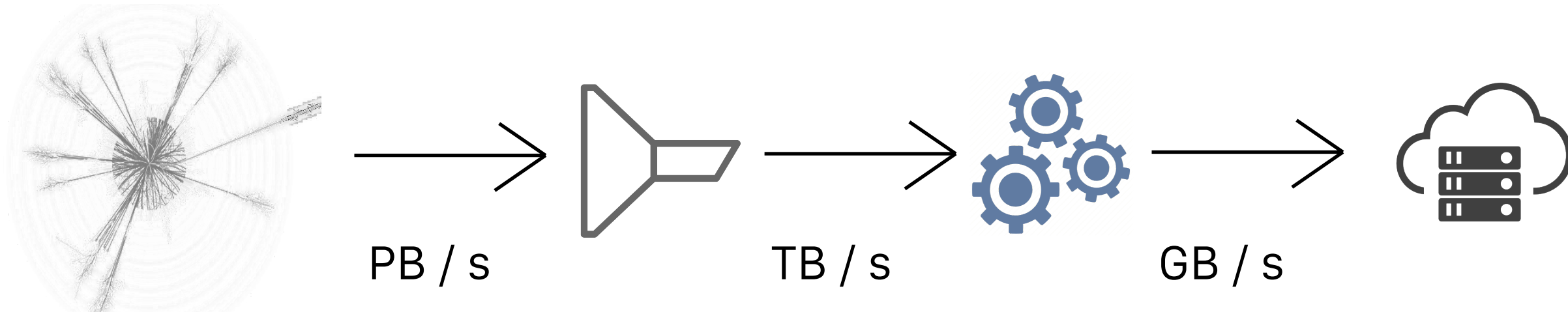


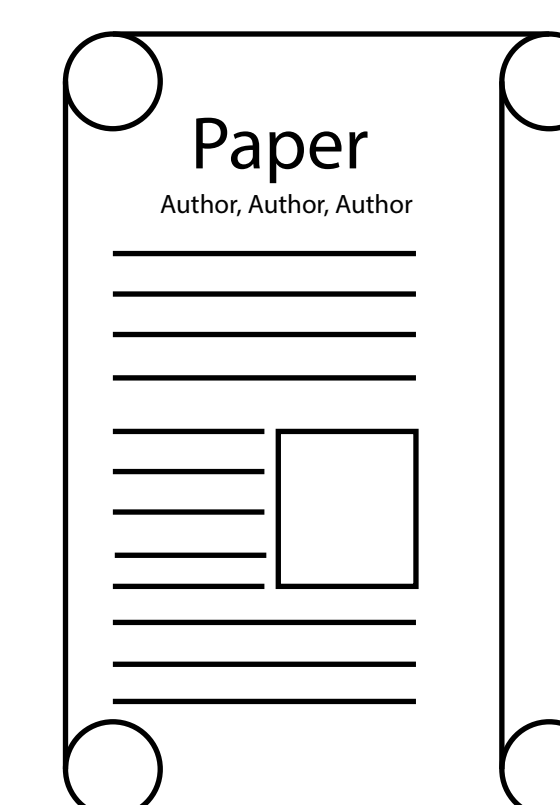
## HEP Experiments

All starts with an experiment, and the data collected from it.  
The amount of data that can get generated from a H.E.P. experiment can be up to 6 PB / s.



## Data analysis

The data analysis is performed by a custom and complex workflow, and dozens of dockerized intermediate steps.



## Paper creation

After analyzing data with a custom workflow, a paper is written. These papers contain insights on some subset of the original data, and are highly dependent on the workflow that researchers choose to generate the results.

