

Data and Analysis Reinterpretation

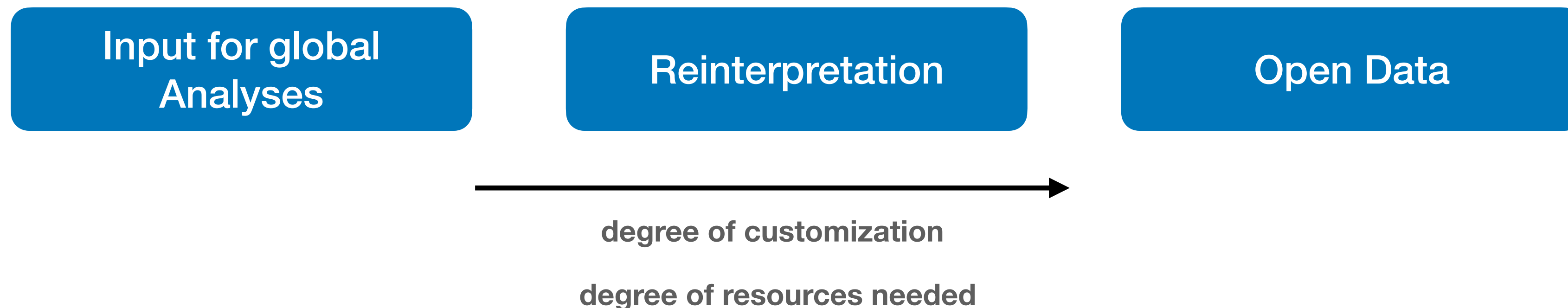
PUNCH4NFDI Open Data Workshop

Lukas Heinrich, CERN

Maximally exploiting LHC analyses

The LHC data is unique and ~fixed. But we can extract new physics results in **three distinct ways**

$$\text{result} = f_{\text{analysis}}(\text{theory}, \text{data})$$




Maximally exploiting LHC analyses

The LHC data is unique and ~fixed. But we can extract new physics results in **three distinct ways**

Inclusion of existing results in global analyses / fits

- **no re-analysis of data required**
- **requires high quality archival of results**

$$\text{result} = f_{\text{analysis}}(\text{theory}, \underset{\text{LHC data}}{\text{data}})$$


$$\text{result}' = \text{combo}(\text{result}, \text{result}_2, \dots, \text{result}_n)$$

Maximally exploiting LHC analyses

The LHC data is unique and ~fixed. But we can extract new physics results in **three distinct ways**

Reinterpreting existing analysis in light of a new theory.

- no new development - reuse event selection procedure
- only re-run new signal - reuse common background estimates
- **requires high-quality archival of analysis pipeline** $f_{\text{analysis}}(\cdot, \text{data})$

$$\boxed{\text{result} = f_{\text{analysis}}(\text{theory}, \underset{\text{LHC data}}{\text{data}})} \longrightarrow \text{result}' = f_{\text{analysis}}(\text{theory}', \text{data})$$

Maximally exploiting LHC analyses

The LHC data is unique and ~fixed. But we can extract new physics results in **three distinct ways**

Open Data:

