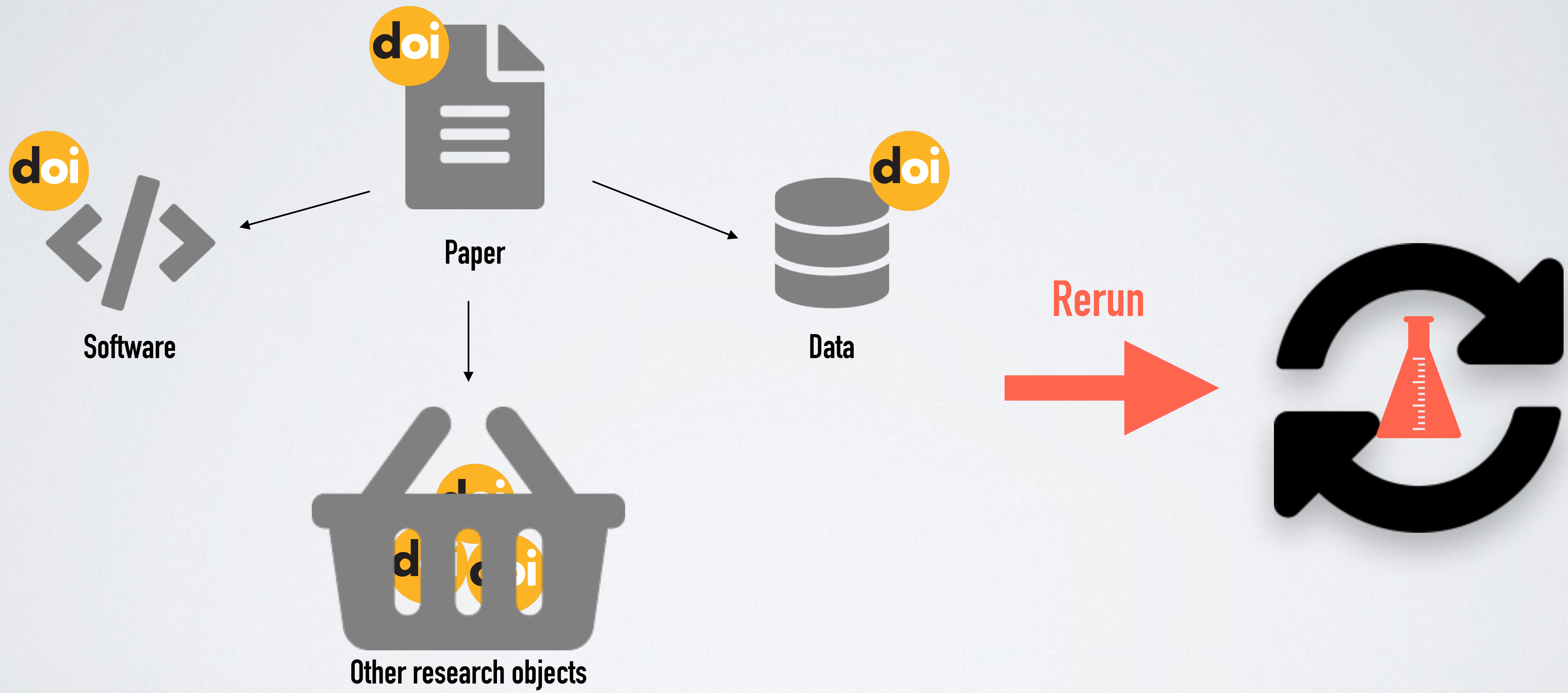


https://www.fosteropenscience.eu/content/what-open-science-introduction

OPEN RESEARCH



VALIDATE AND REUSE RESEARCH



WHERE TO STORE THESE OBJECTS?

WHERE TO STORE THESE OBJECTS?

In a Digital Repository!

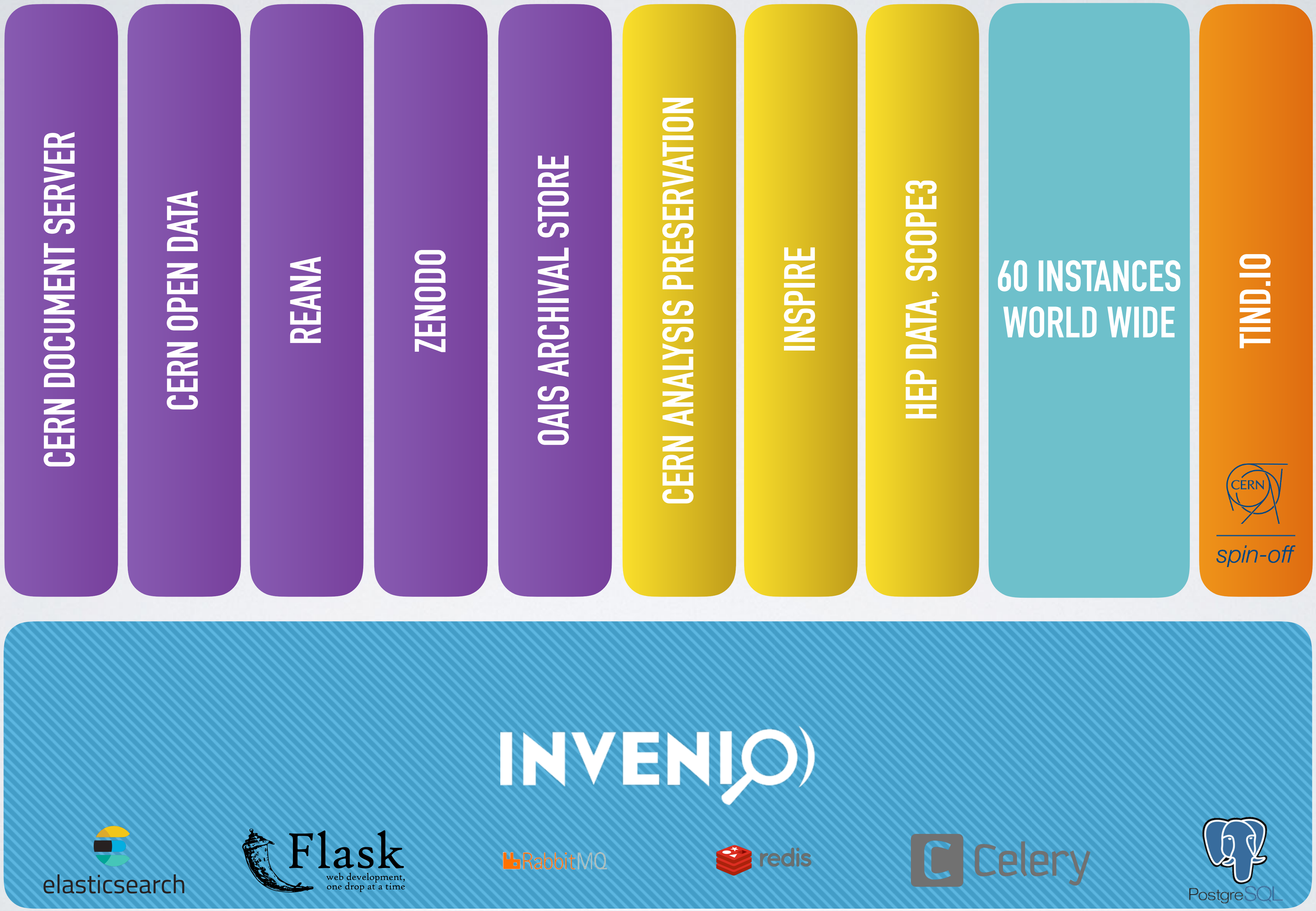
WHY A DIGITAL REPOSITORY?

PUBLISH OR PERISH?

20%

store data in a digital archive

OPEN SCIENCE @ CERN

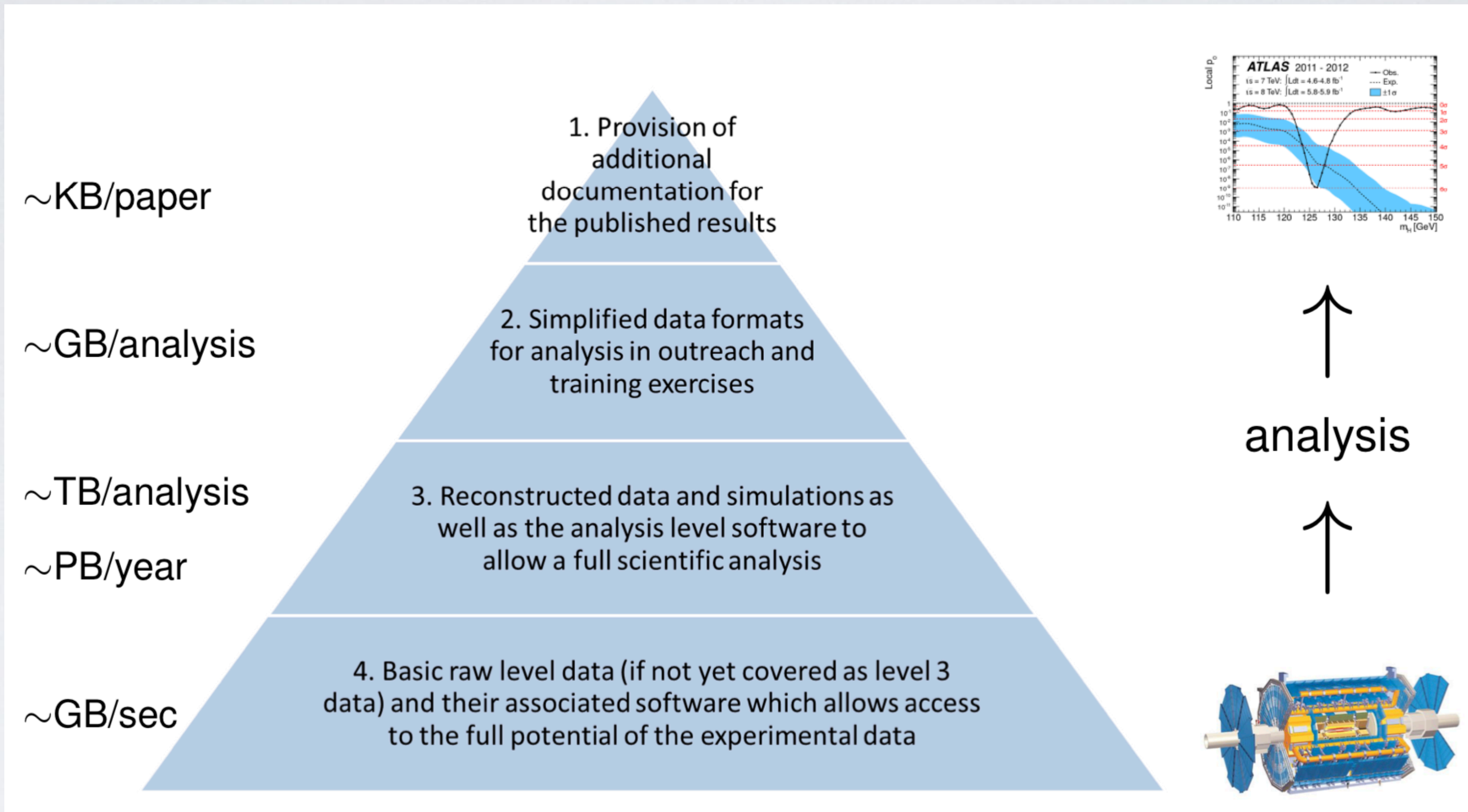


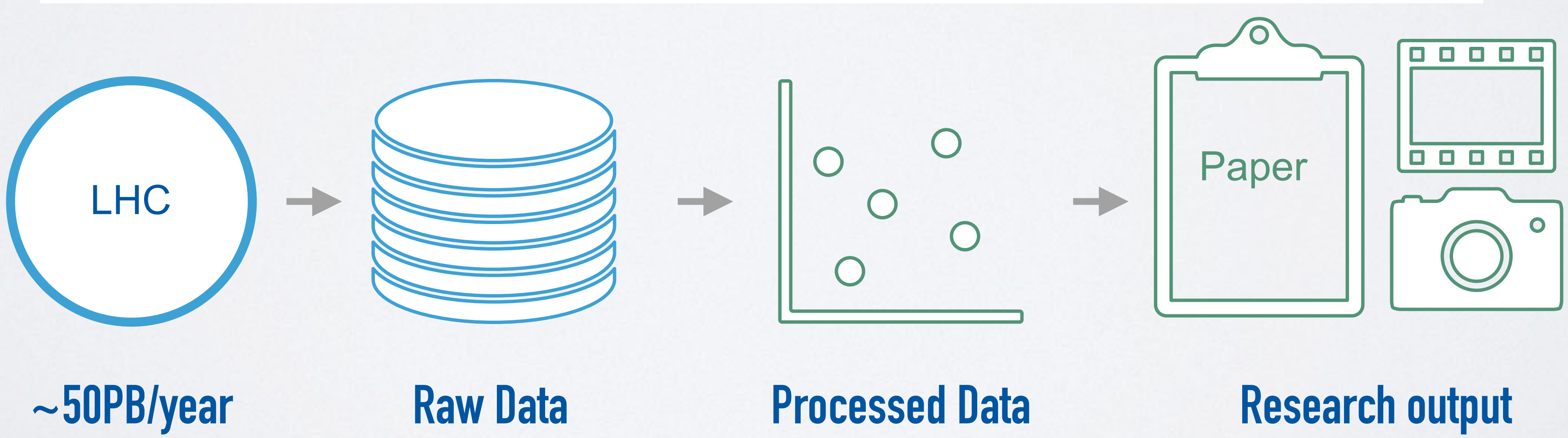
INVENIOSOFWARE.ORG

ABOUT INVENIO

- Born at **CERN**
- **Free Open** Source Software
- **Core** for any digital repository
- **Modern** & reliable technology
- **Flexible** and modular
- Handling **100M+** records
- Developed with **PB** in mind
- **Fast** uploads & search