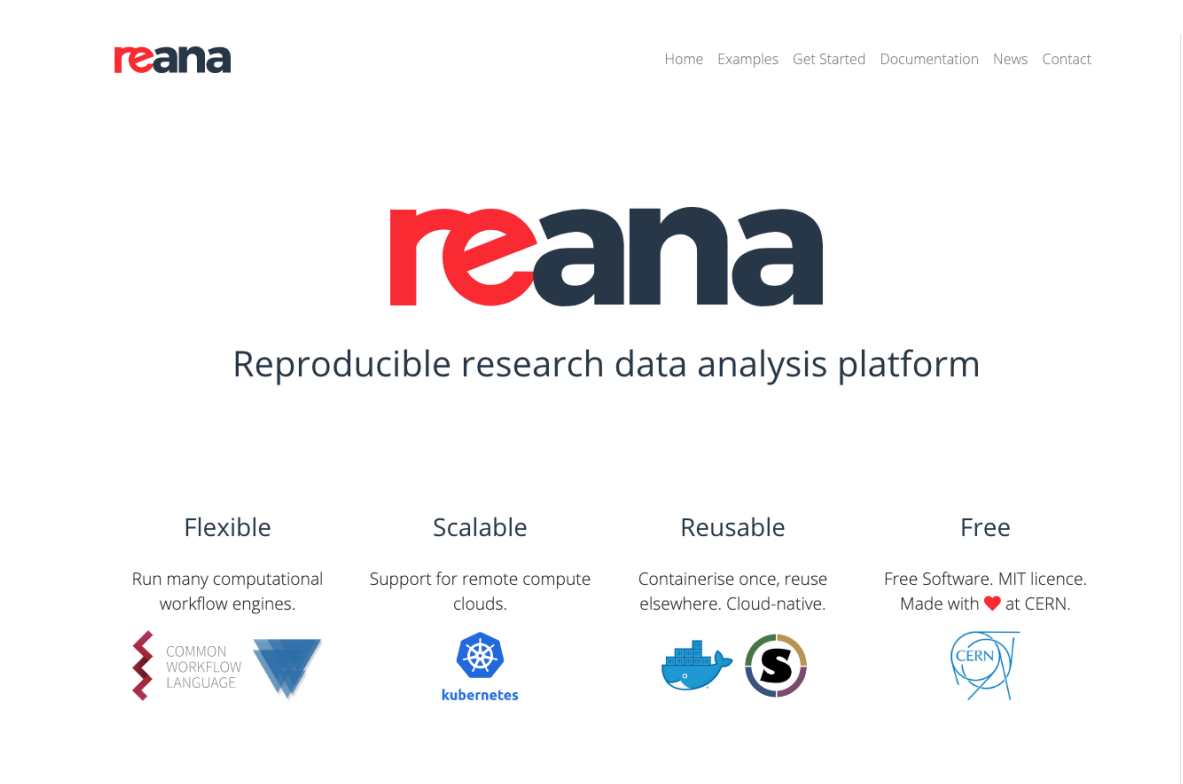
Once the paper is published, we would also like to associate additional products (derived data, figures, tables, likelihoods...). These services are provided by various cyberinfrastructure components.

## Preservation, reproducibility and reinterpretation

*"In Particle Physics context, the immense cost and complexity of the experimental set-up essentially make the independent and complete replication of HEP experiments unfeasible and unhelpful."*

*HEP experiments are set up with unique capabilities, often being the only facility or instrument of their kind in the world; they are also constantly being upgraded to satisfy requirements for higher energy, precision and level of accuracy."*

Chen, X., Dallmeier-Tiessen, S., Dasler, R. et al. Open is not enough. Nature Phys 15, 113–119 (2019)



### REANA

Reusable and reproducible research data analysis platform that helps researchers to structure their input data, code, environments and computational workflows so that the analysis can be instantiated and run on remote compute clouds.

REANA was born to target the use case of particle physics analyses, but is applicable to any scientific discipline.



### Scikit-HEP / pyhf

Compute likelihoods for the running experiments.

A likelihood is a statistical model for multi-bin histogram-based analysis and its interval estimation is based on the asymptotic formulas of "Asymptotic formulae for likelihood-based tests of new physics".

Originally from dianahep



### Git repositories

Dockerized code and workflows specifications are stored in different GitLab and GitHub instances.



### CERN Analysis Preservation (CAP)

CAP is a data preservation service run at CERN currently targeted at preserving analysis information from the four largest LHC experiments.

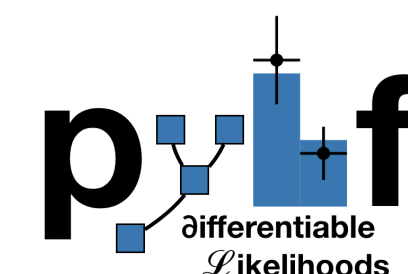
It makes it easier for physicists who are meant to have access to this analysis information to find, understand and use it just as they have done before but with less effort involved.