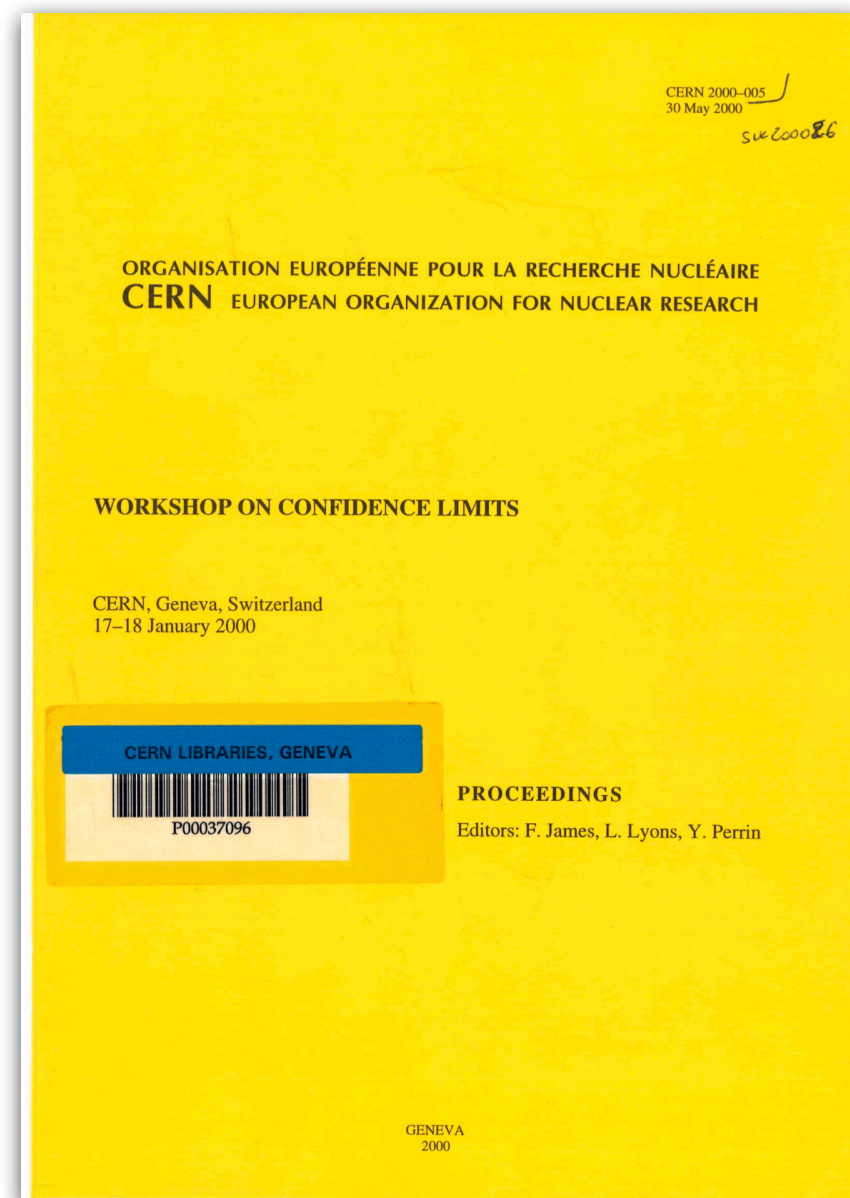
- no constraints from any existing analysis - complete freedom
- but development of pipeline $f'_{\text{analysis}}(\cdot, \text{data})$ is a significant commitment
- requires high-quality archival of
 - data → Open Data
 - software & tools by expt's to analyze it → Open Source Frameworks

$$\boxed{\text{result} = f_{\text{analysis}}(\text{theory}, \underset{\text{LHC data}}{\text{data}})} \longrightarrow \text{result}' = f'_{\text{analysis}}(\text{theory}', \text{data})$$

