

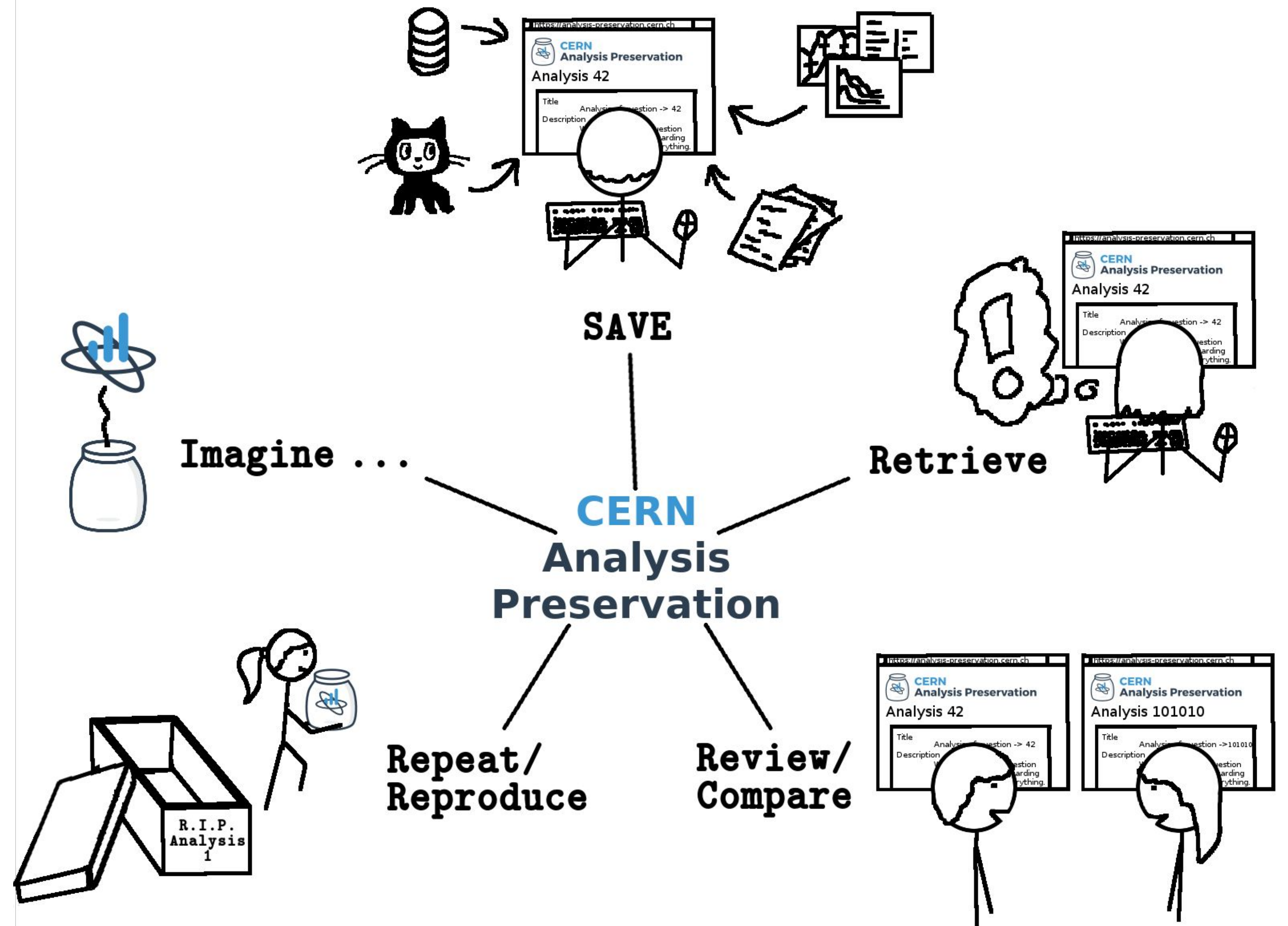


CERN **Analysis Preservation**

Contextualising analyses through data and software preservation

Robin Dasler
WSSSPE5.1
6 September, 2017

Motivation



CERN Analysis Preservation

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

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

Technology

CAP is built on the **Invenio digital library** framework

(used in CERN Document Server, INSPIREHEP, CERN Open Data and many others)

