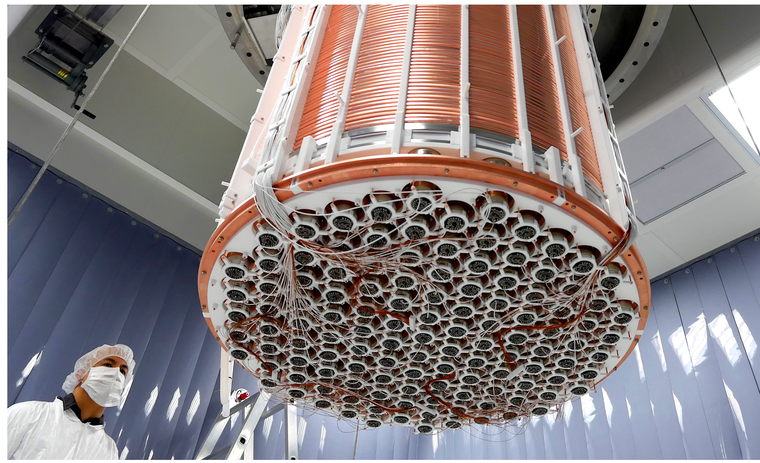
ESCAPE &
DM Test Science Project
and HEP Open Data

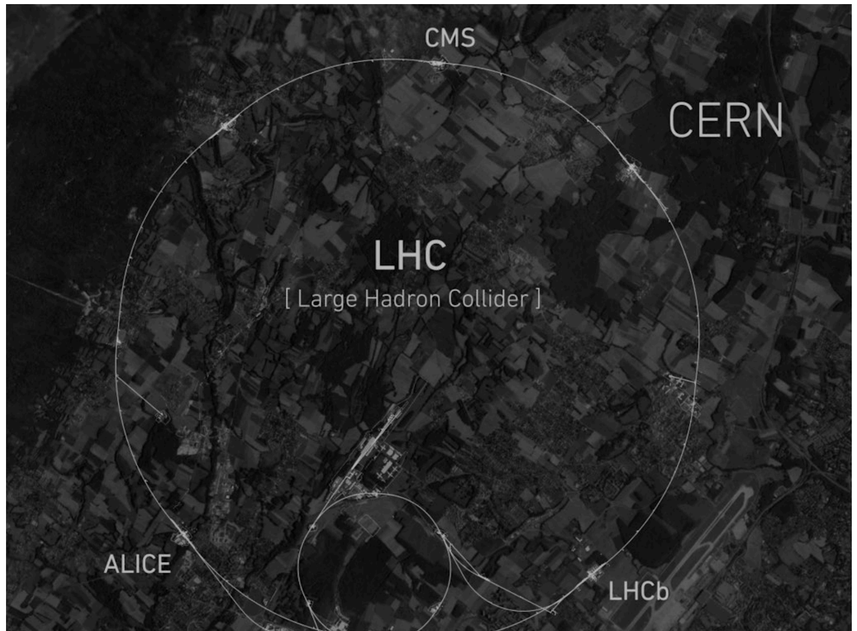
Lukas Heinrich, TUM

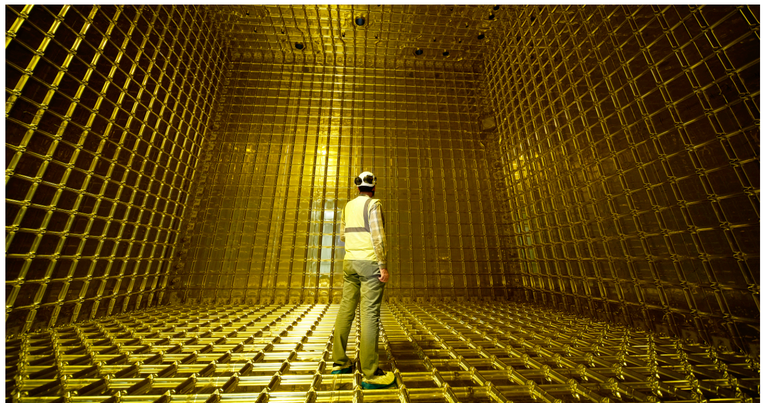
Open Science in the Big Science Era

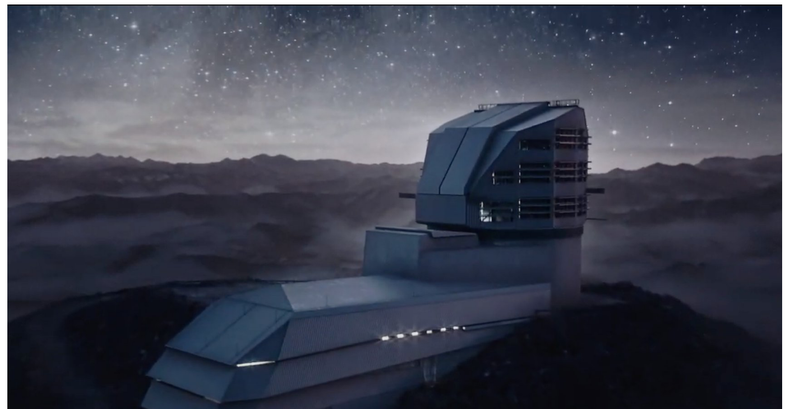
Cutting Edge Scientific Experiments are increasingly unique

- complexity often only allows for 1 or a few instruments worldwide





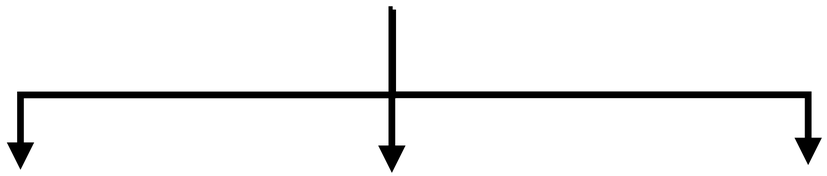




Open Science in the Big Science Era

Given this setup, it's imperative that we try to maximize the scientific output for those instruments