Open Likelihoods

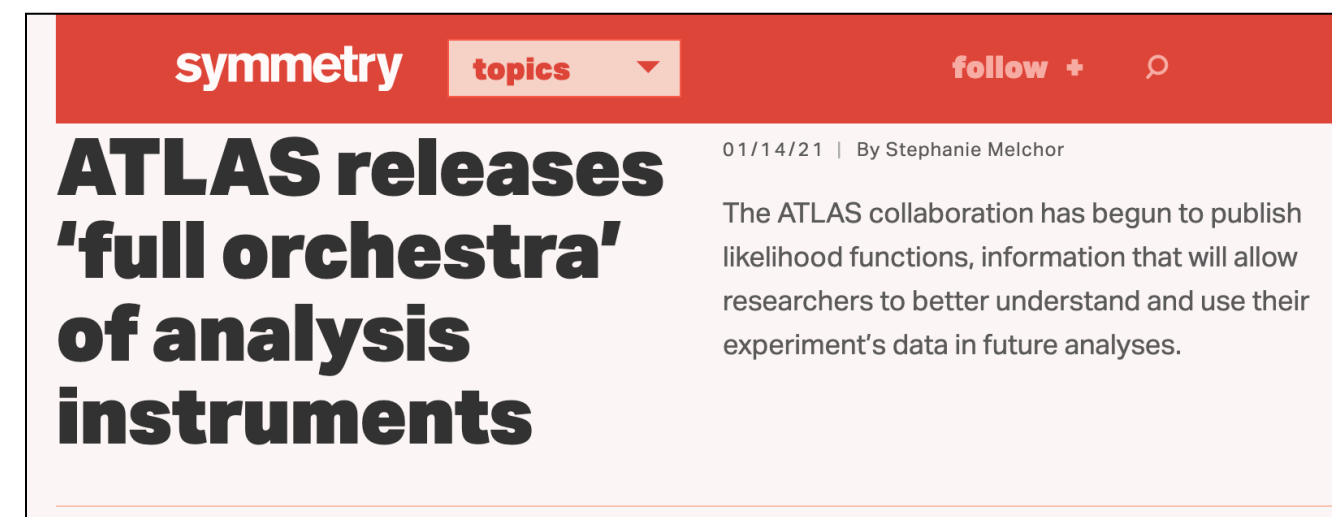
a new "data product" for HEP

2000

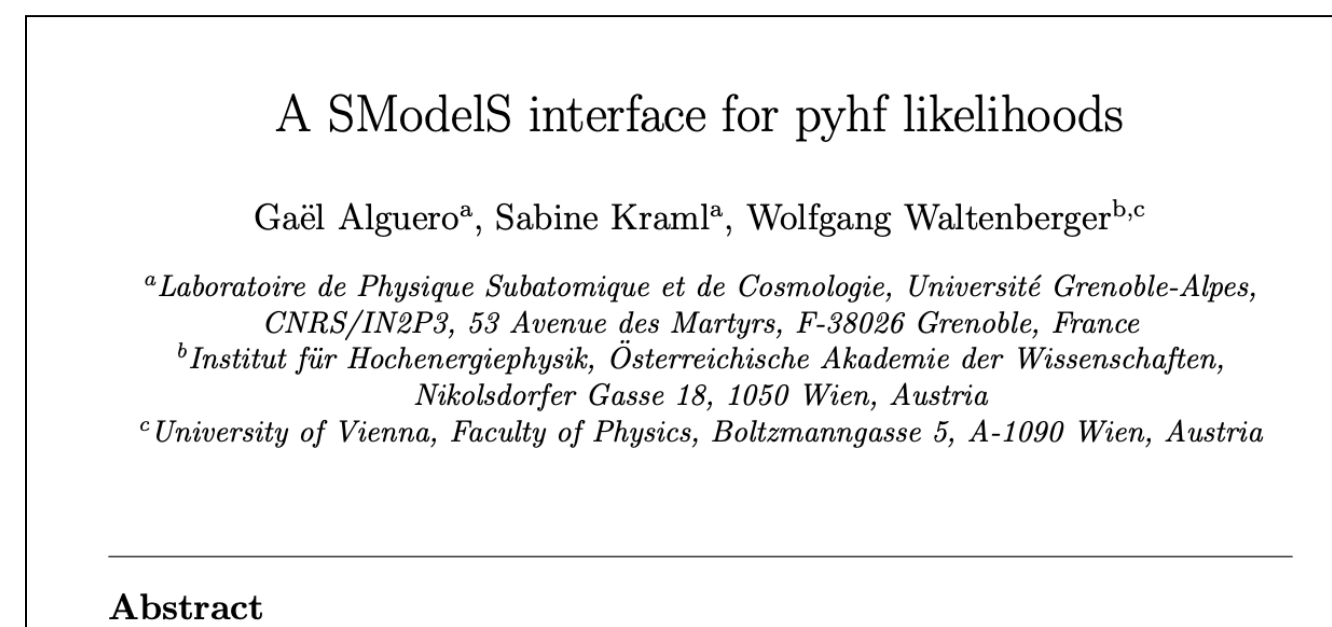


"...everybody would agree that [...] experiments should give their likelihood function"
"Carried unanimously"

2020



ATLAS has begun to publish likelihood functions.



First analyses [...] show how [likelihoods] can help theorists make more trustful reinterpretations.

Likelihood function is the "gold standard" to summarize a given measurement.