Data are modelled in **JSON format**

JSON Schema with standard metadata requirements

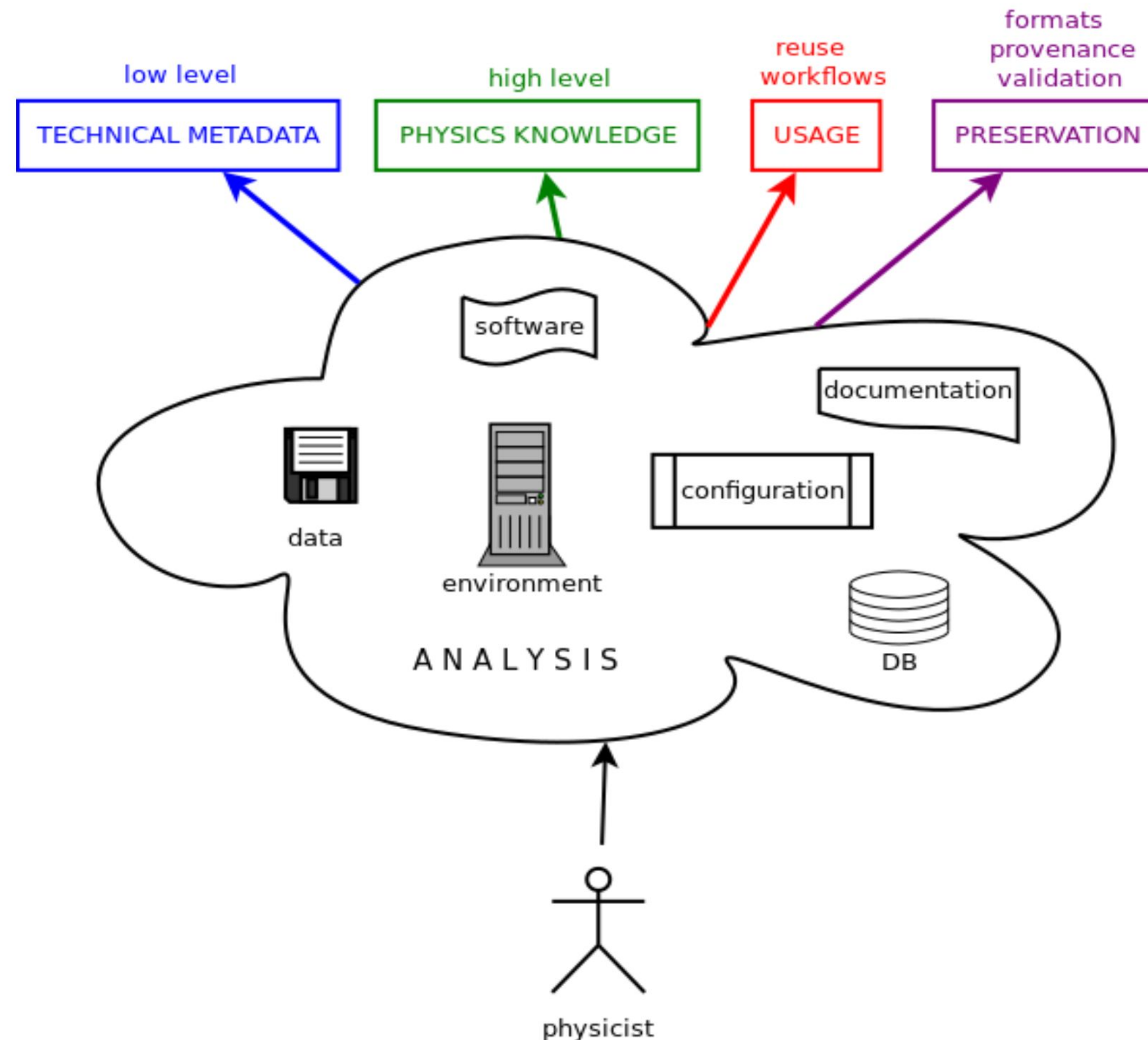
Elasticsearch cluster for indexing and information retrieval needs

Open Archival Information System (OAIS) practices to ensure **long-term preservation**