DATA @ CERN


~KB/paper
~GB/analysis
~TB/analysis
~PB/year
~GB/sec
1. Provision of additional documentation for the published results
2. Simplified data formats for analysis in outreach and training exercises
3. Reconstructed data and simulations as well as the analysis level software to allow a full scientific analysis
4. Basic raw level data (if not yet covered as level 3 data) and their associated software which allows access to the full potential of the experimental data
ATLAS 2011 - 2012
analysis


LHC
Paper
~50PB/year
Raw Data
Processed Data
Research output

DATA @ CERN

Processed Data

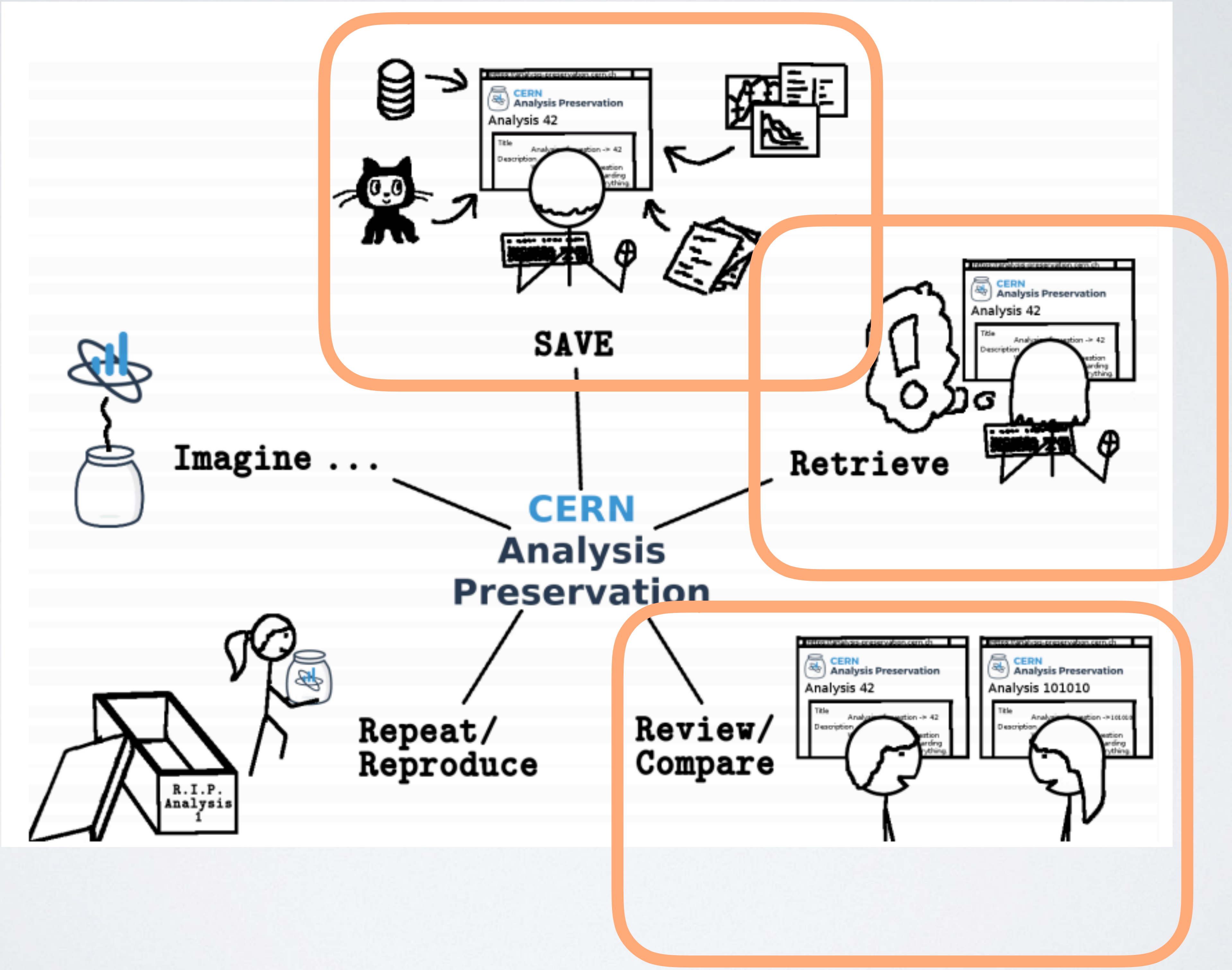
Research output

CERN ANALYSIS PRESERVATION	REANA	CERN OPEN DATA	CERN DOCUMENT SERVER
http://analysispreservation.cern.ch	https://reana.io	http://opendata.cern.ch	http://cds.cern.ch
Capture	**Reproduce**	**Publish**	**Publish and Archive**

Processed Data

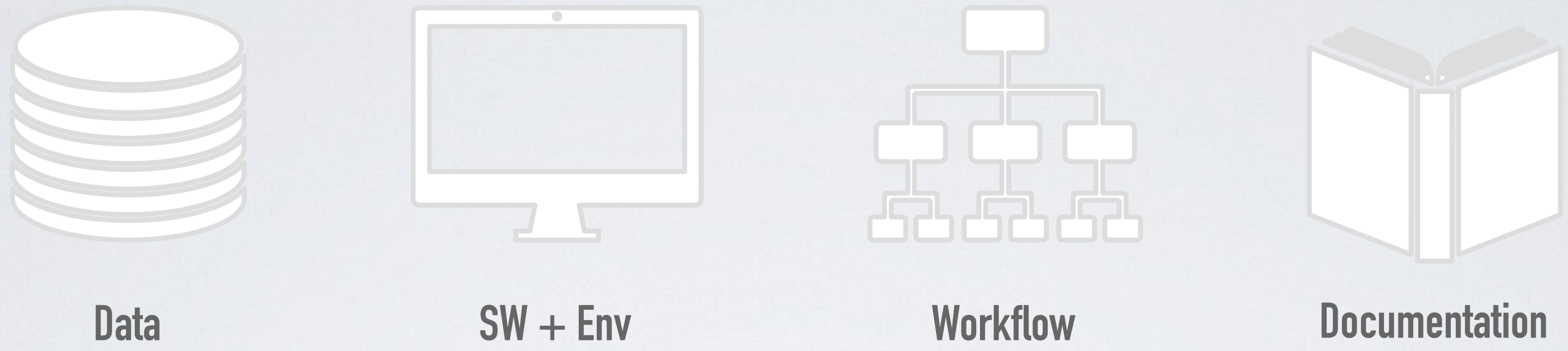
CERN ANALYSIS PRESERVATION	REANA	CERN OPEN DATA
http://analysispreservation.cern.ch	https://reana.io	http://opendata.cern.ch

CERN ANALYSIS PRESERVATION



CERN ANALYSIS PRESERVATION

Capturing all the elements needed to understand and rerun an analysis even several years later