Three broad directions



Creation

create entirely new studies based on existing data

Combination

combine multiple results studying the same question

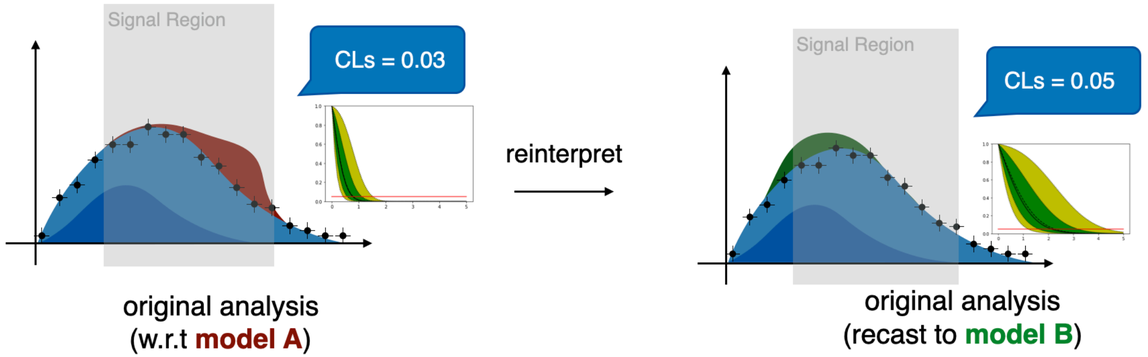
Reinterpretation

repurpose a study prepared for question A for a new question B

Reinterpretation and Combination

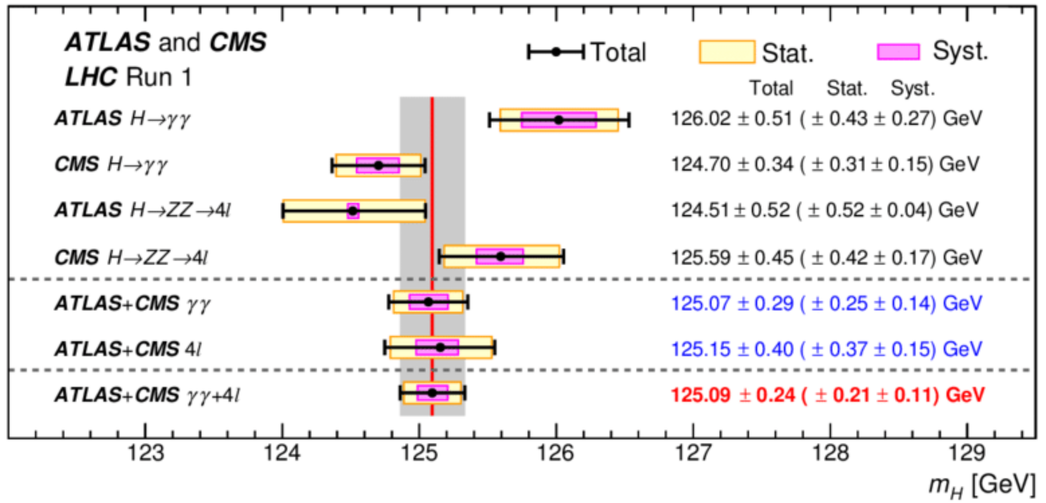
Reinterpretation

$$p(x\,|\,\theta) \rightarrow p(x\,|\,\phi)$$



Combination

$$p_{\mathrm{ana}_1}(x\,|\,\theta) \cdot p_{\mathrm{ana}_2}(x\,|\,\theta)$$

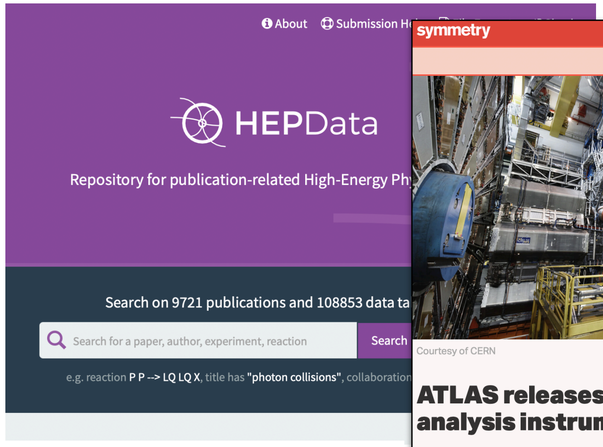


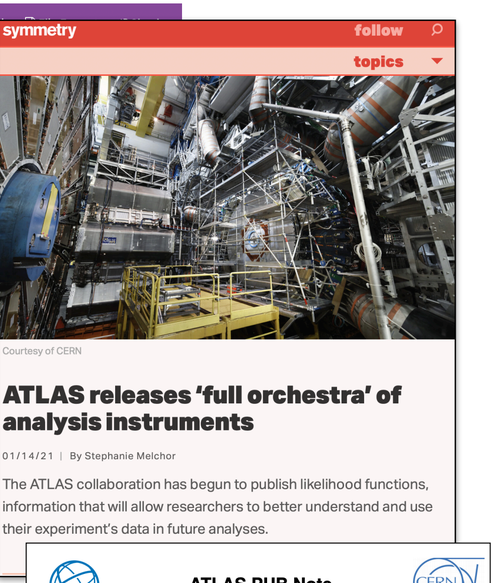
Recent Milestones

A lot of recent progress around the community to enable Open Science

- New HepData with improved web-based APIs
- Likelihood and ML model Publishing
- Systematic RECAST (Reinterpretation) at LHC
- ORIGINS Dark Matter Data Center
- Common CERN Open Data Policy

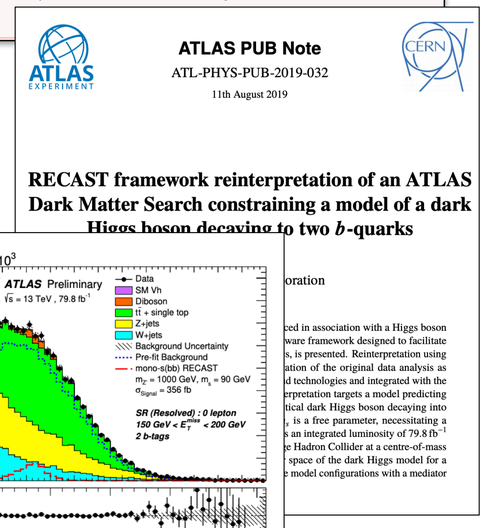
Good time to capitalize on the momentum.