### RECAST

Project to extend the results of HEP analysis with reusable workflows.

It allow scientists to pick any of the available workflows, and substitute any step by the desired dockerized code. Thanks to this, fine adjustments can be done in future runs of the same workflow.



### Analysis Database

An analysis database linking multiple workflow runs with a specific paper, and multiple papers with specific runs.

It will serve as the bridge between the dynamic reinterpretation and the static information of a paper.

Work in progress

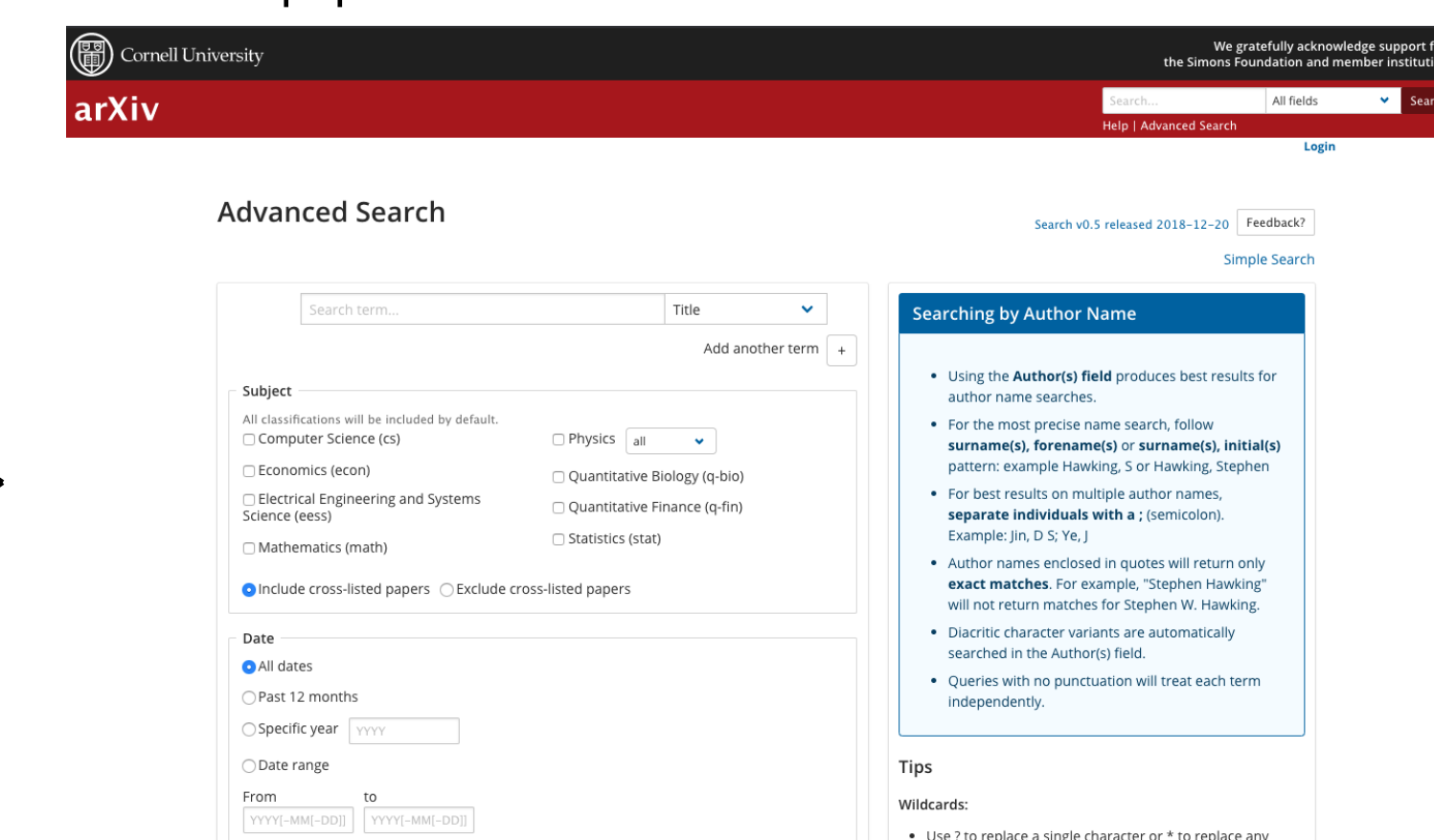
New results

New figures

Likelihoods!