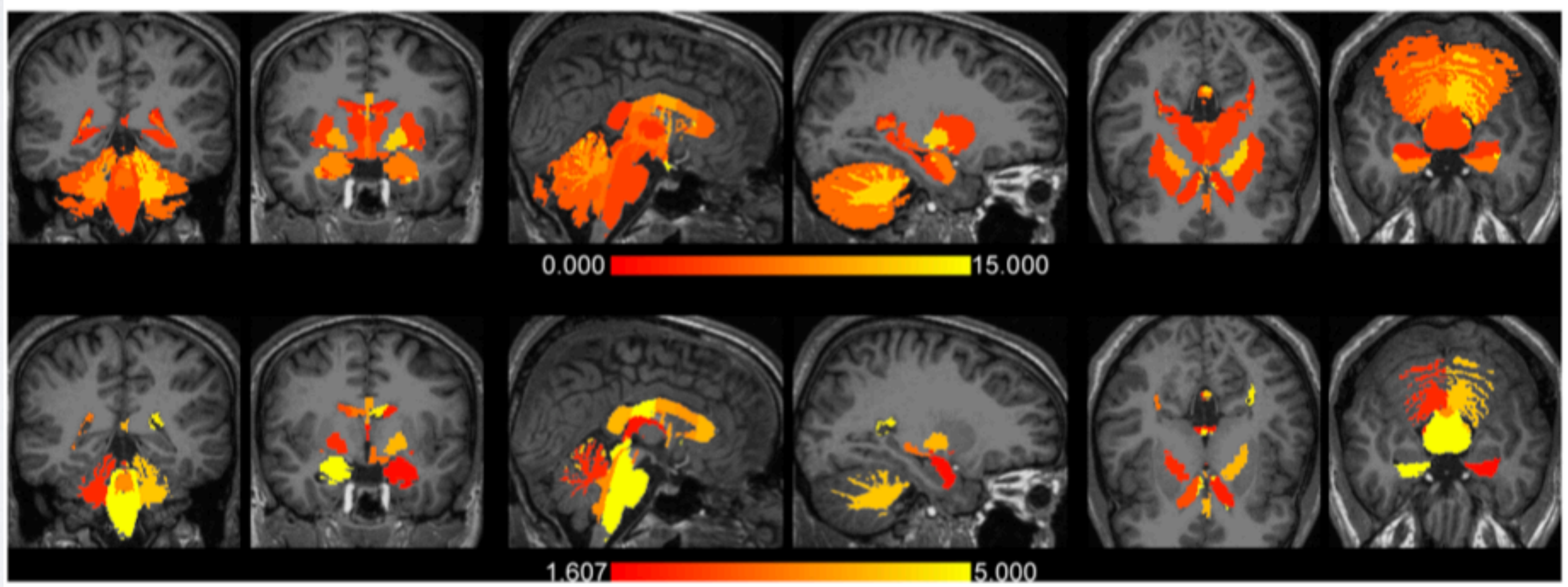


CODE + DATA IS NOT ENOUGH

The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements

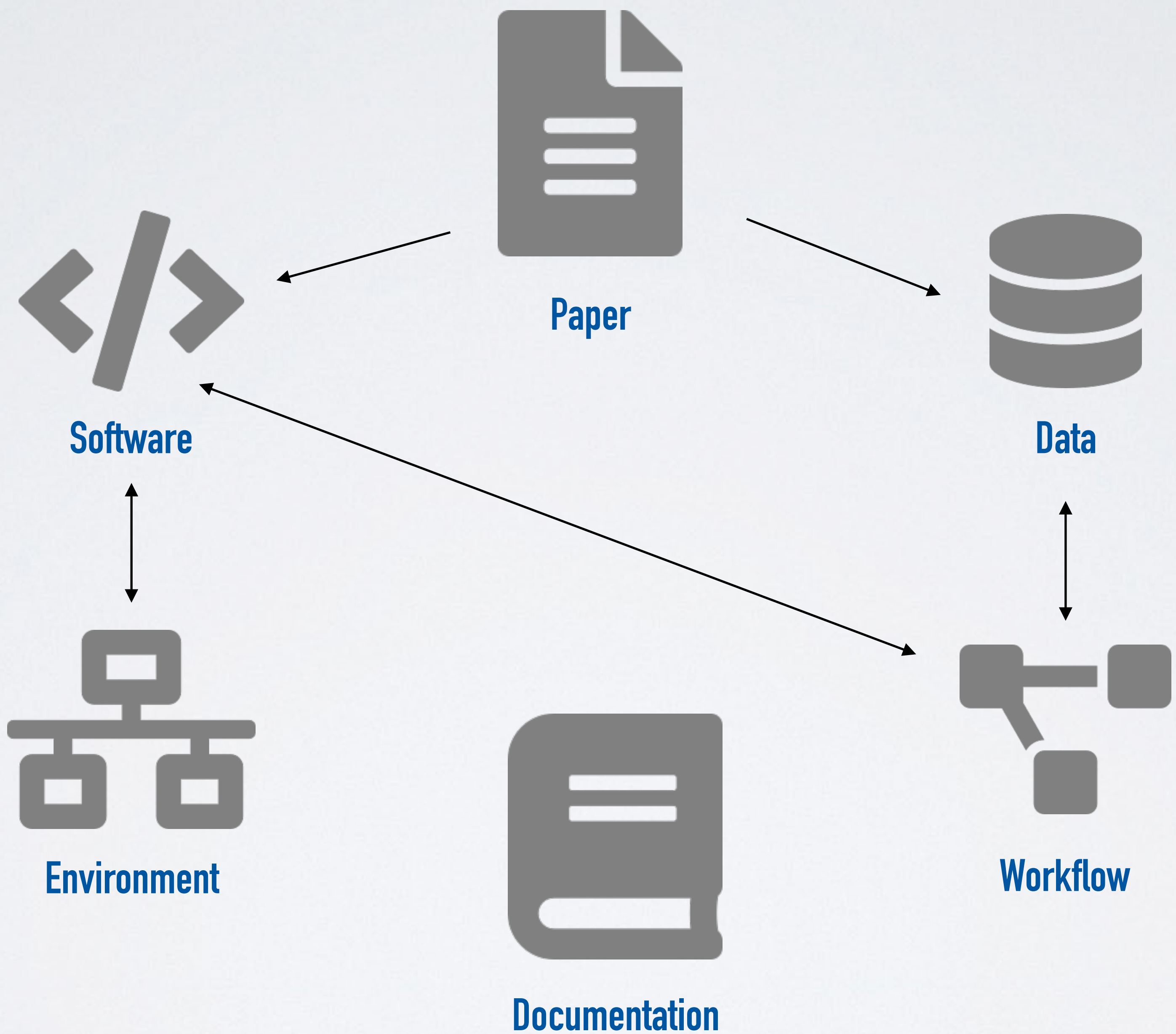
Ed H. B. M. Gronenschild, Petra Habets, Heidi I. L. Jacobs, Ron Mengelers, Nico Rozendaal, Jim van Os, Machteld Marcelis

Published: June 1, 2012 • DOI: 10.1371/journal.pone.0038234

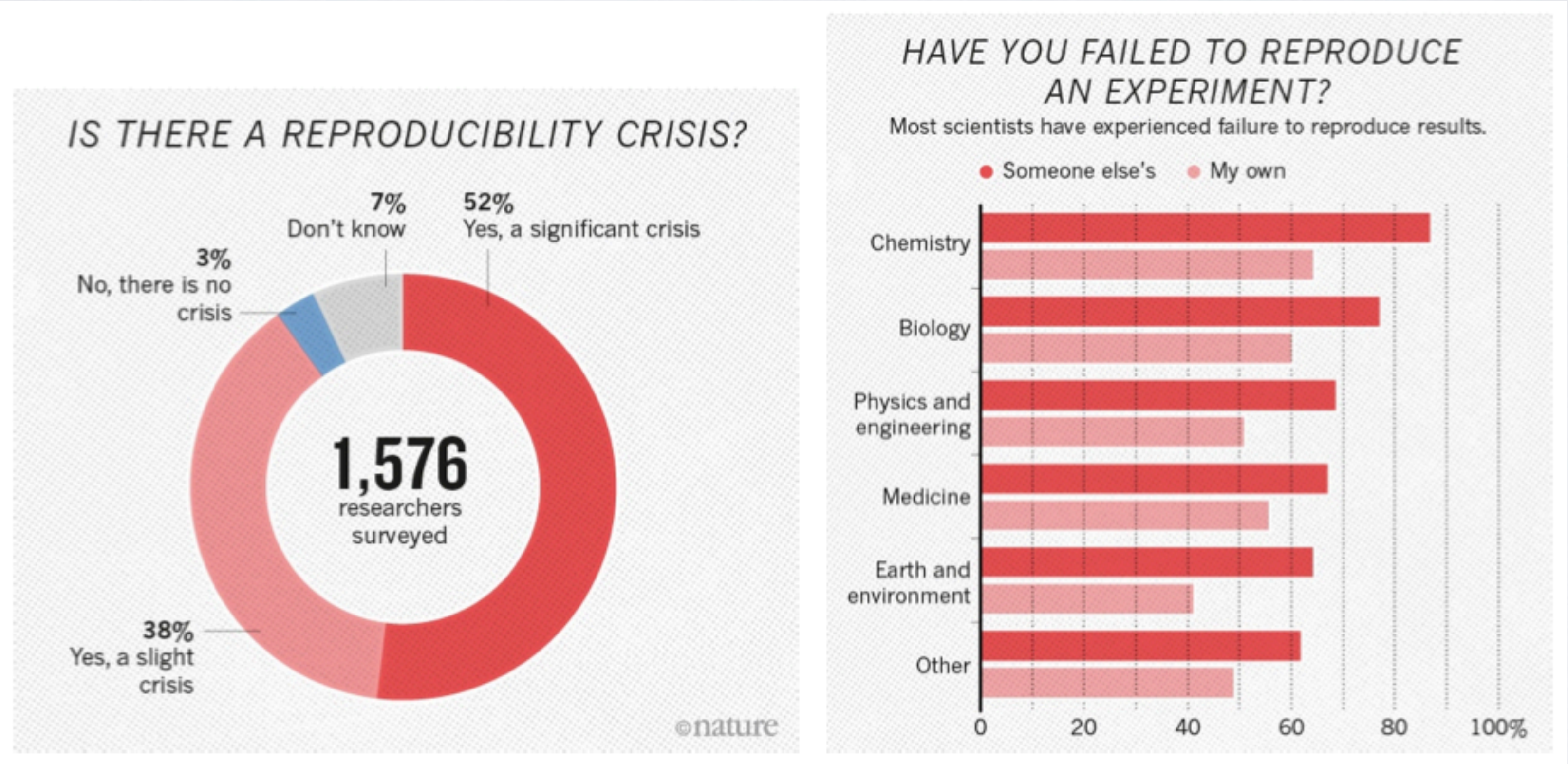


Software changes: 8.8±6.6% (volume) and 2.8±1.3% (thickness)
Operating system changes: factor two smaller

THE FULL RESEARCH MUST BE CAPTURED



REPRODUCIBILITY IS HARD



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

Half of researchers cannot reproduce their own results

Processed Data

CERN ANALYSIS PRESERVATION	REANA	CERN OPEN DATA
http://analysispreservation.cern.ch	https://reana.io	http://opendata.cern.ch

REPRODUCIBLE ANALYSIS

reana

Reproducible research data analysis platform

Flexible

Run many computational workflow engines.

COMMON WORKFLOW LANGUAGE

Scalable

Support for remote compute clouds.

kubernetes

Reusable

Containerise once, reuse elsewhere. Cloud-native.