ONNX

DMDC
ORIGINS Dark Matter Data Center



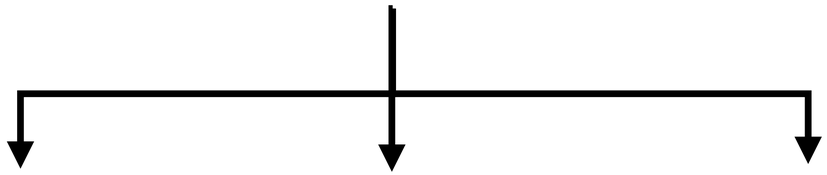
Open Science in the Big Science Era

If we want to extract the most science we need to go beyond confines of a single experiment, and share data & workflows openly

(especially for science questions that cut across multiple domains)

Instruments: develop **policies & data formats** designed for reuse

Wider Community: develop **infrastructure** that makes it actually FAIR



Open Creation

study-independent Open Data

Open Combination

study-dependent Open Data (data products)

Open Reinterpretation

study-dependent Open Analysis Workflows

European Open Science Cloud

EUROPEAN OPEN SCIENCE CLOUD

[EOSC]

Long-Term Effort by the EU to develop infrastructure for Open Science

Mission:

to provide European researchers […] and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services […]

Concretely: Not yet a “place” (a la commercial clouds) but rather a project / funding stream to develop “Web of FAIR Data and services”

- 250M€ 2018-2020, ~1B€ next 7 years

ESCAPE

HNEP & Astro traditionally very strong in building community infrastructure (e.g. WWW, Worldwide LHC Computing Grid, Zenodo, INSPIRE, …)

- EU HNEP institutions deeply involved in using contributing our experience & institutional knowledge to shape of EOSC
- ESCAPE: Umbrella project to develop Open Science Research Infrastructure for Astro & HEP communities and data

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

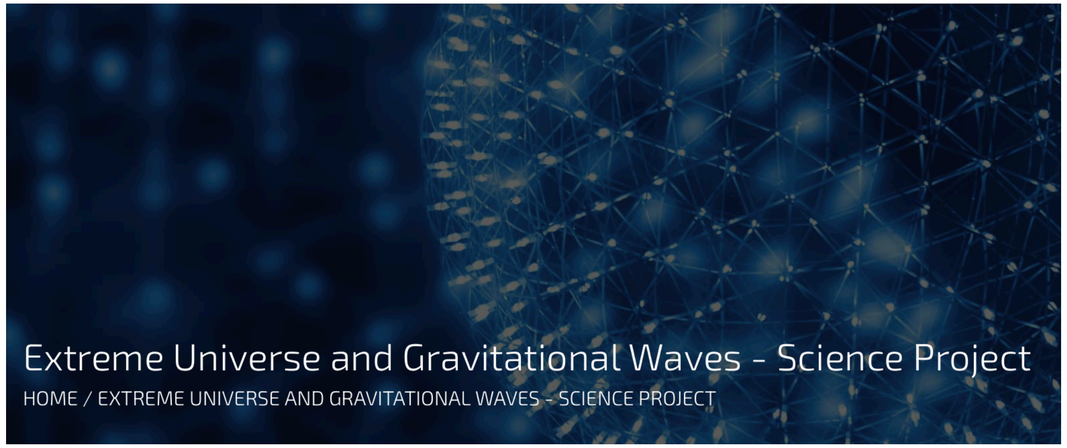
ESCAPE Services

ESCAPE aims to develop the full gamut of required infrastructure (data & software repositories, analysis platforms, authentications, …)

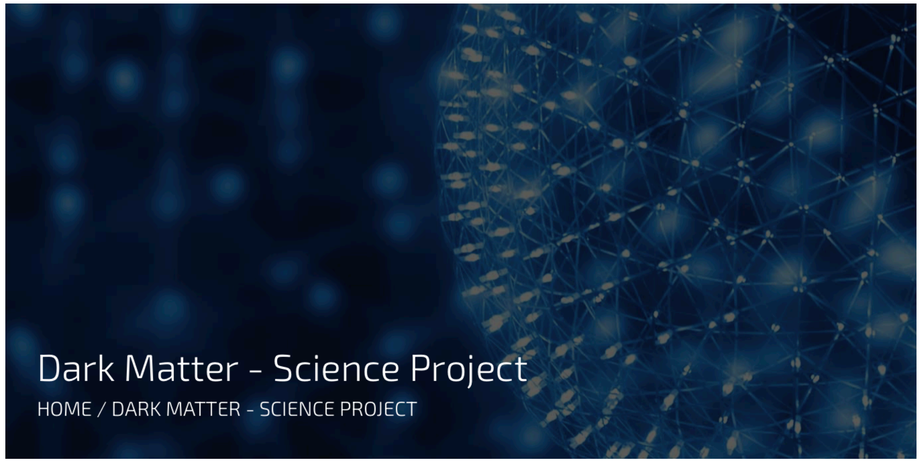
One Key Metric: can we use this infrastructure to extract new science openly?

Test Science Projects

ESCAPE has two “Test Science Projects” to serve as proxies to demonstrate the multi-domain integration and usability of the ESCAPE infrastructure

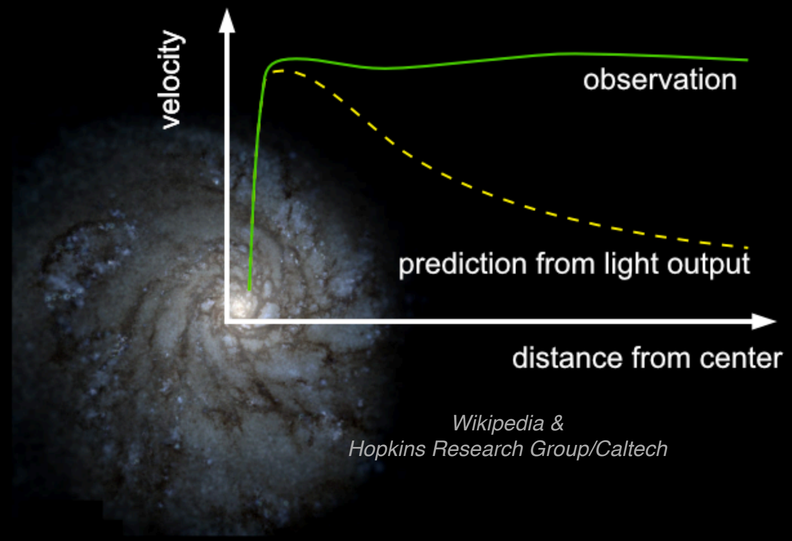


Combined data-sources to provide integrated platform for Multi-messenger Astronomy