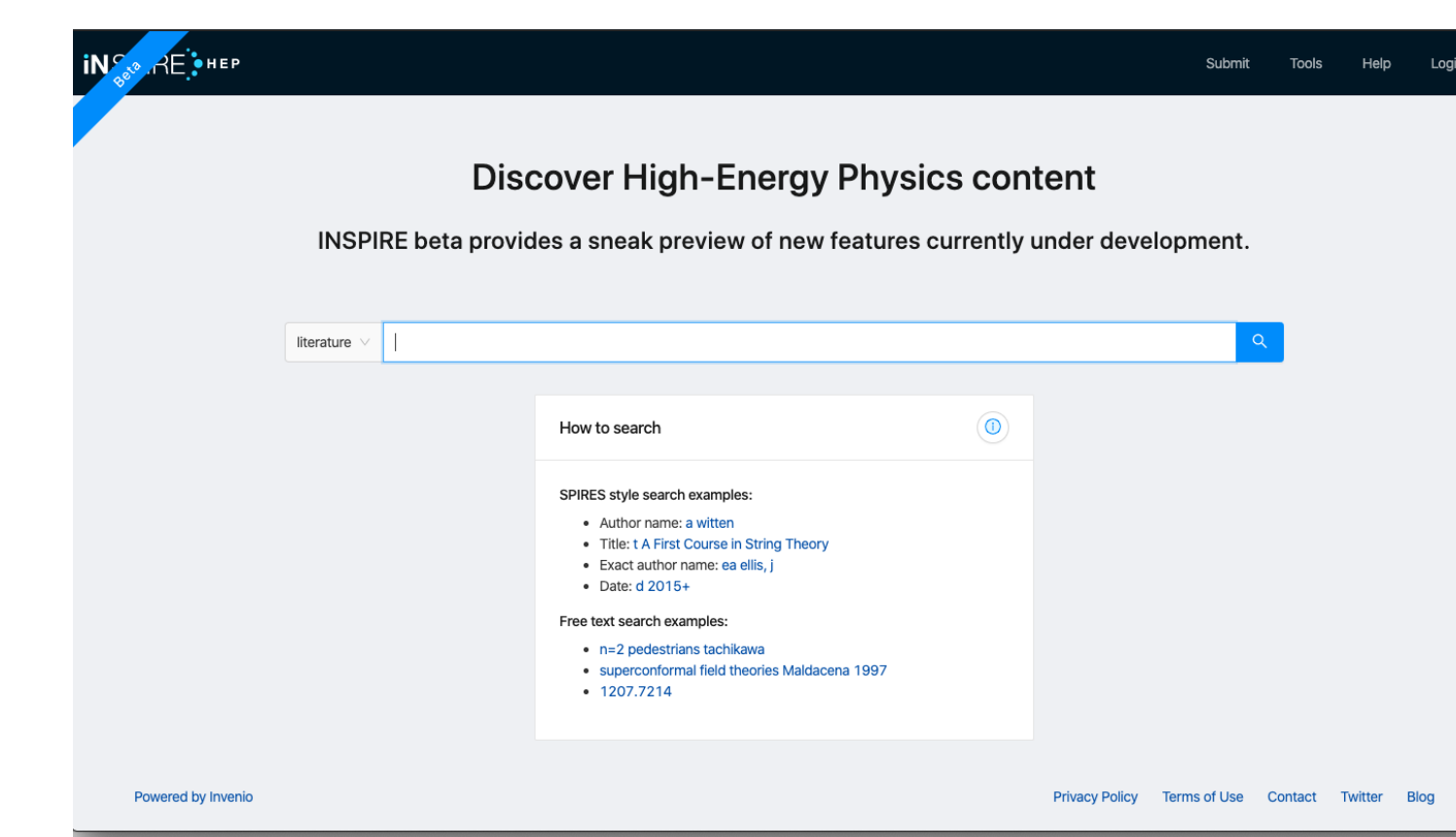
### ArXiv

Stores the paper PDF and text versions.



### INSPIRE

Stores the paper metadata (authors, citations...) serving as hub. It contains links to each paper data spread into other components.



### HEP Data

Stores the paper results and figures.