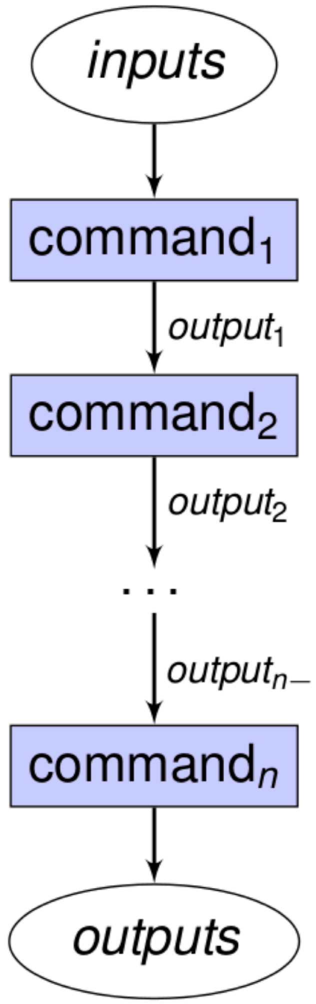
Free

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

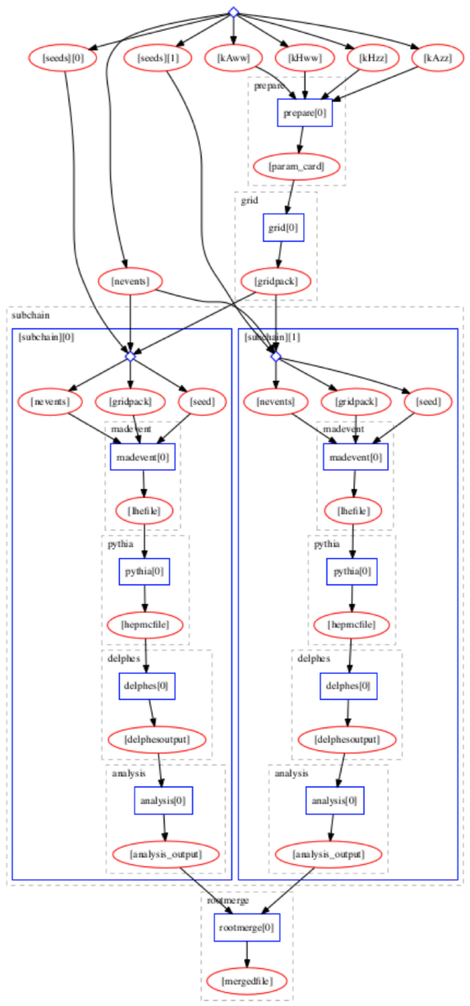
CERN

reana.io

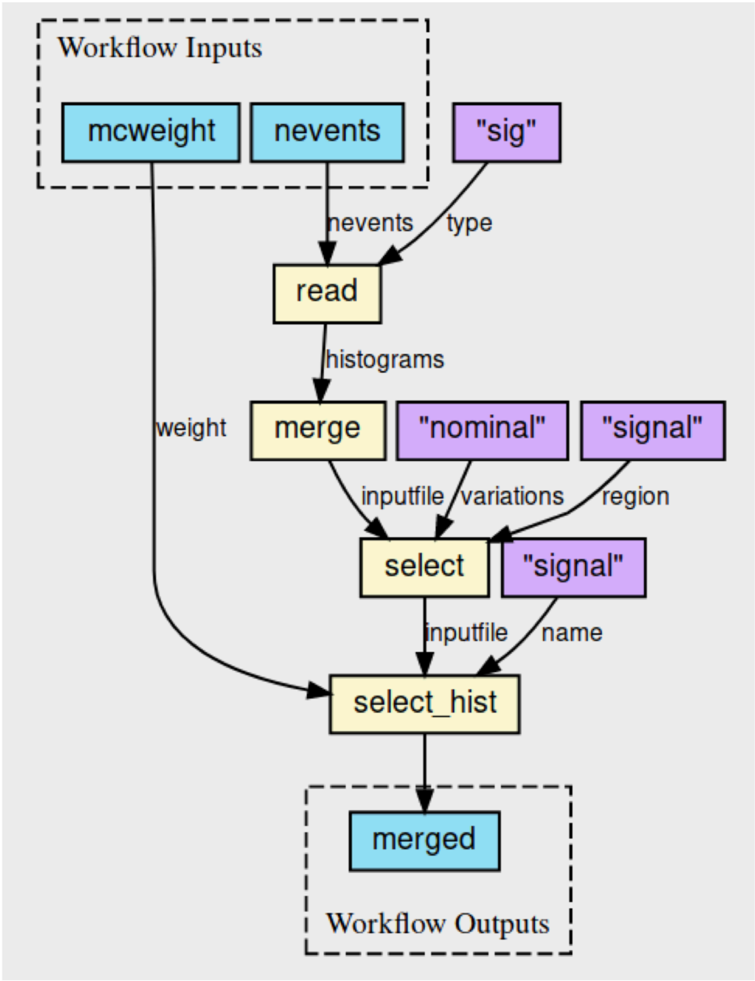
REPRODUCIBLE WORKFLOWS



Serial

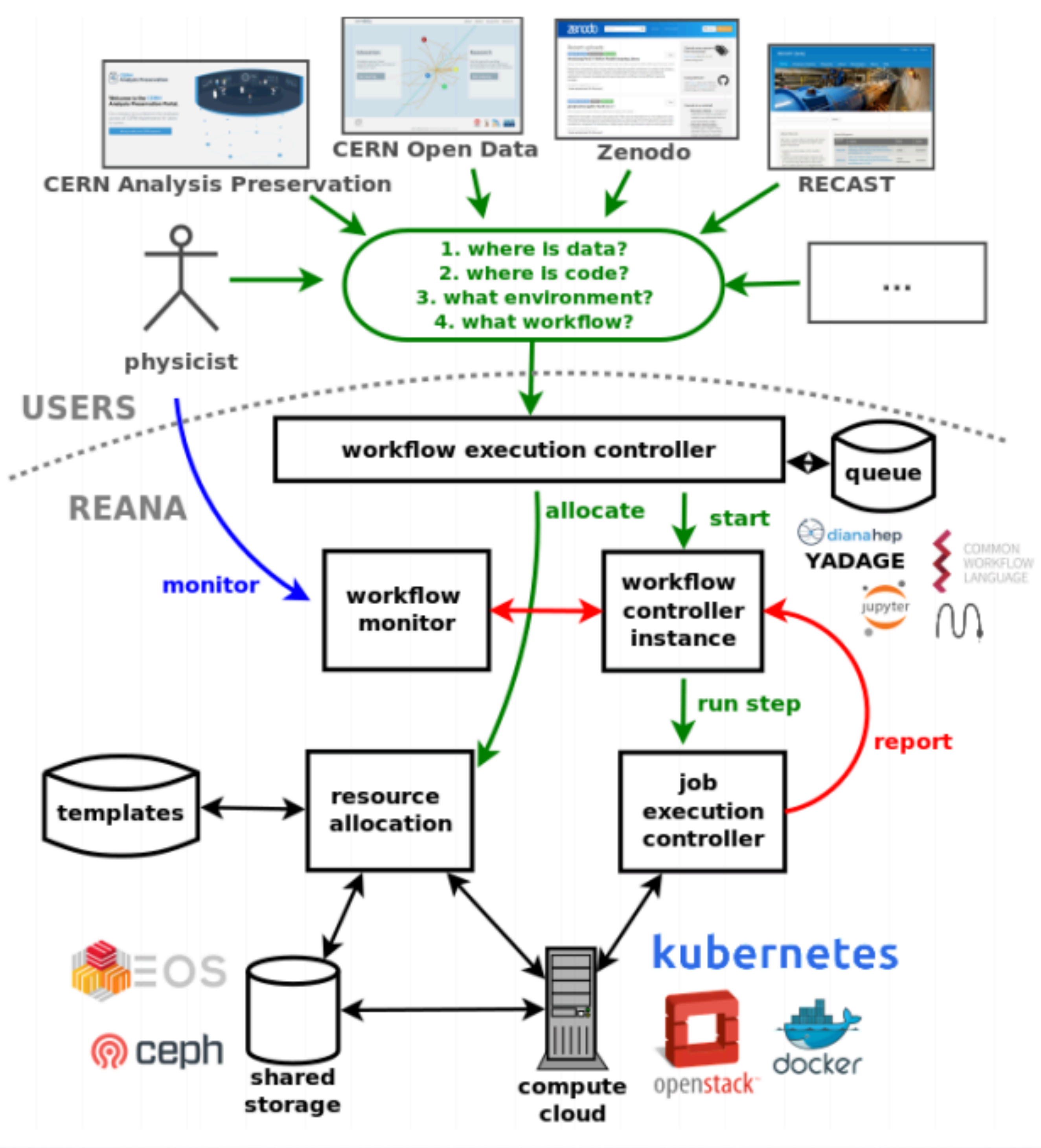


Yadage



CWL

REANA


CERN Analysis Preservation
CERN Open Data
Zenodo
RECAST
1. where is data?
2. where is code?
3. what environment?
4. what workflow?
...
physicist
USERS
REANA
workflow execution controller
queue
allocate
start
monitor
workflow monitor
workflow controller instance
YADAGE
run step
report
templates
resource allocation
job execution controller
kubernetes
EOS
ceph
shared storage
compute cloud
openstack
docker

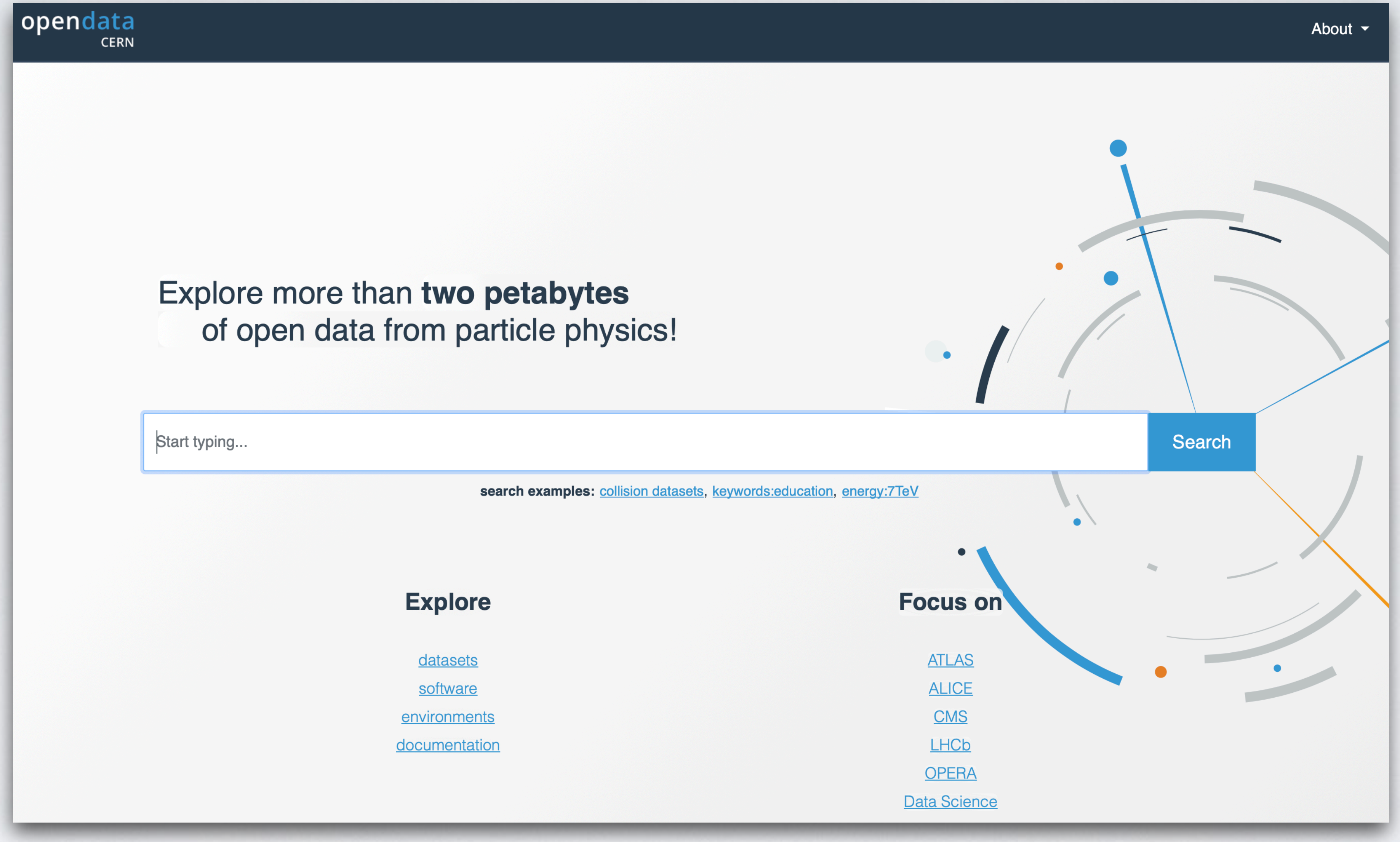
Processed Data

CERN ANALYSIS PRESERVATION	REANA	CERN OPEN DATA
http://analysispreservation.cern.ch	https://reana.io	http://opendata.cern.ch

CERN OPEN DATA PORTAL

2 PB data

Publicly-accessible site for curated releases of CERN data sets and software



CERN OPEN DATA PORTAL

opendata CERN | Search | About

Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD)

/Mu/Run2010B-Apr21ReReco-v1/AOD, CMS collaboration

Cite as: CMS collaboration (2014). Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.B8MR.C4A2

Dataset | Collision | CMS | Collision energy 7TeV | Accelerator CERN-LHC | Parent Dataset: /Mu/Run2010B-v1/RAW

Description

Mu primary dataset in AOD format from RunB of 2010

Notes

This dataset contains all runs from 2010 RunB. The list of validated runs, which must be applied to all analyses, can be found in

CMS list of validated runs Cert_136033-149442_7TeV_Apr21ReReco_Collisions10_JSON_v2.txt

Related Datasets

/Mu/Run2010B-v1/RAW

Characteristics

Dataset: **32376291** events **2979** files **3.2 TB** in total

System Details

Global tag: FT_R_42_V10A::All

Recommended release for analysis: CMSSW_4_2_1_patch1

How were these data selected?

There are four categories of triggers in the Mu dataset (with significant overlaps):

~70% inclusive single muon triggers with varying trigger pt threshold 3,5,7,9,11,13,15,17,19,21 GeV plus a few with loosened quality cuts.

~20% isolated single muon triggers with varying trigger pt threshold 9,11,13,15,17 GeV.

~10% inclusive dimuon triggers with varying trigger pt threshold 3,5 GeV plus one Z->mumu trigger with loosened quality cuts.

~20% combinations of muon triggers with various pt thresholds 3,5,7,8,9,11 GeV with some EM/e/gamma or hadronic/jet energy deposit with thresholds 6-100 GeV.

How were these data validated?

During data taking all the runs recorded by CMS are certified as good for physics analysis if all subdetectors, trigger, lumi and physics objects (tracking, electron, muon, photon, jet and MET) show the expected performance. Certification is based first on the offline shifters evaluation and later on the feedback provided by detector and Physics Object Group experts. Based on the above information, which is stored in a specific database called Run Registry, the Data Quality Monitoring group verifies the consistency of the certification and prepares a json file of certified runs to be used for physics analysis. For each reprocessing of the raw data, the above mentioned steps are repeated. For more information see:

CMS data quality monitoring: Systems and experiences

The CMS Data Quality Monitoring software experience and future improvements

The CMS data quality monitoring software: experience and future prospects

How can you use these data?

You can access these data through the CMS Virtual Machine. See the instructions for setting up the Virtual Machine and getting started in

How to install the CMS Virtual Machine

Getting started with CMS open data

Filename	Size	Download	EOS Link
CMS_Run2011A_BTag_AOD_12Oct2013-v1_00000_file_index.txt	122 Bytes		
CMS_Run2011A_BTag_AOD_12Oct2013-v1_20000_file_index.txt	59.5 kB		