1 Describing an analysis

- ❑ W3C DCAT
- ❑ JSON Schema
- ❑ domain-specific fields

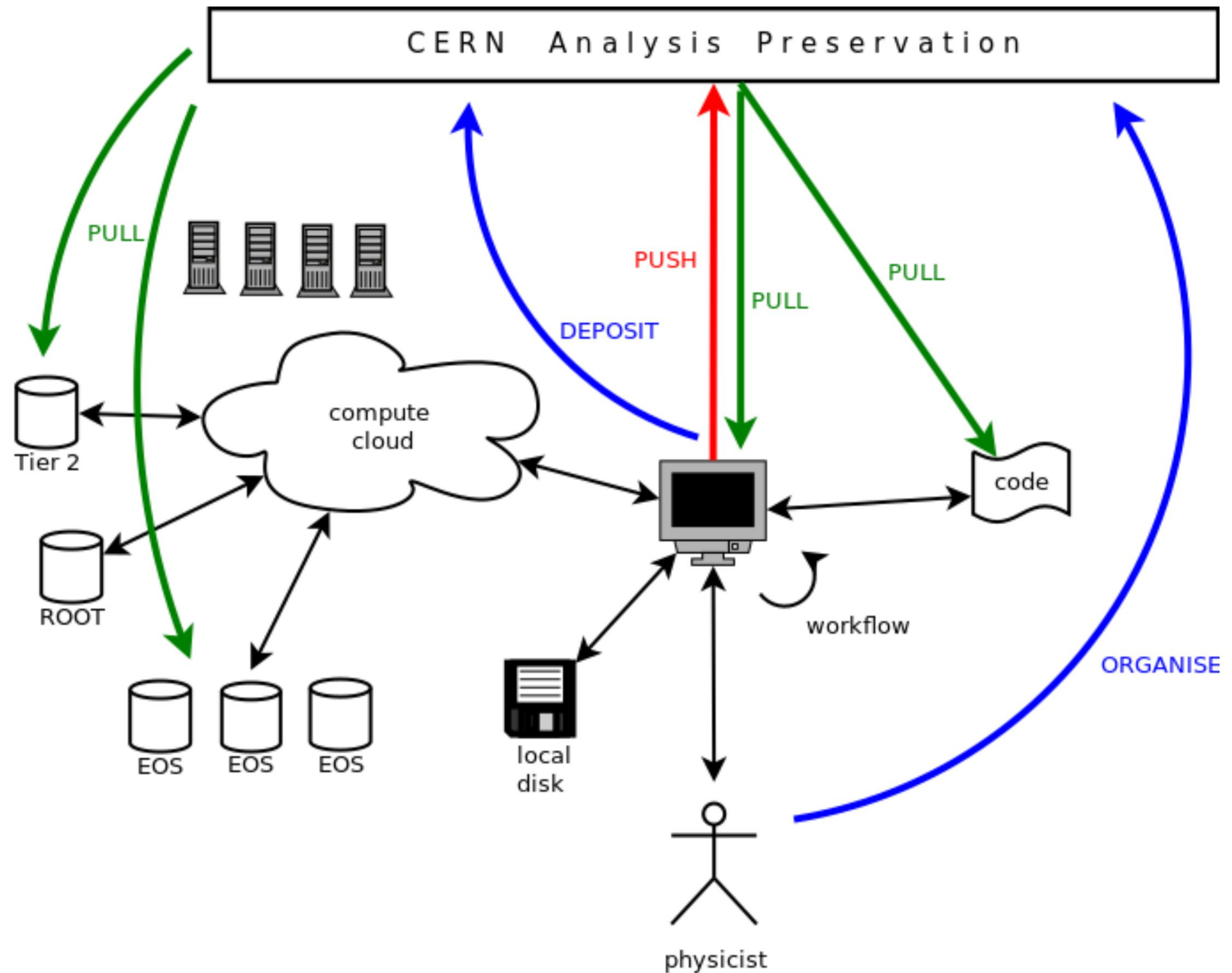
INVENIO)



Structuring knowledge behind research data analysis

2 Capturing an analysis

- datasets: local storage, cloud storage
- software: Git, SVN
- information: DBs (WG, Bookeeping, Data dependency, etc), TWikis
- protocols: HTTP, XRootD



Taking consistent snapshot of analysis assets at a certain time

2 Capturing an analysis

Submission form
with auto-complete functionality
(based on connections made to
existing LHCb databases)

CERN

Analysis Preservation

Search

Create

Home

Collaboration

Search

My Deposits

My shared deposits

My drafts

Create

LHCb Analysis

Hit ? for shortcuts

LHCb Analysis 7/2/2017, 2:41:02 PM

Save

Basic Information

Basic Information

Analysis Name

Measurement

Proponents

Status

Reviewers

Review eGroup

Working Group

Keywords

Stripping/Turbo selection

ntuple/userDST-production

User Analysis

Additional Resources

WARNING: This is just a **DEMO**. Data saved is **NOT** backed-up at the moment and might be lost during any system upgrade