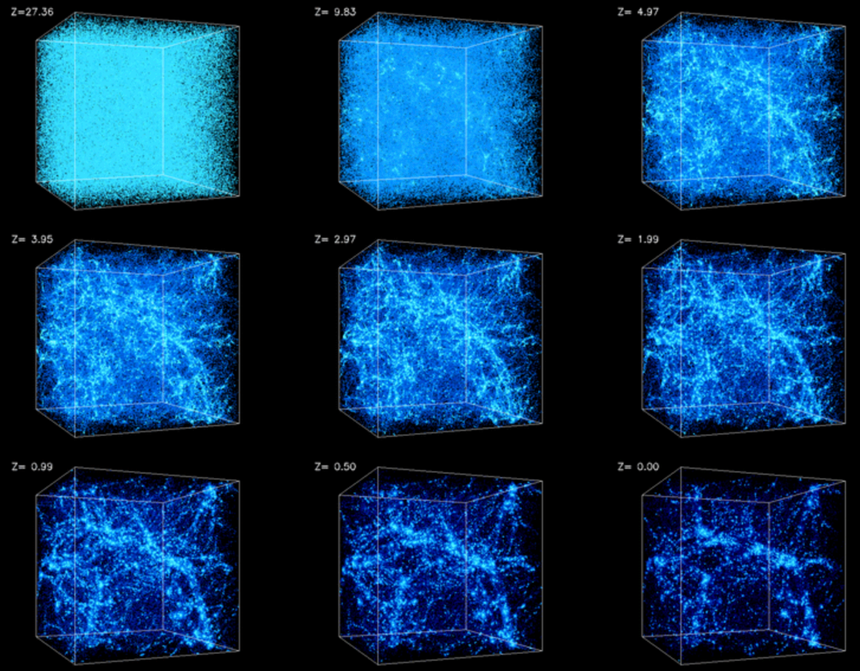
Cross-Experiment Open Analysis of Dark Matter data across Collider, Direct Detection, Neutrino, …

Big science question: Dark Matter



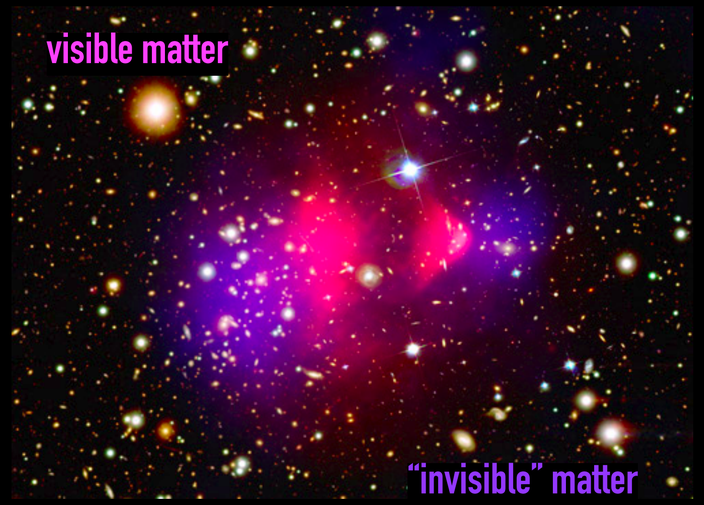
Wikipedia &
Hopkins Research Group/Caltech

Vera Rubin,
© Washington Times & Zuma

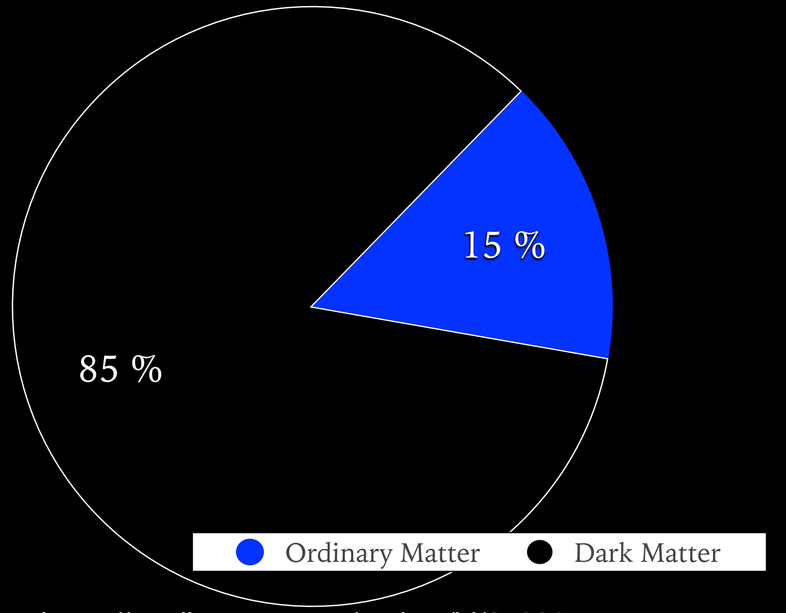


Simulations were performed at the National Center for Supercomputer Applications by A. Kravtsov and A. Klypin.

→ many possible DM scenarios
→ many ways to detect it
→ many data sources



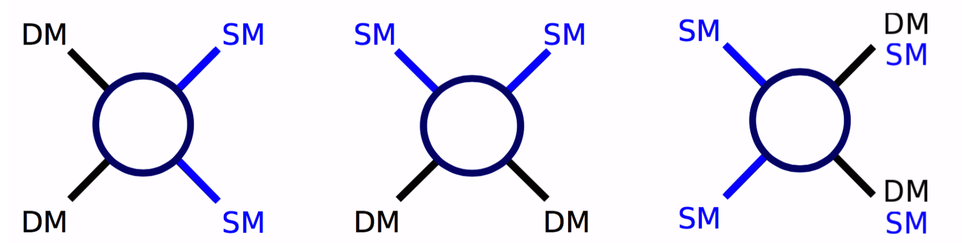
NASA/CXC/M. Weiss



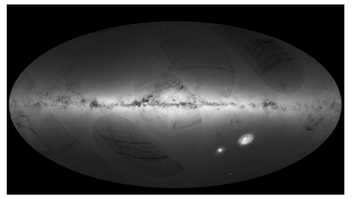
https://cordis.europa.eu/project/id/679305

Dark Matter TSP

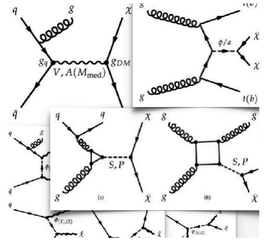
Dark Matter research is distributed across many communities