FirstPrevious1NextLast

Datasets

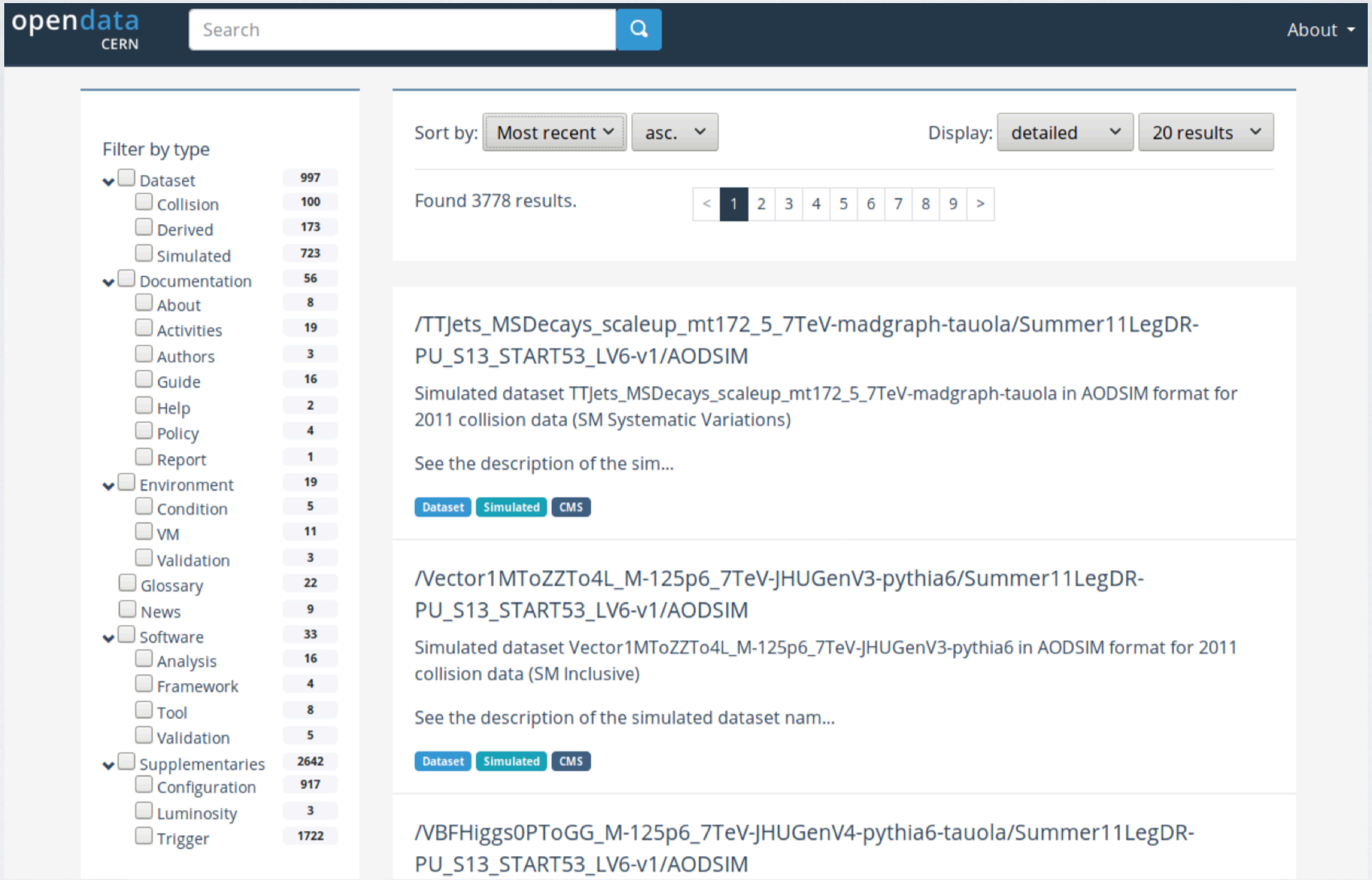
Filename	Size	Download	EOS Link
802CF580-BB46-E311-8D89-00261894388D.root	886.7 MB		
00376186-543E-E311-8D30-002618943857.root	3.9 GB		
0080432E-043E-E311-B4CB-00248C0BE01E.root	2.8 GB		
00867474-453E-E311-A450-003048FFD7C2.root	3.9 GB		
02012C2B-323E-E311-897E-003048FFD736.root	2.2 GB		
02116E88-003E-E311-A1A9-0025905964BA.root	4.2 GB		
0216066B-3A3E-E311-ABD0-003048FFD732.root	3.9 GB		
02477509-3D3E-E311-A230-00261894389A.root	3.9 GB		
02581093-3E3E-E311-8235-00248C55CC3C.root	3.9 GB		
0297C037-2D3E-E311-83A2-00259059649C.root	4.2 GB		

FirstPrevious12345NextLast

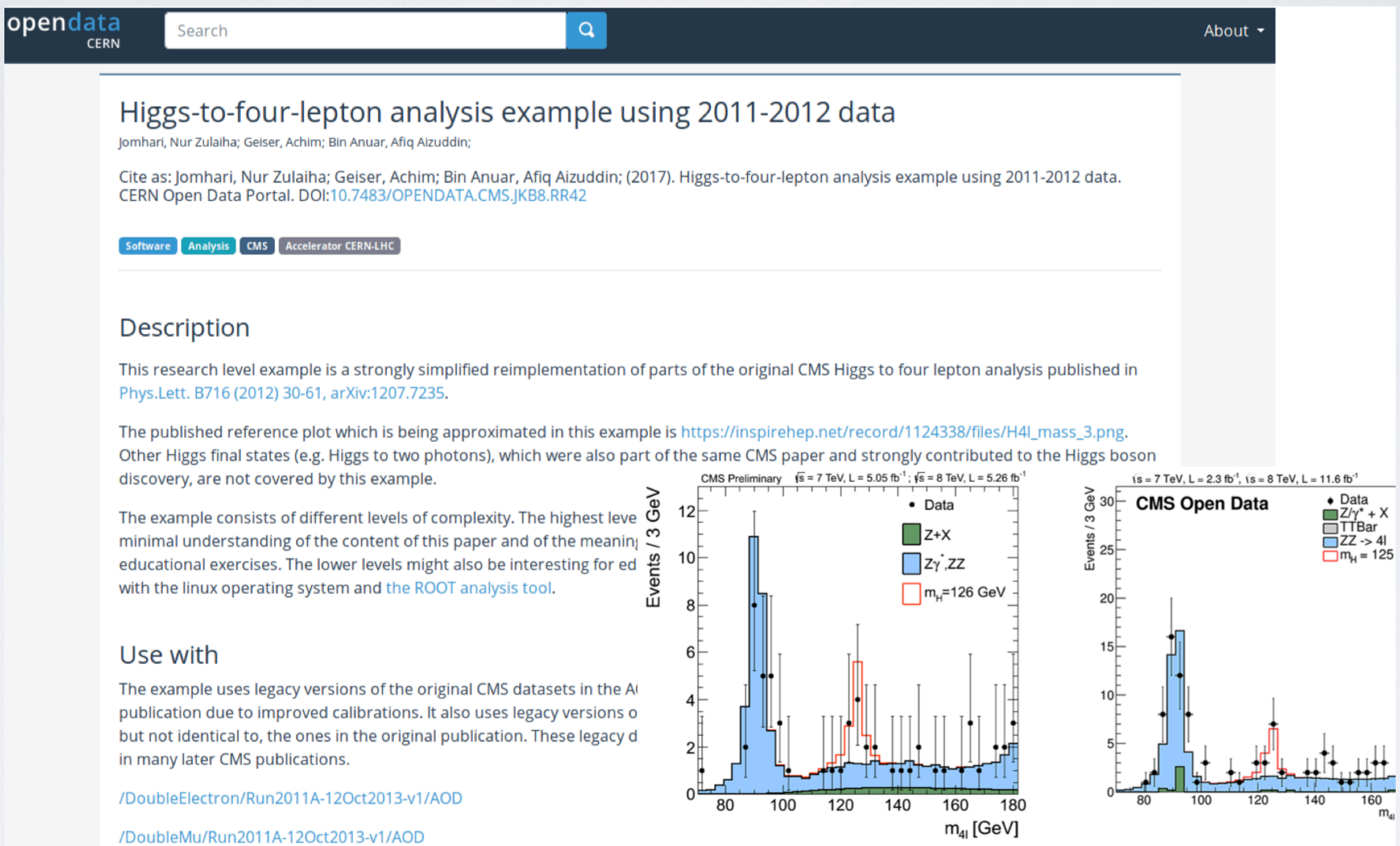
Disclaimer

The open data are released under the Creative Commons CC0 waiver. Neither CMS nor CERN endorse any works, scientific or otherwise, produced using these data. All releases will have a unique DOI that you are requested to cite in any applications or publications.

CERN OPEN DATA PORTAL

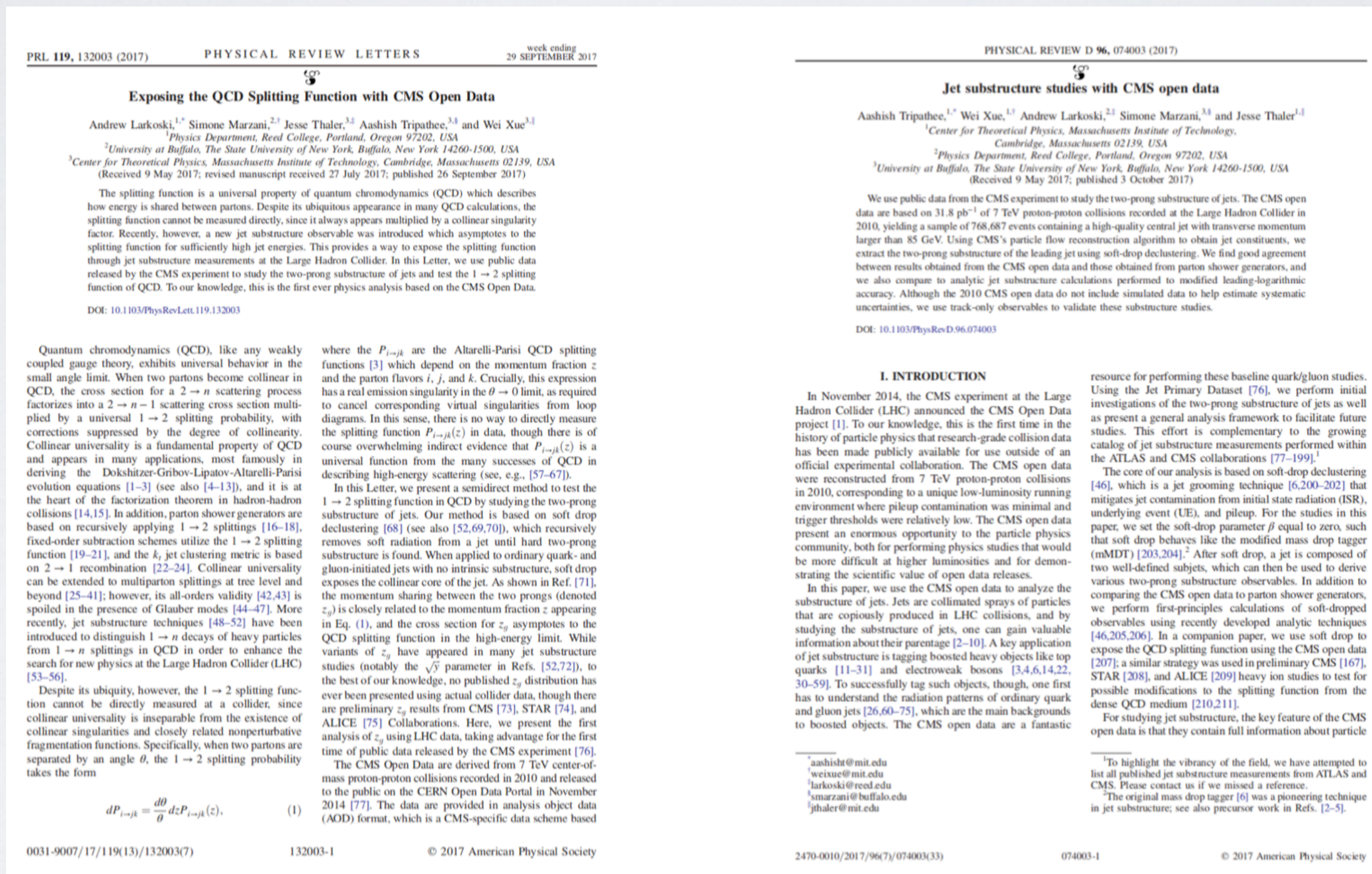


CERN OPEN DATA PORTAL

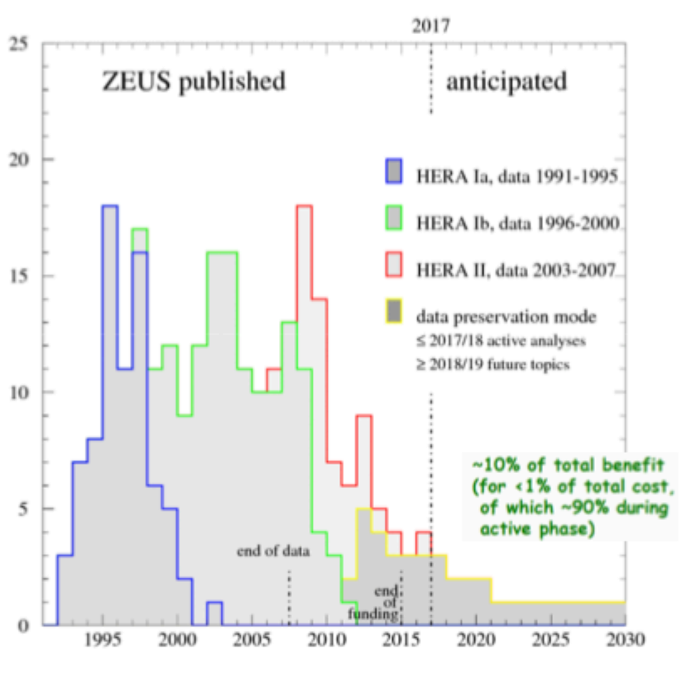


RESEARCH USE CASES

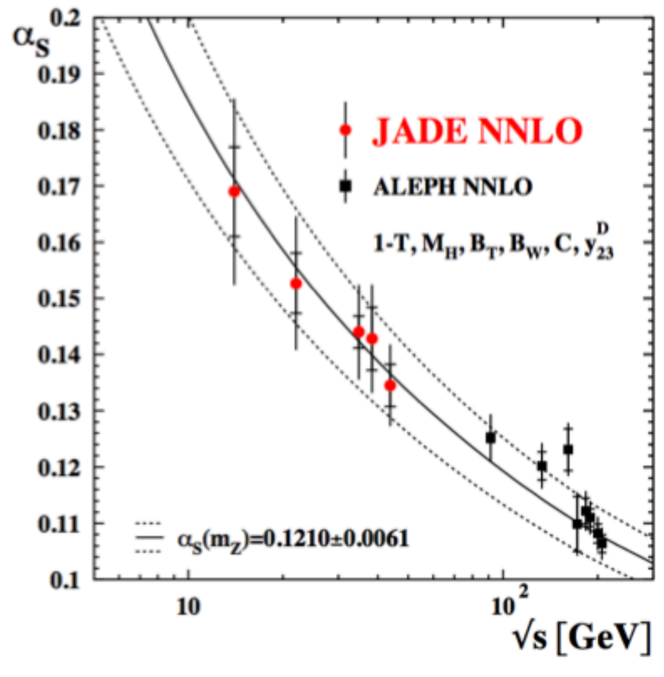
Independent analyses by theorists (Jesse Thaler et al, MIT)