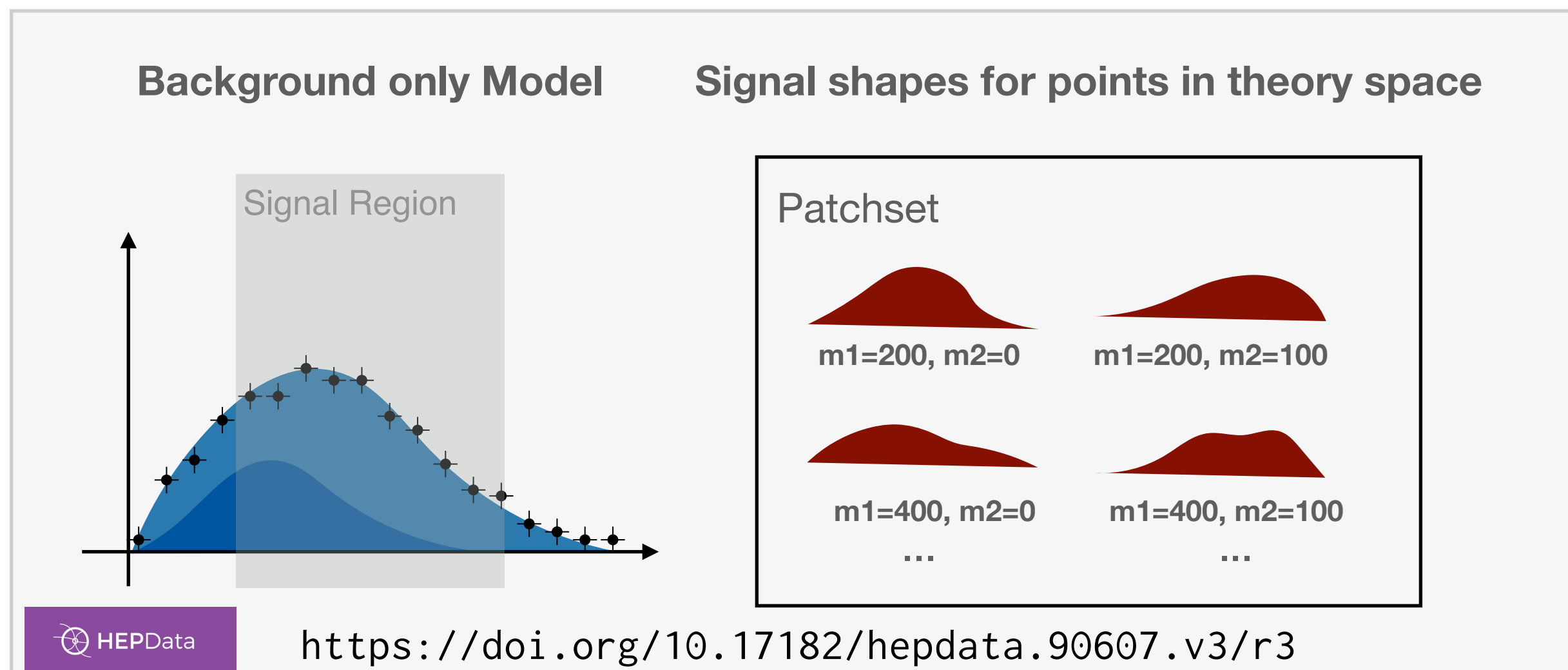
- full set of systematics
- correlations
- reusable (new priors)