BASIC INFORMATION

Please provide some information relevant for all parts of the Analysis here

</> JSON

Analysis Name

Provide a name for your analysis. This will be displayed as an analysis title when shared.

obser

✓

Observation of excited Omega_c

First **observation** of the DCS decay Xi_c+ -> p phi

B+ -> J/psi K+ K- pi+ **observation**

Bs-> J/psi phi eta **observation**

B0s -> xc0 (pi+ pi-) phi (K+ K-) **observation**

B0s->Psi(2S) K pi **observation**

Lambda_b -> Jpsi p pi- **observation**

B0s -> Xc1 (J/psi gamma) pi+ pi- **observation**

B+ ->Jpsi 5pi and B+ -> psi(2S) 3pi **observation**

B0 -> Jpsi pi0 **observation**

Lambda_b -> psi(2S) p pi- **observation**

Bs -> Jpsi Ks K+/- pi-/+ **observation**

Observation of Bc -> Jpsi p pbar pi

Observation of the Resonant Character of the Z(4430) State

Observation of Bc -> Jpsi K

First **observation** of the B*s2(5840) ->B*+K- decay and properties of the orbitally excited B0s mesons

Observation of Bs -> Jpsi K* K*

Observation of Eta_c(2S) -> ppbar decays and search for X(3872) -> p pbar

Observation of Xib2JpsiKp

Observation of Bc -> Bs pi

CERN ANALYSIS PRESERVATION - 2017

3 Reusing an analysis

Instantiating preserved analysis on the cloud

Reproduce an analysis even many years after its initial publication

How can we help you to rerun/reinstantiate your analysis in many years to come?

What tools do you use already, what tools do we need to use to make this happen?

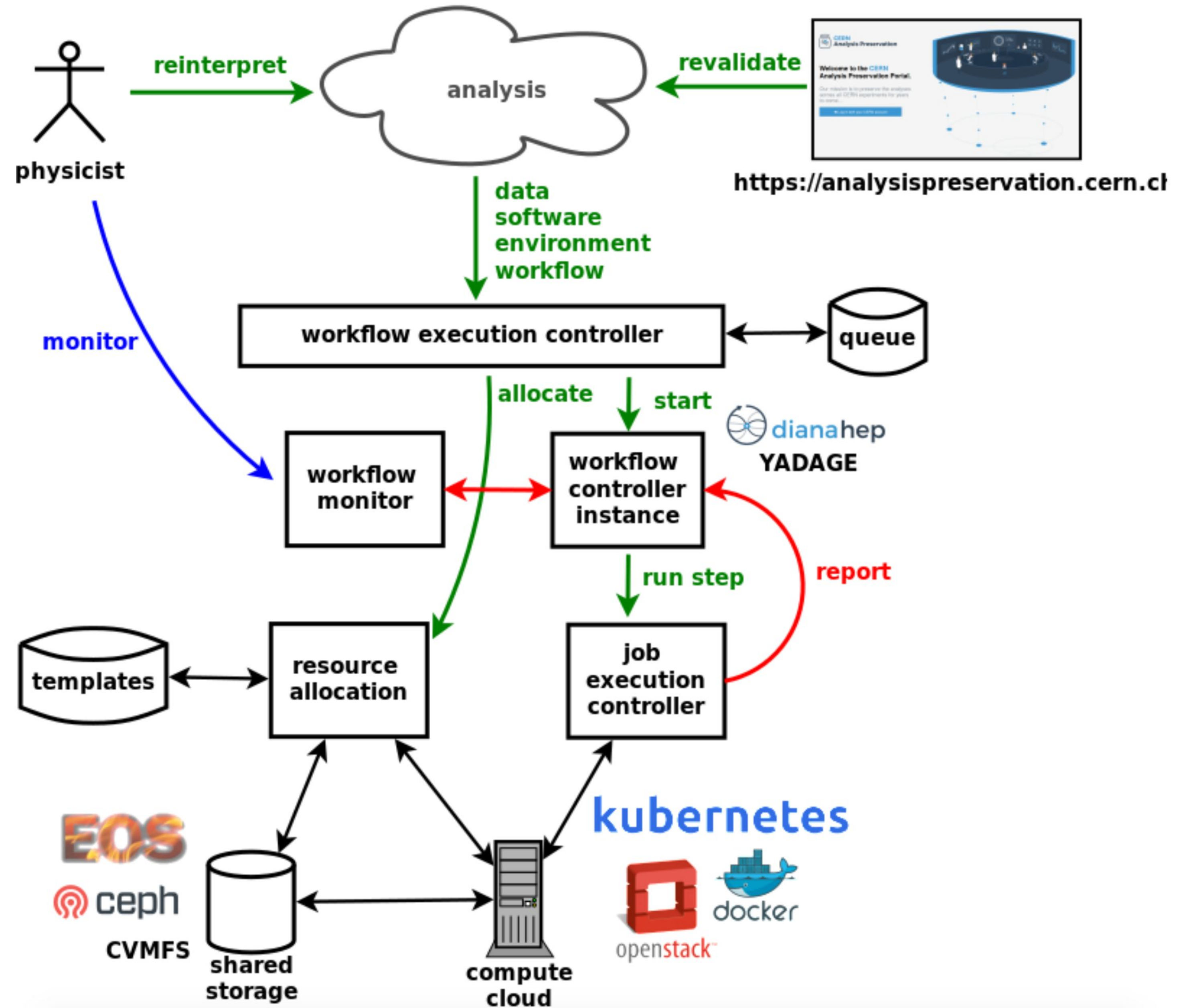
What are the blockers?