LONG TERM VALUE OF DATA



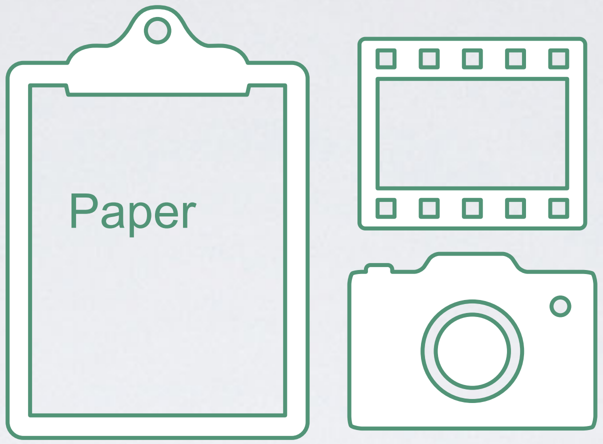
Achim Geiser https://indico.cern.ch/e/588219



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

Collaborations publish papers even ~15 years after data taking ends

JADE data (1979-1986) still unique even ~35 years later

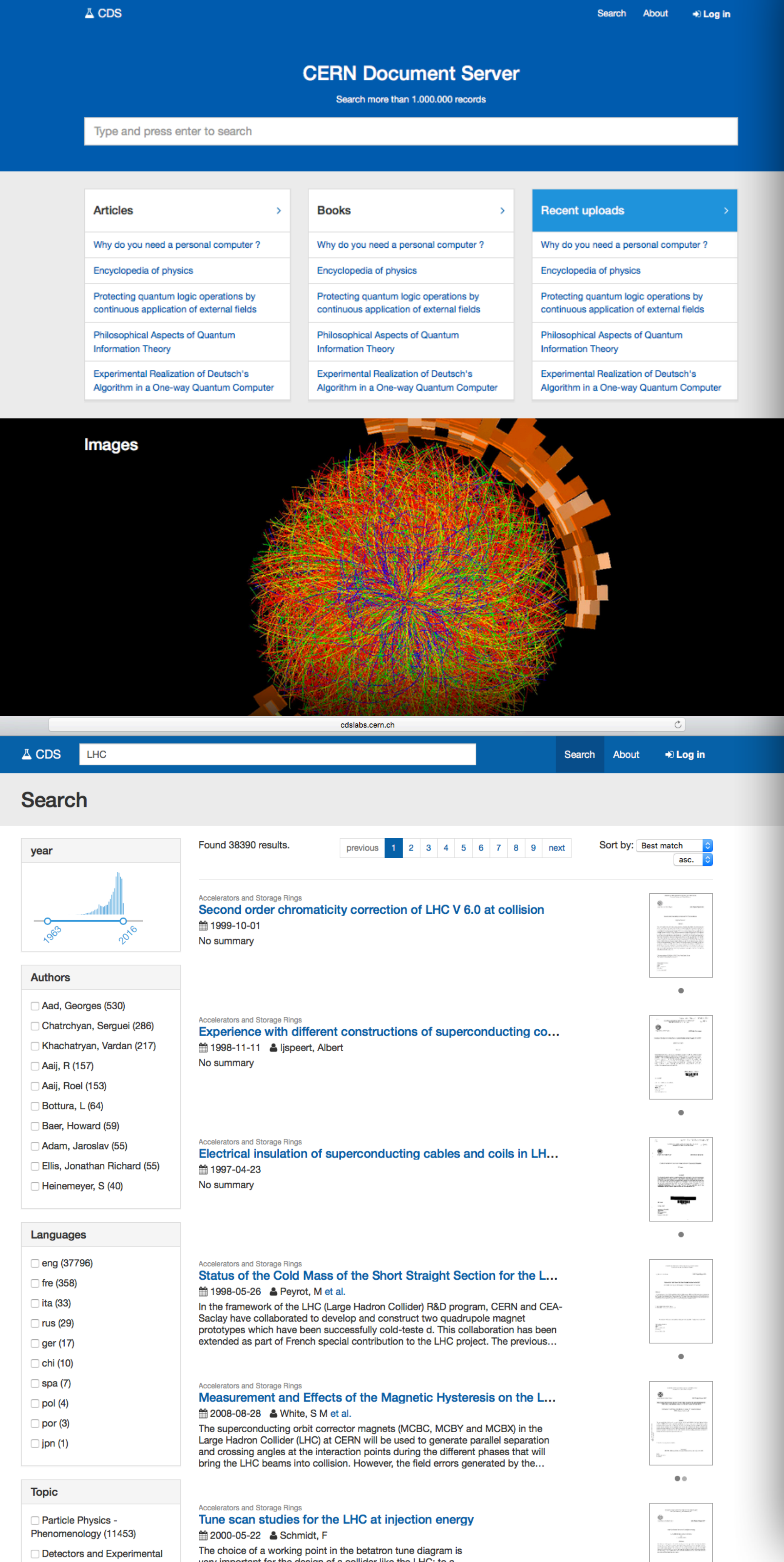
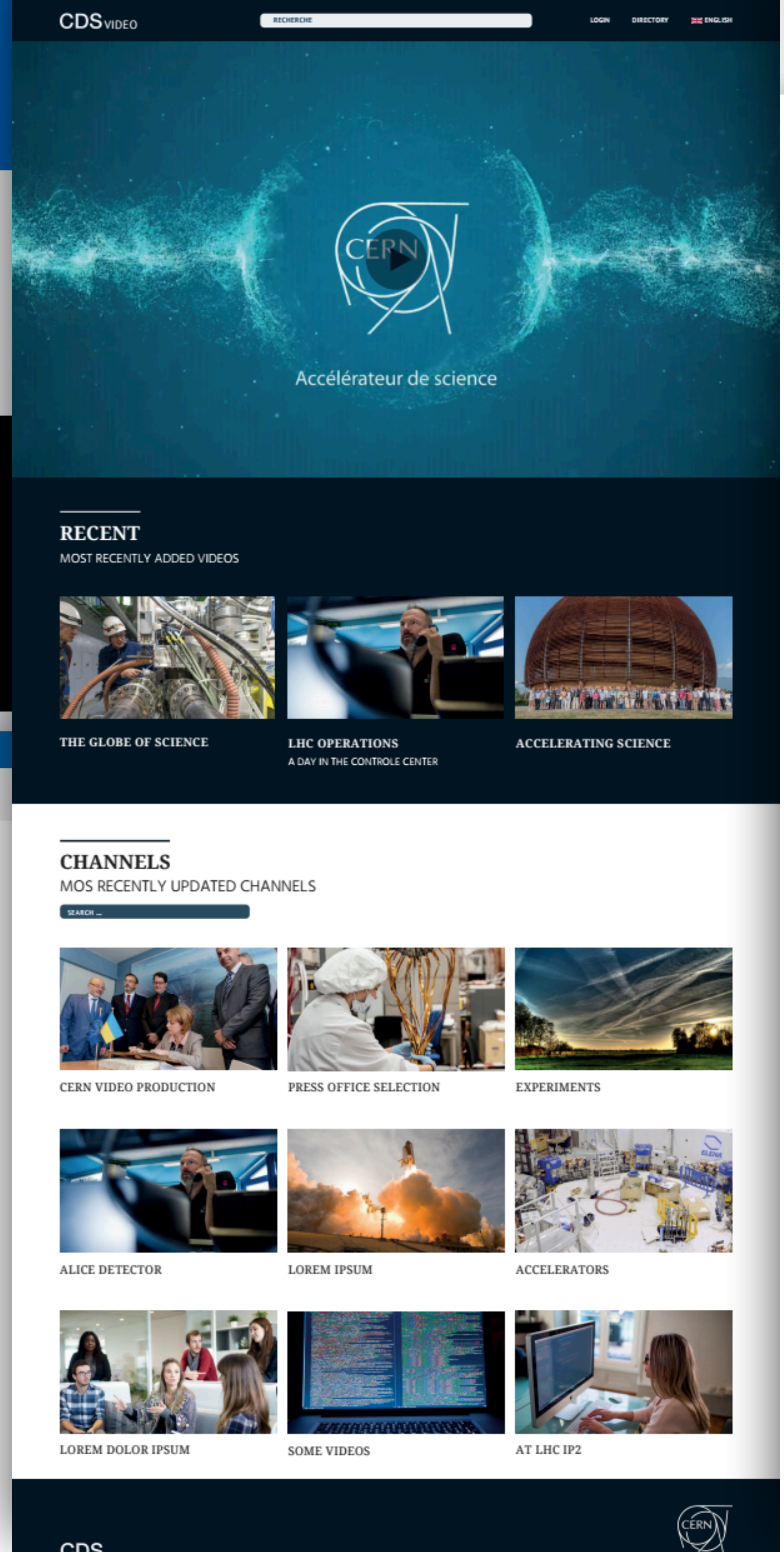
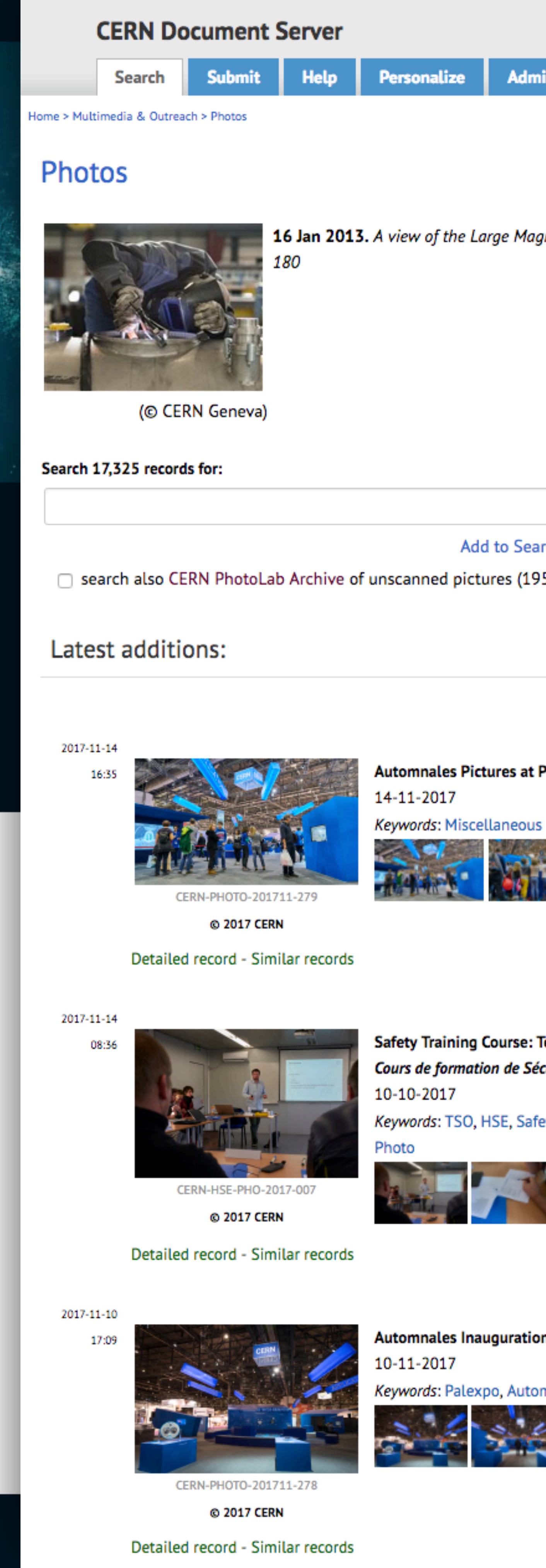