also true beyond HEP



After 20yr effort: Open HEP Likelihoods

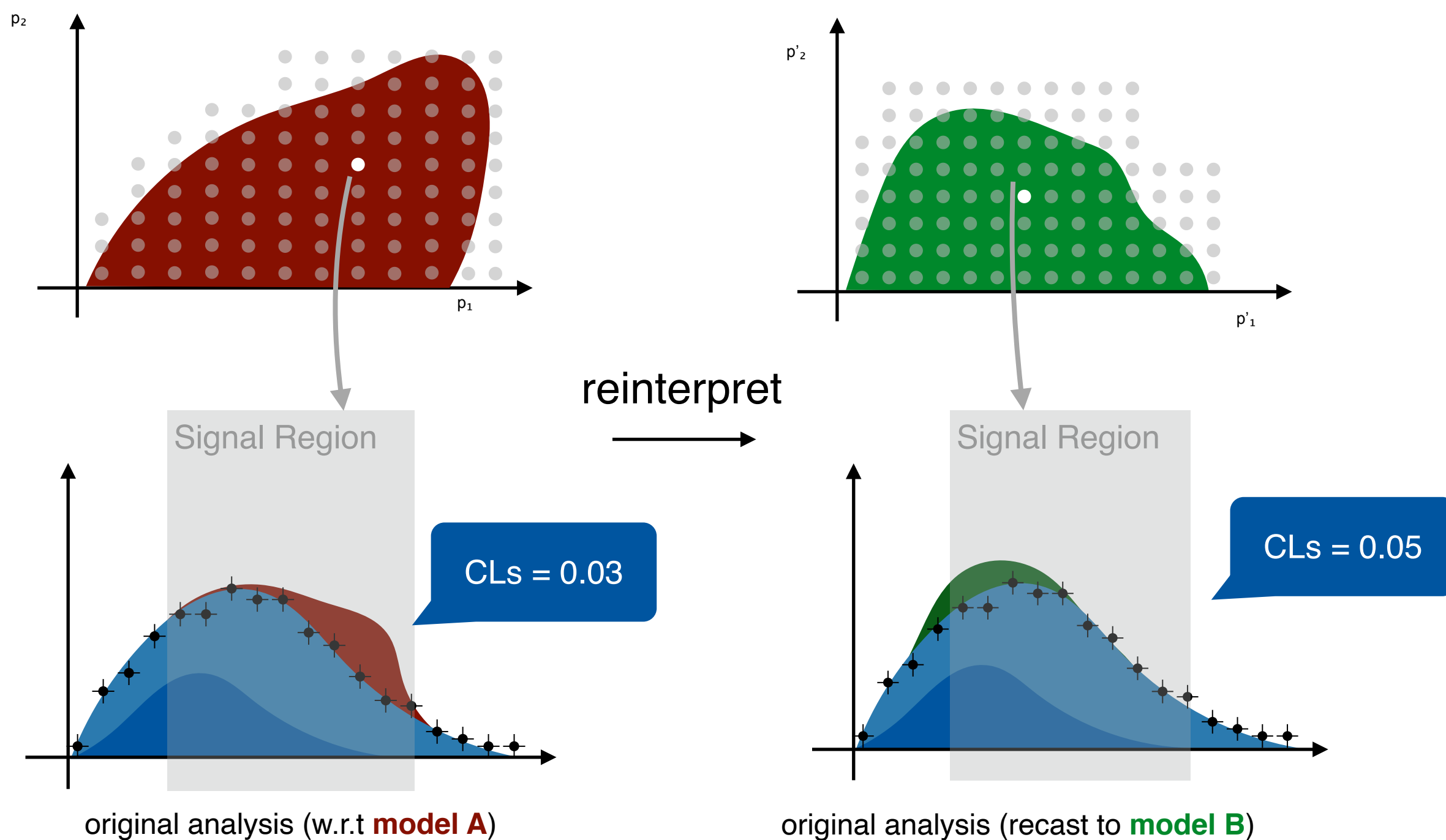
- Technology-agnostic JSON files (**Interoperable** of FAIR)
- Permanent Identifiers (DOIs) (**Findable** of FAIR)
- immediately (re-)used by wider HEP community (**Reusable** of FAIR)



Reinterpretation

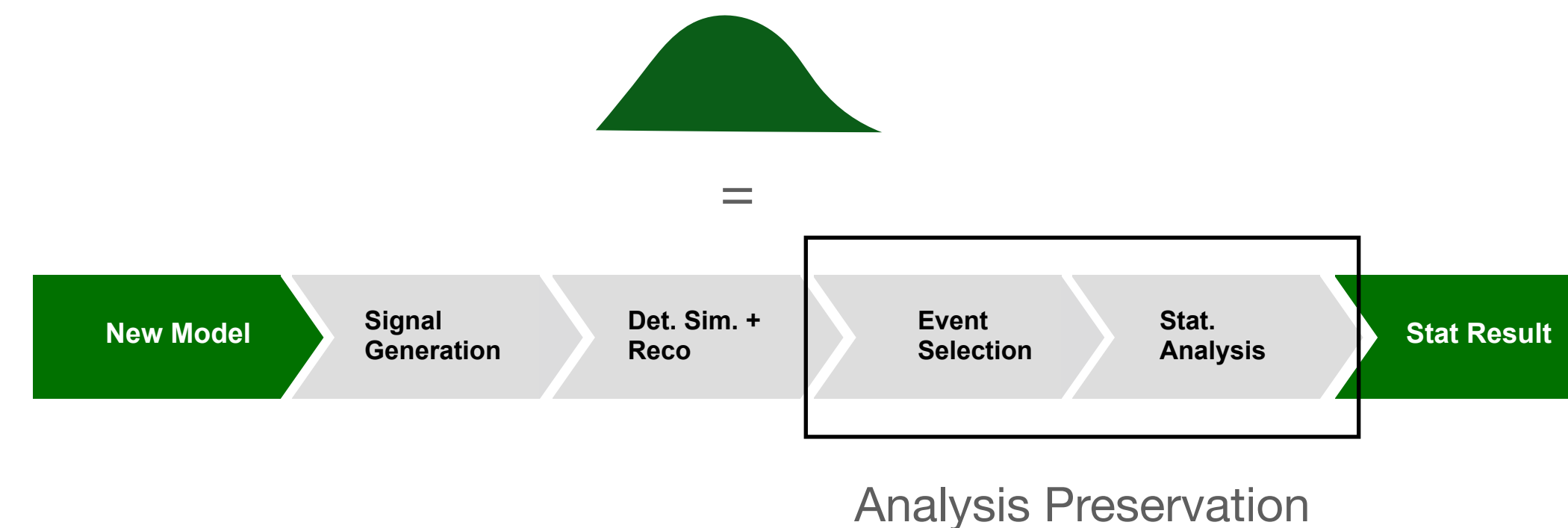
Analysis Likelihood provide base likelihood + results for given set of theories