Indirect Detection | Direct Detection | Colliders

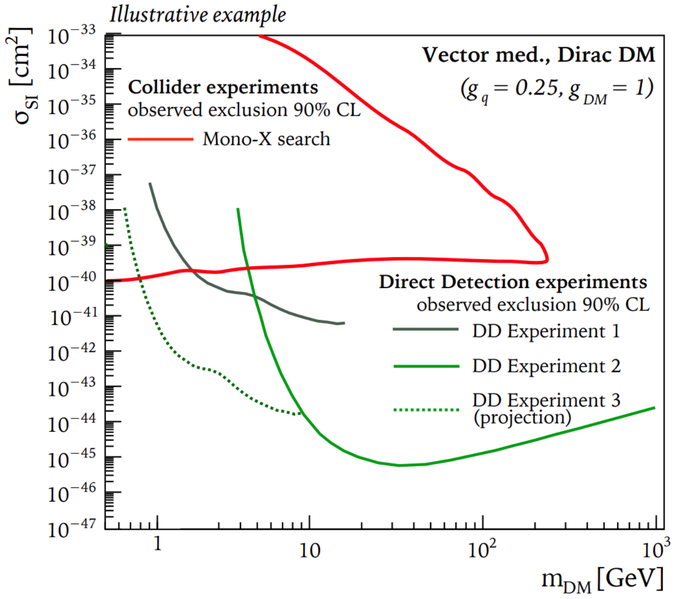


Astrophysics



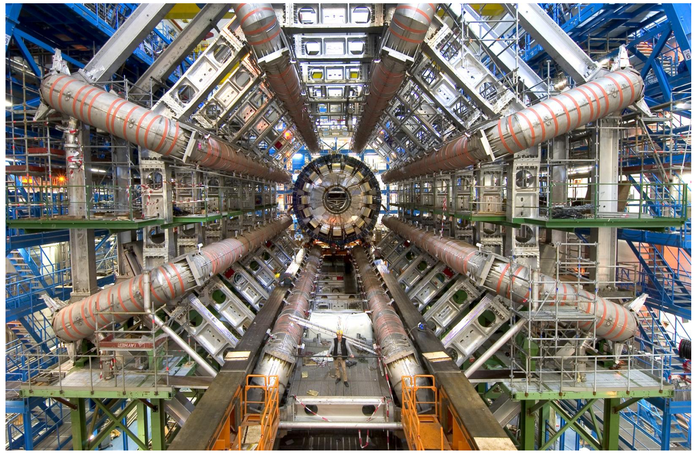
Theory

- slated to benefit particularly from the Open Science Infrastructure developed by ESCAPE
- **Domain Science Result:** new, combined exclusion bou from multiple data sources
- **Open Science Result:** demonstrate ESCAPE usability, feasibility of science w/ open infrastructures

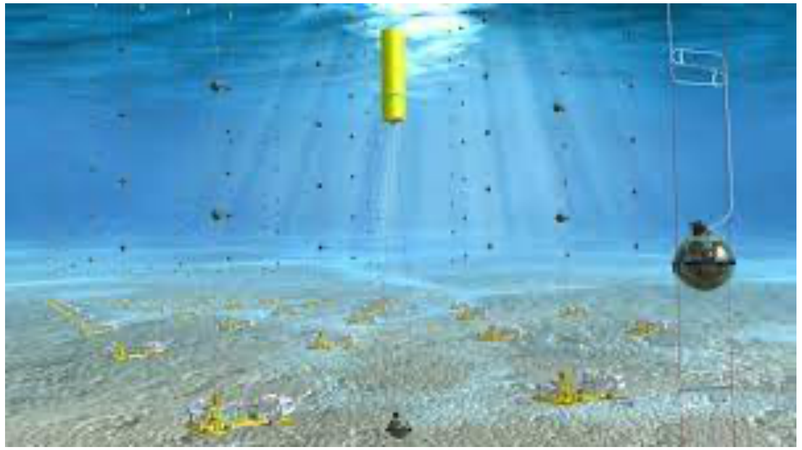


Participants in DM TSP

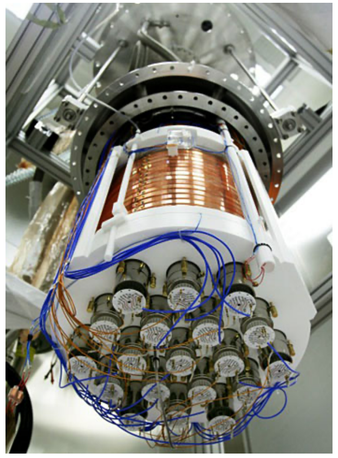
Experiments



ATLAS



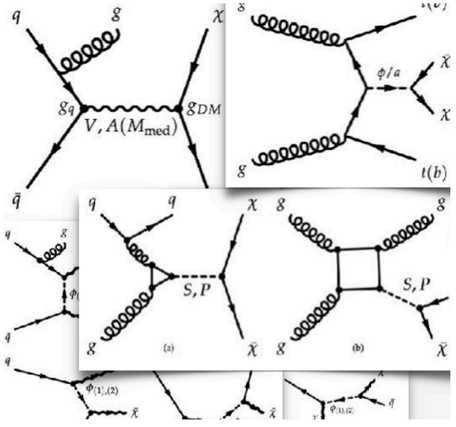
KM3NeT



DarkSide

Computing:

Theory:



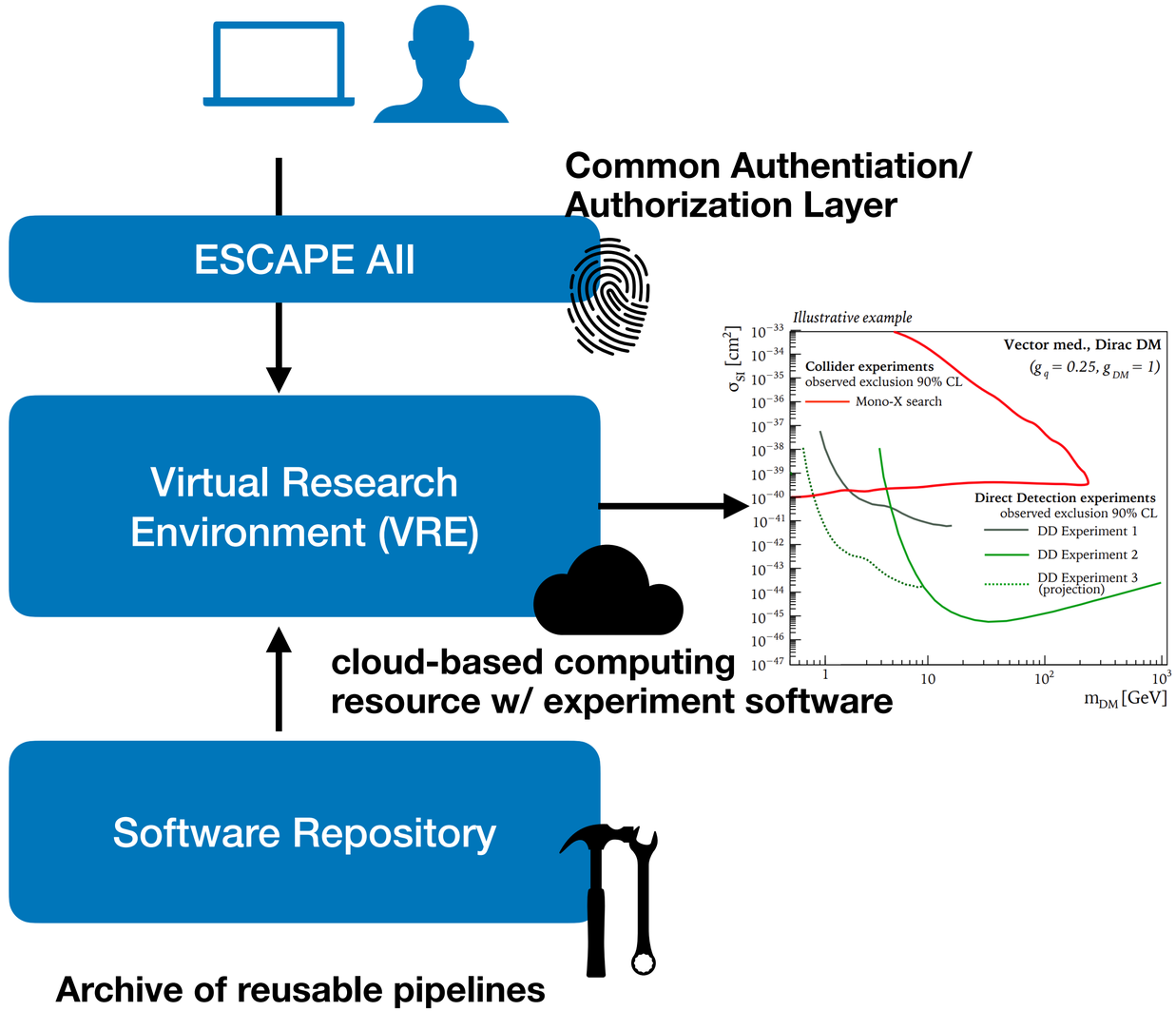
Used Infrastructure

DM TSP integrates with a number of infrastructure components from ESCAPE

Data Lake: RUCIO

VRE: jupyter, reana, kubernetes

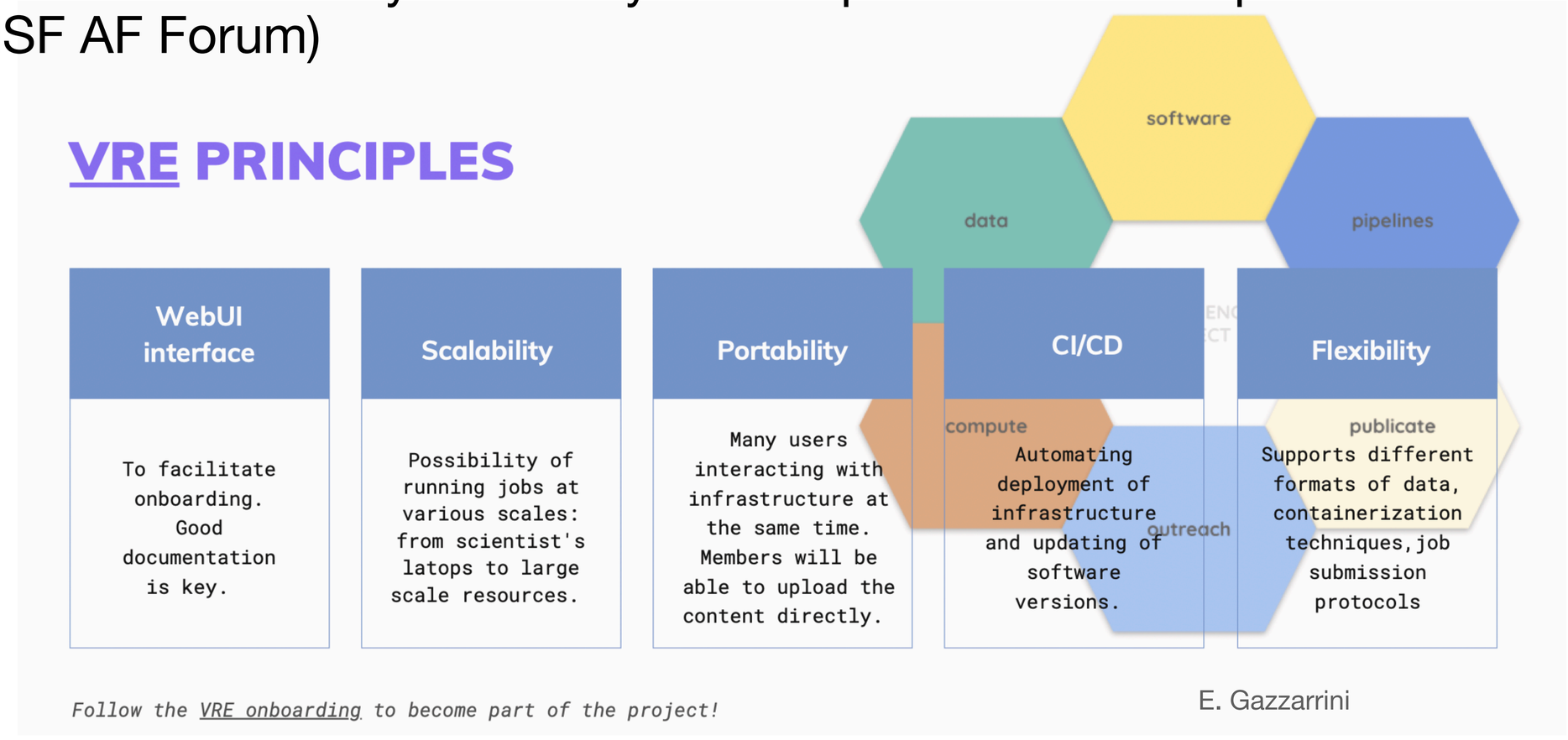
Software Repo: zenodo



Virtual Research Environment

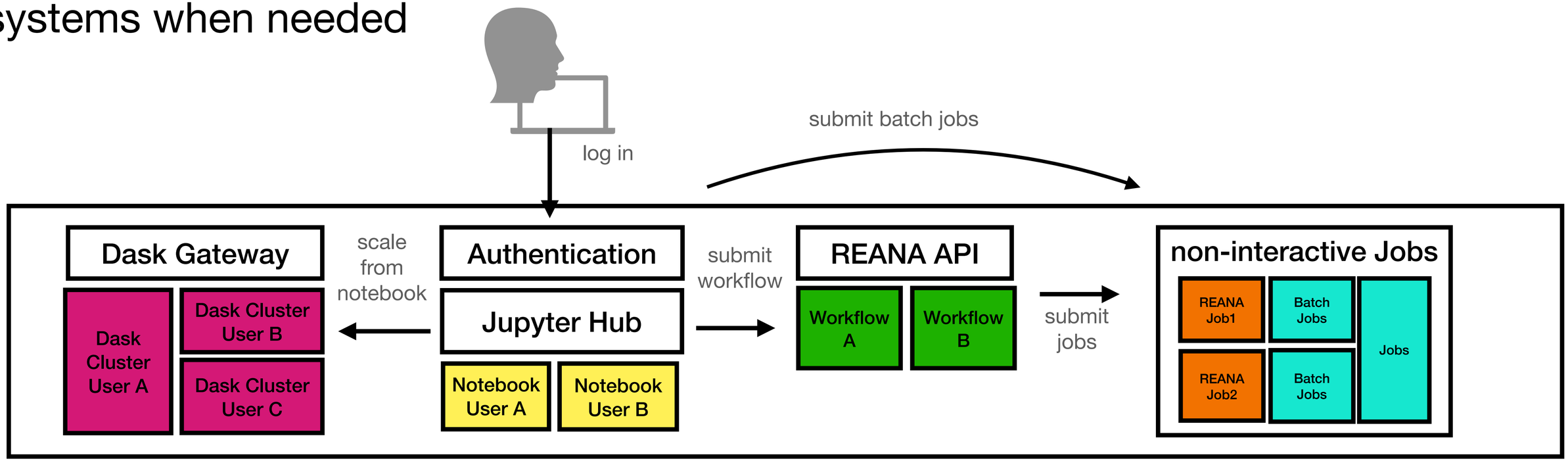
Development of a common, shared cloud-based infrastructure with native integration to Open Science data & software

Closely connected to “Analysis Facility” developments within experiments (see e.g. HSF AF Forum)



Virtual Research Environment