Research output

CERN DOCUMENT SERVER

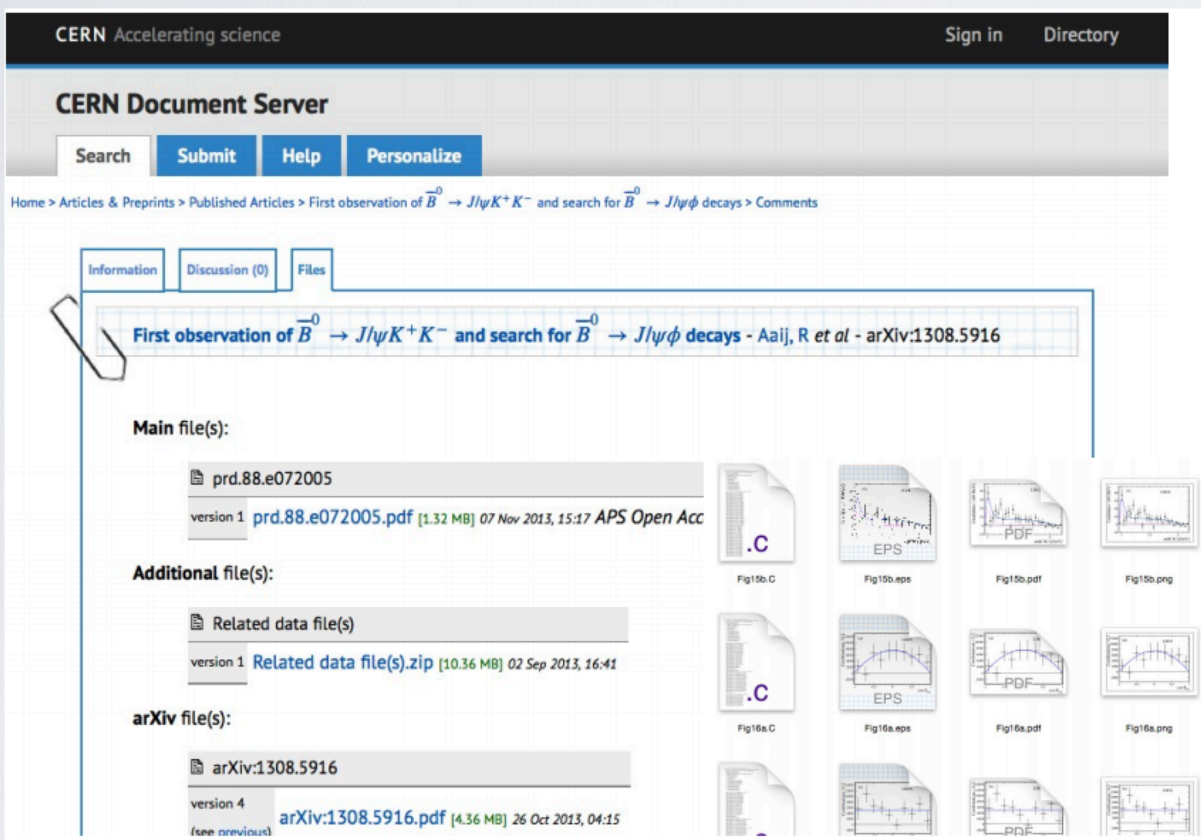
http://cds.cern.ch

Publish and Archive


CDS
Search
About
Log in
CERN Document Server
Search more than 1.000.000 records
Type and press enter to search
Articles
Books
Recent uploads
Why do you need a personal computer ?
Encyclopedia of physics
Protecting quantum logic operations by continuous application of external fields
Philosophical Aspects of Quantum Information Theory
Experimental Realization of Deutsch's Algorithm in a One-way Quantum Computer
Images
cdslabs.cern.ch
LHC
Search
Found 38390 results.
previous 1 2 3 4 5 6 7 8 9 next
Sort by: Best match
asc.
year
1963
2016
Authors
Aad, Georges (530)
Chatrchyan, Serguei (286)
Khachatryan, Vardan (217)
Aaij, R (157)
Aaij, Roel (153)
Bottura, L (64)
Baer, Howard (59)
Adam, Jaroslav (55)
Ellis, Jonathan Richard (55)
Heinemeyer, S (40)
Languages
eng (37796)
fre (358)
ita (33)
rus (29)
ger (17)
chi (10)
spa (7)
pol (4)
por (3)
jpn (1)
Topic
Particle Physics - Phenomenology (11453)
Detectors and Experimental Techniques (11156)
Accelerators and Storage Rings
Second order chromaticity correction of LHC V 6.0 at collision
1999-10-01
No summary
Experience with different constructions of superconducting co...
1998-11-11 Ijspeert, Albert
Electrical insulation of superconducting cables and coils in LH...
1997-04-23
Status of the Cold Mass of the Short Straight Section for the L...
1998-05-26 Peyrot, M et al.
In the framework of the LHC (Large Hadron Collider) R&D program, CERN and CEA-Saclay have collaborated to develop and construct two quadrupole magnet prototypes which have been successfully cold-teste d. This collaboration has been extended as part of French special contribution to the LHC project. The previous...
Measurement and Effects of the Magnetic Hysteresis on the L...
2008-08-28 White, S M et al.
The superconducting orbit corrector magnets (MCBC, MCBY and MCBX) in the Large Hadron Collider (LHC) at CERN will be used to generate parallel separation and crossing angles at the interaction points during the different phases that will bring the LHC beams into collision. However, the field errors generated by the...
Tune scan studies for the LHC at injection energy
2000-05-22 Schmidt, F
The choice of a working point in the betatron tune diagram is very important for the design of a collider like the LHC: to a great extent the performance of the collider depends on the...

CDS VIDEO
RECHERCHE
LOGIN
DIRECTORY
ENGLISH
CERN
Accélérateur de science
RECENT
MOST RECENTLY ADDED VIDEOS
THE GLOBE OF SCIENCE
LHC OPERATIONS
A DAY IN THE CONTROLE CENTER
ACCELERATING SCIENCE
CHANNELS
MOS RECENTLY UPDATED CHANNELS
CERN VIDEO PRODUCTION
PRESS OFFICE SELECTION
EXPERIMENTS
ALICE DETECTOR
LOREM IPSUM
ACCELERATORS
LOREM DOLOR IPSUM
SOME VIDEOS
AT LHC IP2
CDS

CERN Document Server
Search
Submit
Help
Personalize
Home > Multimedia & Outreach > Photos
Photos
16 Jan 2013. A view of the Large Magnet
180
(© CERN Geneva)
Search 17,325 records for:
Add to Search
search also CERN PhotoLab Archive of unscanned pictures (1952-
Latest additions:
2017-11-14
16:35
Automnales Pictures at Pale
14-11-2017
Keywords: Miscellaneous
CERN-PHOTO-201711-279
© 2017 CERN
Detailed record - Similar records
2017-11-14
08:36
Safety Training Course: Terr
Cours de formation de Sécuri
10-10-2017
Keywords: TSO, HSE, Safety
Photo
CERN-HSE-PHO-2017-007
© 2017 CERN
Detailed record - Similar records
2017-11-10
17:09
Automnales Inauguration at
10-11-2017
Keywords: Palexpo, Automna
CERN-PHOTO-201711-278
© 2017 CERN
Detailed record - Similar records

CERN DOCUMENT SERVER

ALSO USED FOR SMALL DATA

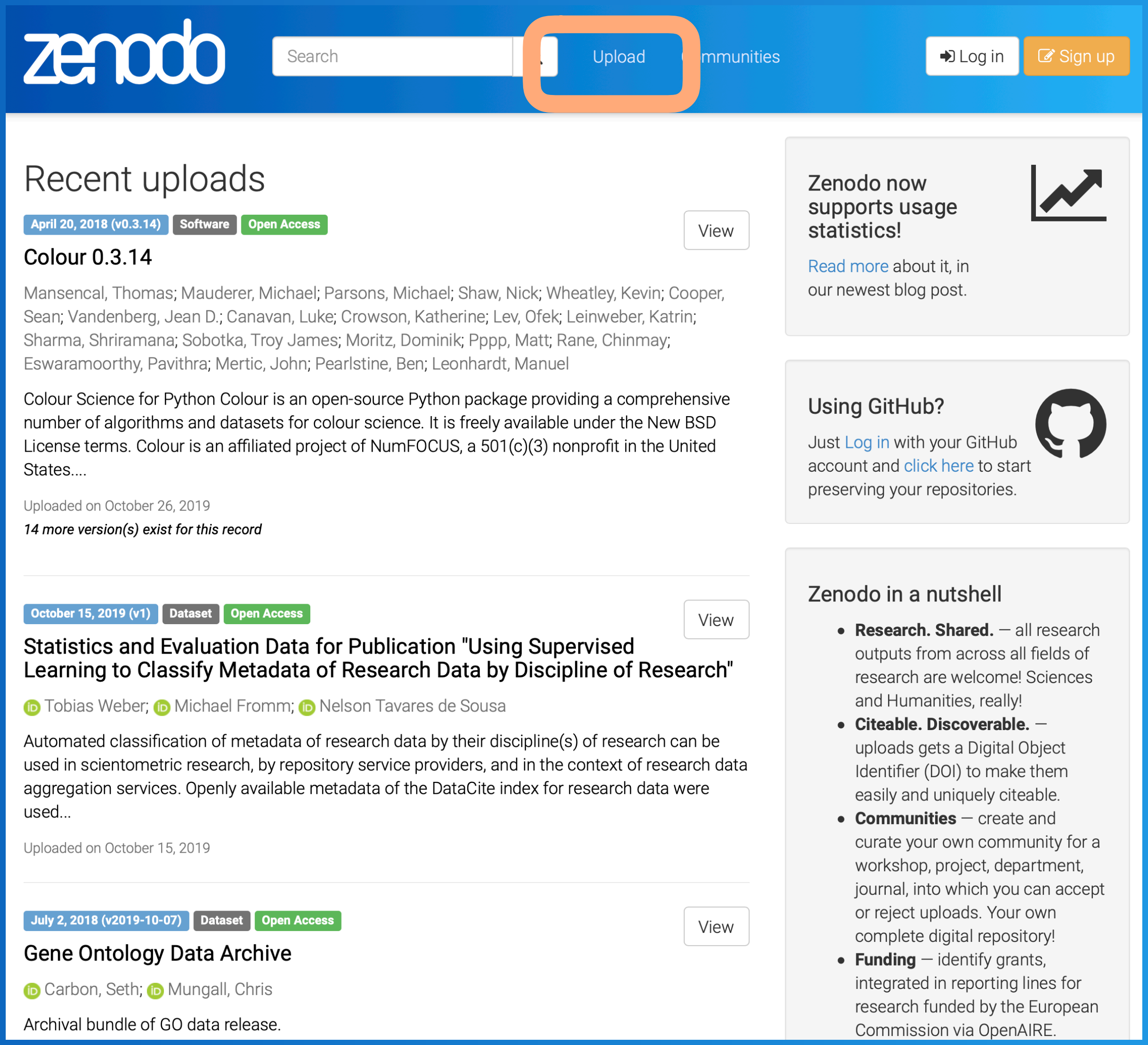


- 2.3 Million records
- 6.5 Million files
- 140 TB of data
- 5'000 unique visits / day