Can we do more? Yes provide machinery to compute new "patches"



Need to run full analysis pipeline for a new theory.

Crucial: **Analysis Preservation!**



Reinterpretation

LHC Analyses involve many people, many codebases, complex workflows

- need to archive software, job templates, workflow

capture software

archive analysis code incl.
dependencies

capture commands

what do with the
captured software

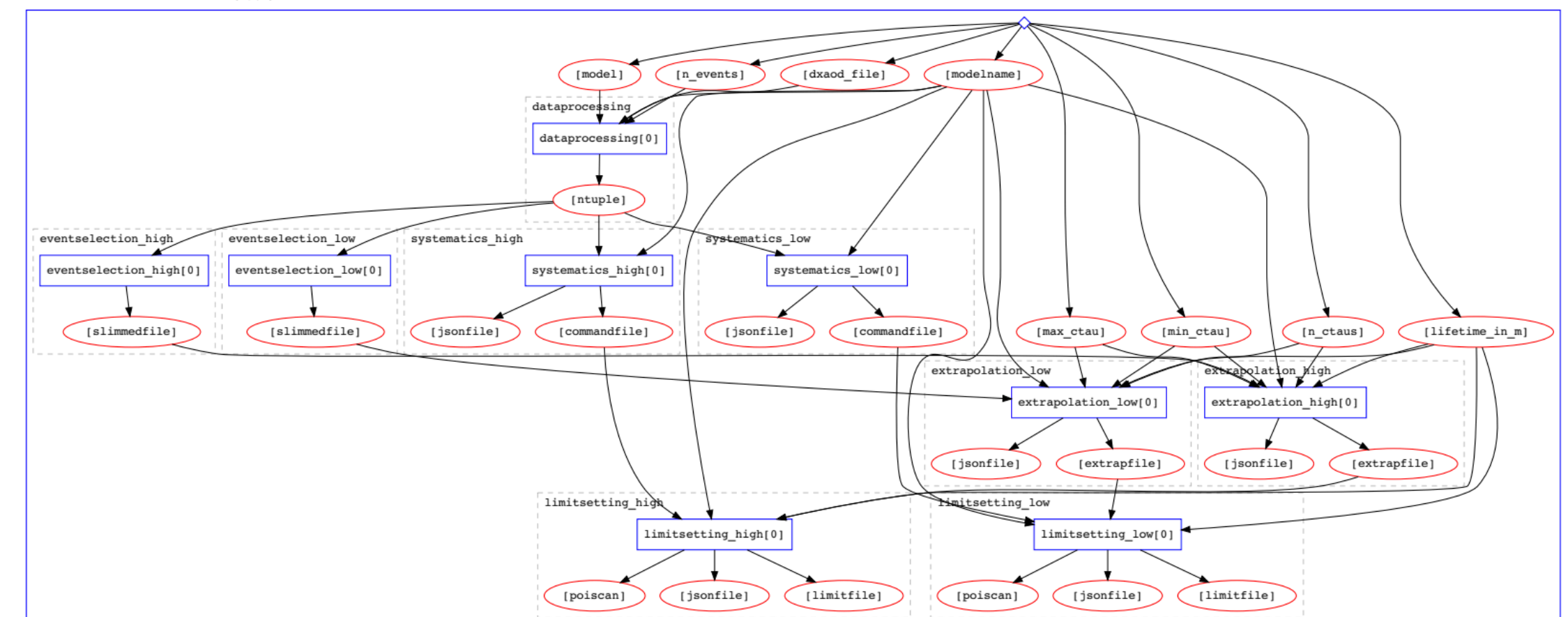
capture workflow

order of individual steps

data assets

input data needed
to run the analysis

- Modern Technology makes this feasible
 - Containers (Docker, ..)
 - Workflow Languages