Jupyter-based interface allows a convenient entry-point and immediate resource for small-scale computation but also interface to scale-out & batch systems when needed



DASK

in-application scale-out

jupyter

Interactive Sessions

reana

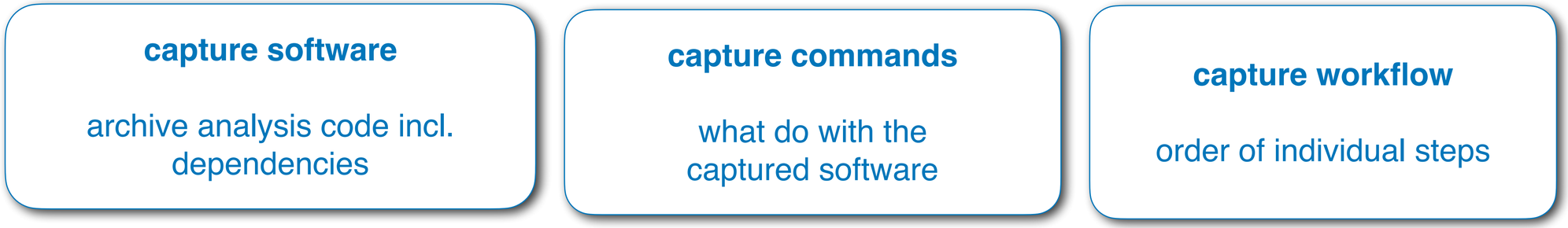
Workflows

Batch

Workflows

Most scientific computations are “workflows”: sequence/graph of interdependent steps: Often only know “within the minds” of analyzers

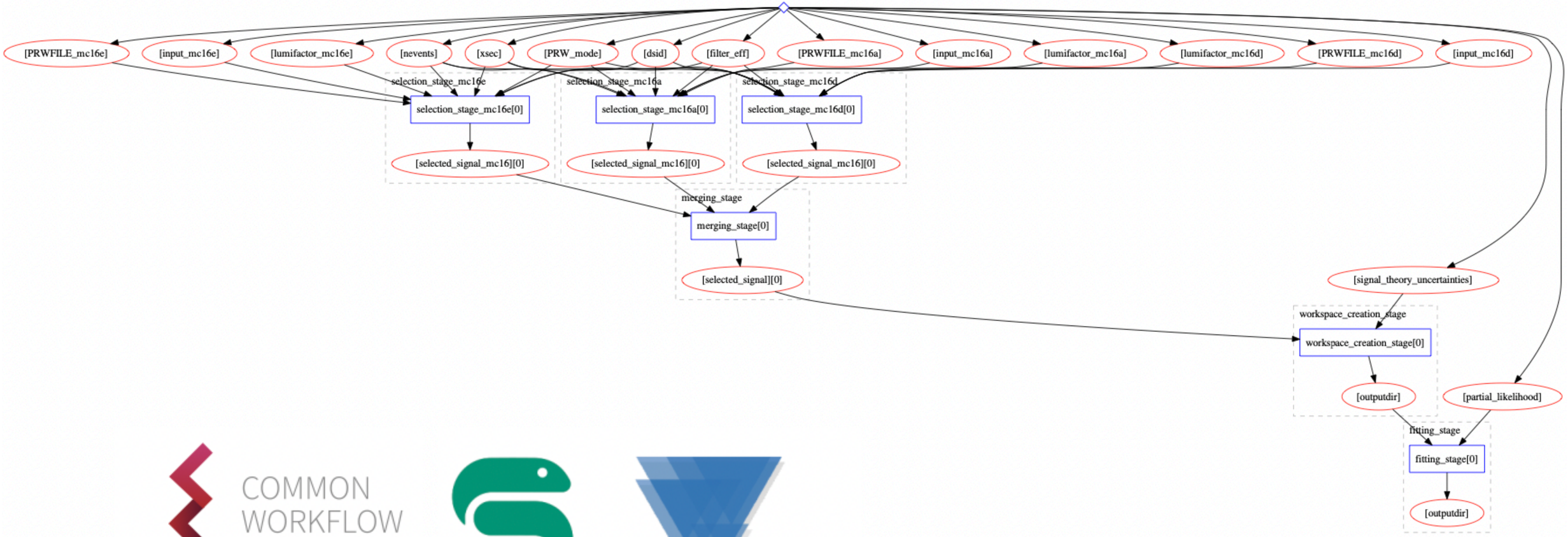
But to be reusable, we need to capture these explicitly. Workflow languages (e.g. CWL, yadage, snakemake, Luigi, …) allow to archive the workflow as a **data product in its own right**