zenodo

Zenodo

http://zenodo.org



CERN Data Center

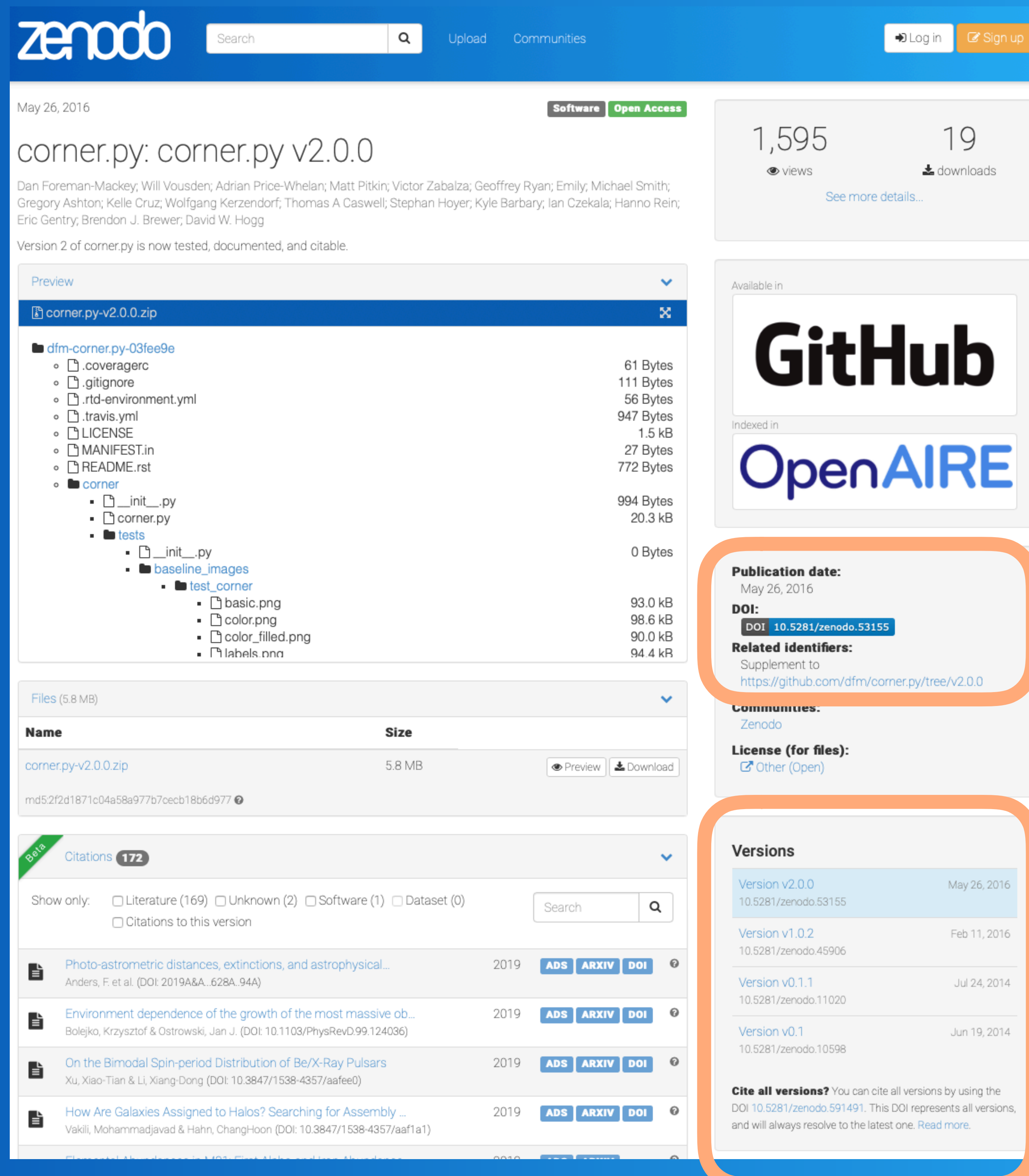
http://github.com/zenodo

open source

INVENIO

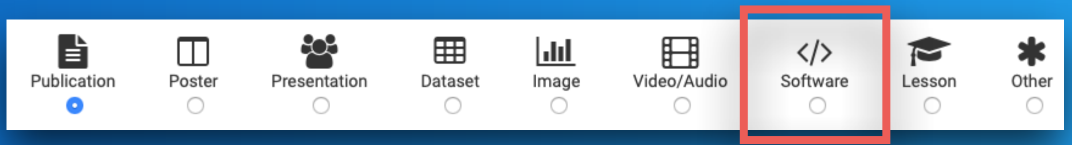
zenodo

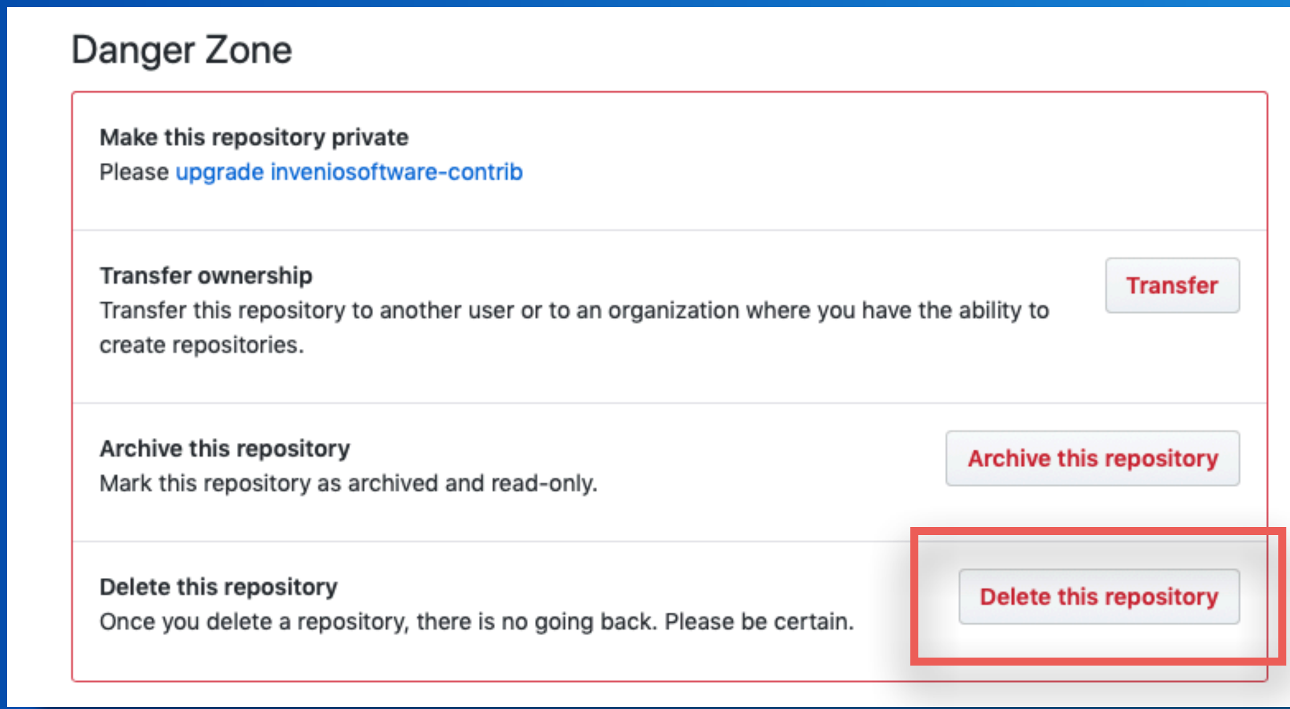
Catch-all for the long-tail of science

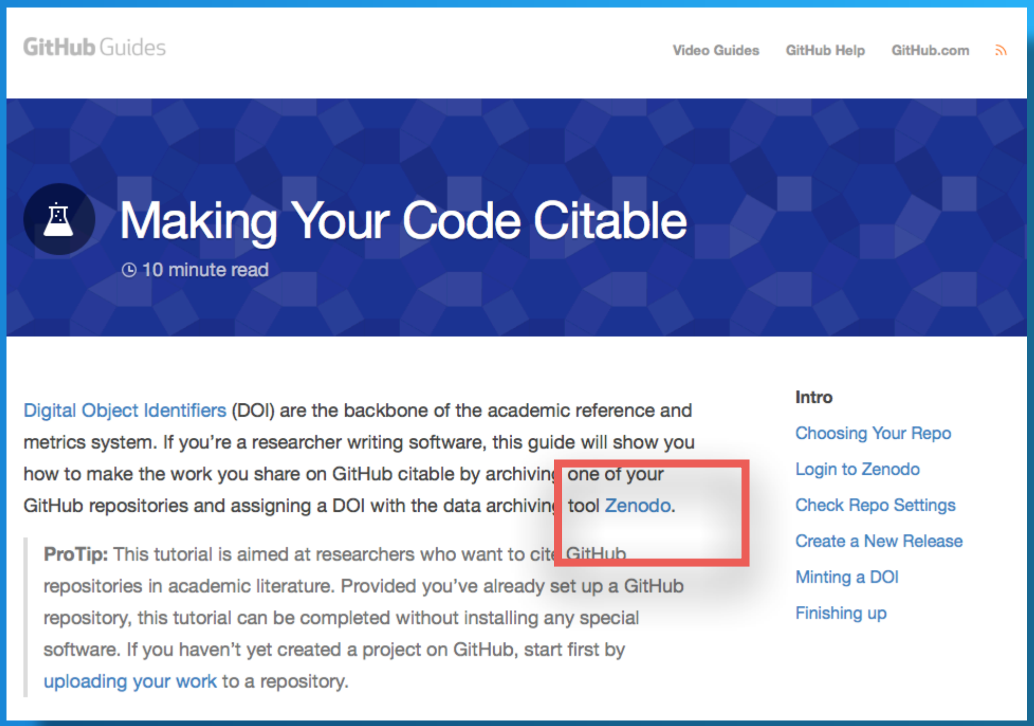

zenodo
Search
Upload
Communities
Log in
Sign up
May 26, 2016
Software
Open Access
corner.py: corner.py v2.0.0
Dan Foreman-Mackey; Will Vousden; Adrian Price-Whelan; Matt Pitkin; Victor Zabalza; Geoffrey Ryan; Emily; Michael Smith; Gregory Ashton; Kelle Cruz; Wolfgang Kerzendorf; Thomas A Caswell; Stephan Hoyer; Kyle Barbary; Ian Czekala; Hanno Rein; Eric Gentry; Brendon J. Brewer; David W. Hogg
Version 2 of corner.py is now tested, documented, and citable.
Preview
corner.py-v2.0.0.zip
dfm-corner.py-03fee9e
.coveragerc 61 Bytes
.gitignore 111 Bytes
.rtd-environment.yml 56 Bytes
.travis.yml 947 Bytes
LICENSE 1.5 kB
MANIFEST.in 27 Bytes
README.rst 772 Bytes
corner
__init__.py 994 Bytes
corner.py 20.3 kB
tests
__init__.py 0 Bytes
baseline_images
test_corner
basic.png 93.0 kB
color.png 98.6 kB
color_filled.png 90.0 kB
Files (5.8 MB)
Name
Size
corner.py-v2.0.0.zip
5.8 MB
Preview
Download
md5:2f2d1871c04a58a977b7cecb18b6d977
Beta
Citations 172
Show only:
Literature (169)
Unknown (2)
Software (1)
Dataset (0)
Citations to this version
Search
Photo-astrometric distances, extinctions, and astrophysical...
Anders, F. et al. (DOI: 2019A&A...628A..94A)
2019
ADS
ARXIV
DOI
Environment dependence of the growth of the most massive ob...
Bolejko, Krzysztof & Ostrowski, Jan J. (DOI: 10.1103/PhysRevD.99.124036)
2019
On the Bimodal Spin-period Distribution of Be/X-Ray Pulsars
Xu, Xiao-Tian & Li, Xiang-Dong (DOI: 10.3847/1538-4357/aafee0)
2019
How Are Galaxies Assigned to Halos? Searching for Assembly ...
Vakili, Mohammadjavad & Hahn, ChangHoon (DOI: 10.3847/1538-4357/aaf1a1)
2019
1,595
views
19
downloads
See more details...
Available in
GitHub
Indexed in
OpenAIRE
Publication date:
May 26, 2016
DOI:
DOI 10.5281/zenodo.53155
Related identifiers:
Supplement to
https://github.com/dfm/corner.py/tree/v2.0.0
Communities:
Zenodo
License (for files):
Other (Open)
Versions
Version v2.0.0
May 26, 2016
10.5281/zenodo.53155
Version v1.0.2
Feb 11, 2016
10.5281/zenodo.45906
Version v0.1.1
Jul 24, 2014
10.5281/zenodo.11020
Version v0.1
Jun 19, 2014
10.5281/zenodo.10598
Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.591491. This DOI represents all versions, and will always resolve to the latest one. Read more.

zenodo

Share ≠ Publish ≠ Preserve







Zenodo hosts 75% of world SW DOIs

zenodo

KEEP
CALM
AND
DO
OPEN
SCIENCE

@ludmilamarian

ludmila.marian@iaea.org

Thanks to Jose.Benito.Gonzalez@cern.ch, the CERN Digital Repository Section Leader, for providing valuable input for this presentation.