What is missing?

Extend impact of preserved analyses through validation and recasting services

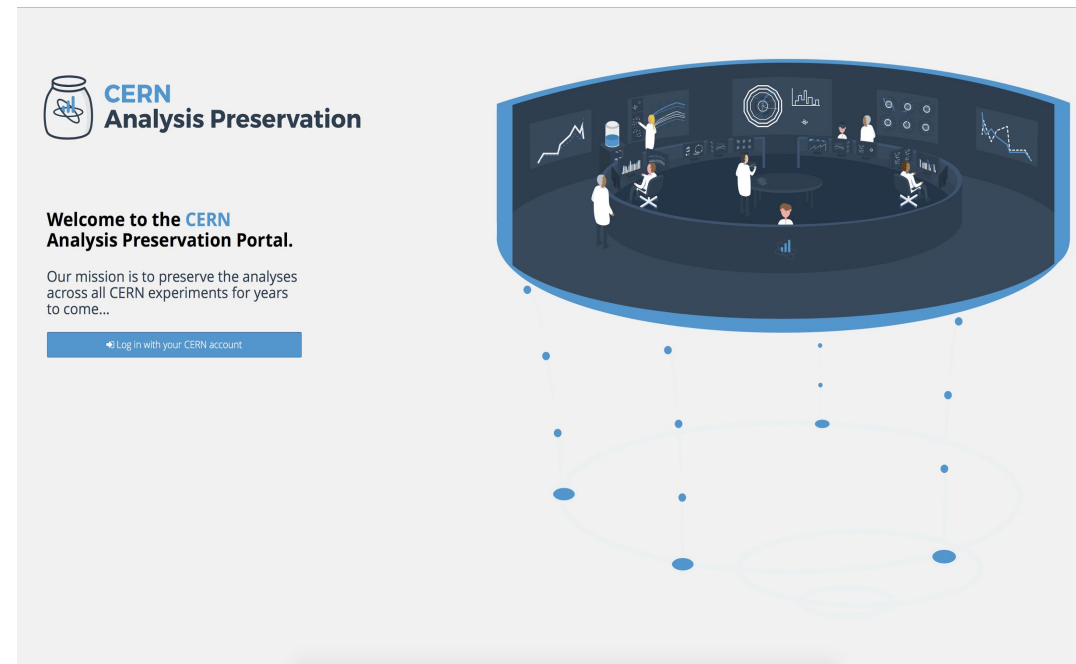
3 Reusing an analysis

CAP/REANA project



Development

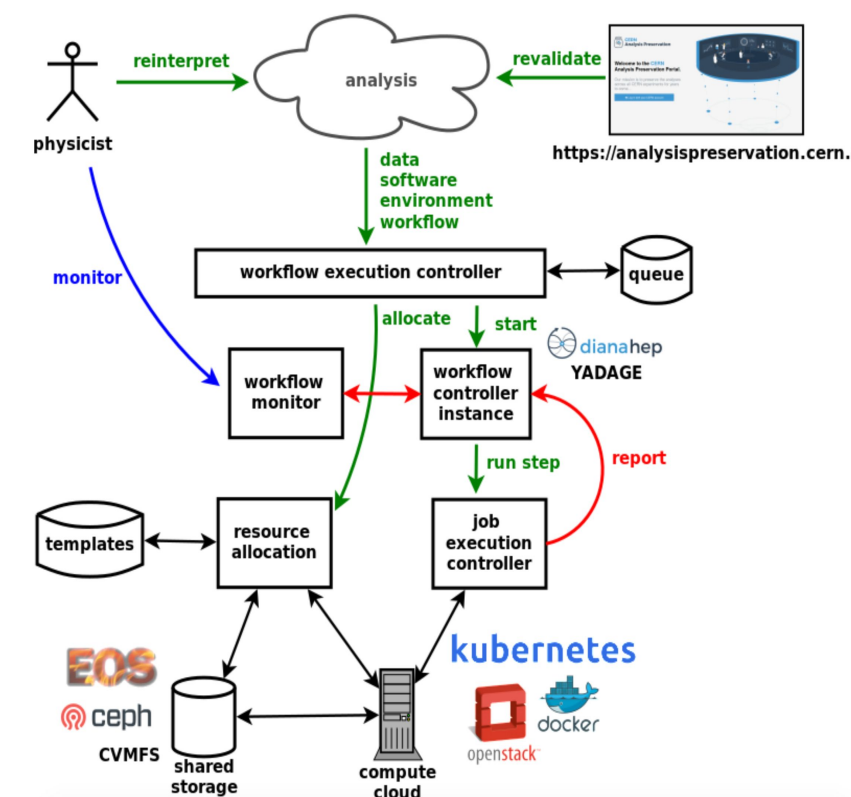
- Open Source
- Openly accessible
- Collaborative
- Transparent roadmap



CERN Analysis Preservation

<http://analysispreservation.cern.ch>

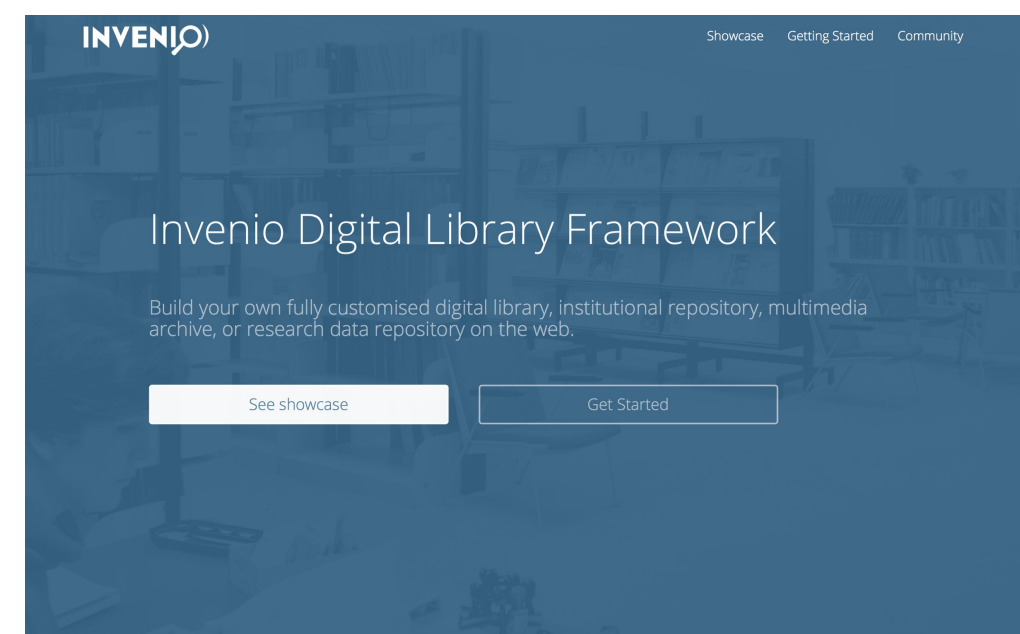
<http://github.com/cernanalysispreservation>
analysis-preservation-support@cern.ch



REANA

<http://reanahub.io>

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



Invenio

<http://inveniosoftware.org>
<http://github.com/inveniosoftware>
[@inveniosoftware](mailto:info@inveniosoftware.org)
info@inveniosoftware.org

Thanks to

S. Dallmeier-Tiessen², R. Dasler², P. Fokianos², J. Kuncar¹,
A. Lavasa², A. Mattmann², D. Rodríguez¹, T. Simko¹, A. Trzcinska², I. Tsanaktsidis²

¹*CERN Information Technology*

²*CERN Scientific Information Service*