Linux Container
(Docker, Kubernetes, …)

Workflow Languages
(yadage, luigi, snakemake, CWL, …)

Workflows Example

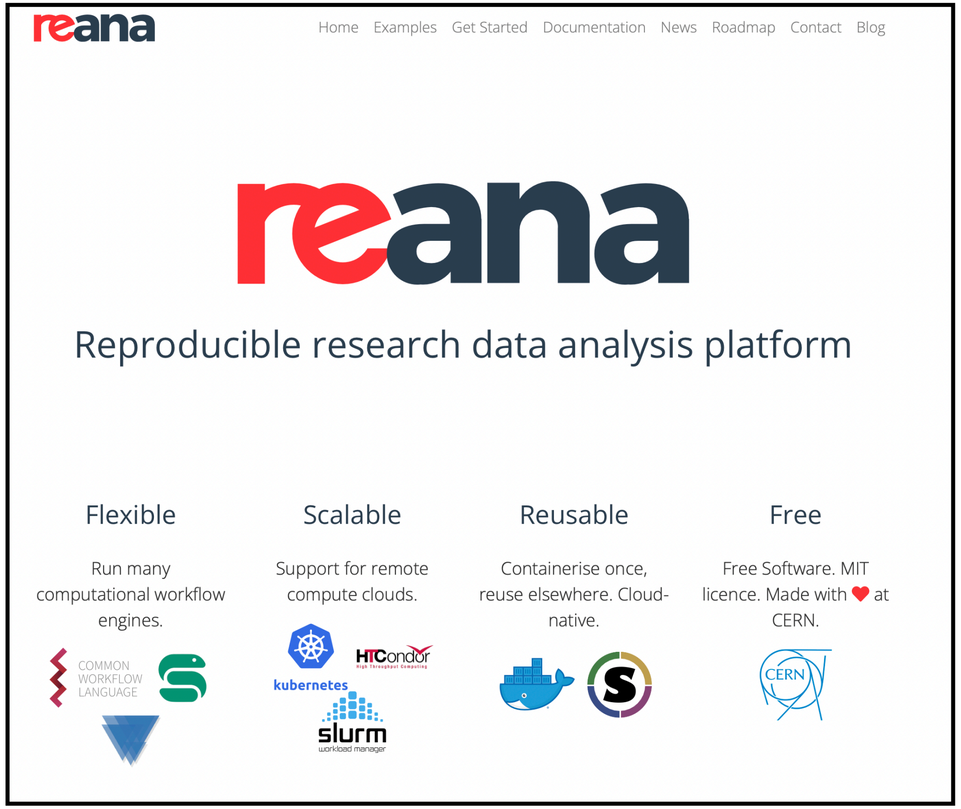


REANA

REANA is a platform to run “workflows-as-a-service”: User submits workflows and input data, REANA manages batch submissions, ordering, …

Deployable as a component inside of the ESCAPE Virtual Research Environment

Will be used to run heavier computations inside of the DM TSP (e.g. ATLAS workflows)



First Peek at working Infrastructure

ATLAS Open Data + ESCAPE Infrastructure = Plots

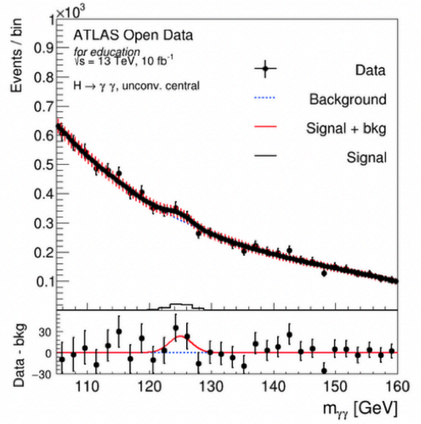
reana.yaml

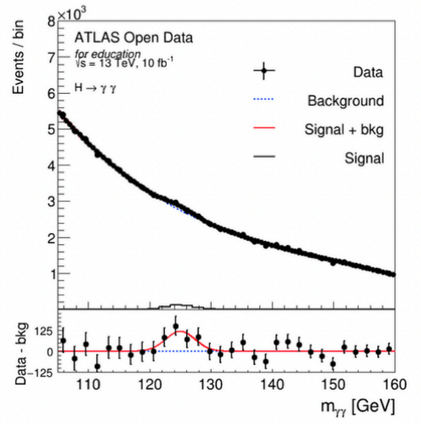
```yaml

# HYY ANALYSIS

version: 0.6.0
inputs:
  directories:
    - Analysis/HyyAnalysis/
    - Plotting/
  ...
  files:
    - HyyAnalysis.C
    - HyyAnalysis.h
    - main_HyyAnalysis.C
    - HyyAnalysisHistograms.h
    - run.sh
...
workflow:
  type: serial
  specification:
    steps:
      - name: atlas-13tev
        environment: 'reanahub/reana-env-root6:6.18.04'

        commands:
        - cd Analysis/HyyAnalysis/ && chmod 755 run.sh && mkdir -p Output_HyyAnalysis && echo -ne "0\n0" | ./run.sh
        - cd Plotting && chmod 755 plotme.sh && mkdir -p histograms && echo -ne "9\n0"  | ./plotme.sh
outputs:
  files:
    - Plotting/histograms/hist_mYY_cat_bin1.png
    - Plotting/histograms/hist_mYY_bin1.png

```

ON TERMINAL

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

$ ana=Hyy
$ reana-client create -n $ana
$ export REANA_WORKON=$ana
$ reana-client upload
$ reana-client start

...     2 minutes later...

$ reana-client download
```

OUTPUT





Outlook & Next Steps

- Lot of momentum in large-scale science projects to push Open Science methods & policies. Supported by a clear policy signals from funding agencies.
- EOSC & ESCAPE are EU-funded efforts to provide common infrastructure
- The Dark Matter TSP is a opportunity to collate the various inputs from Dark Matter searches and demonstrate the added value from a shared Open Science Infrastructure
- Expecting first science results within the next year