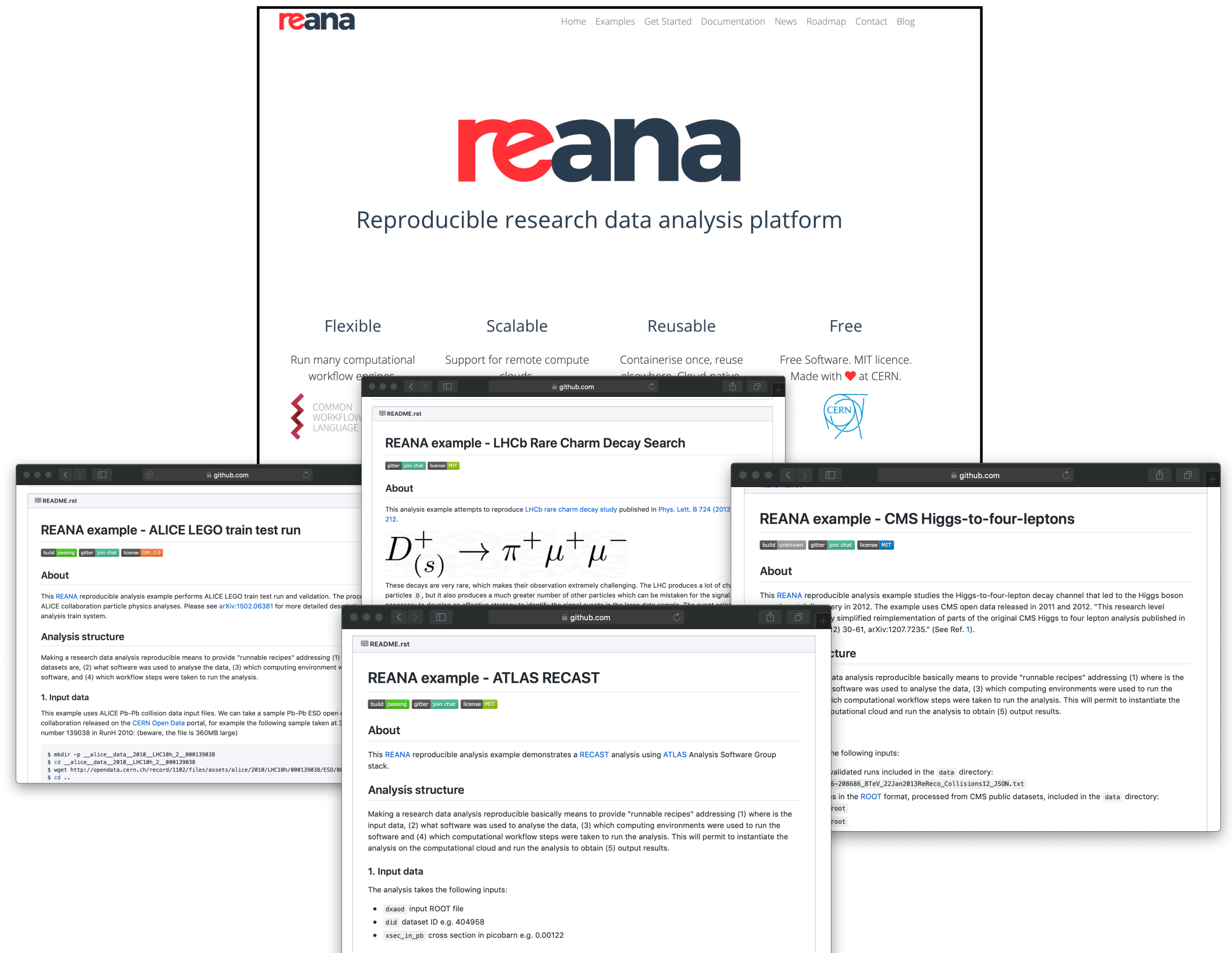
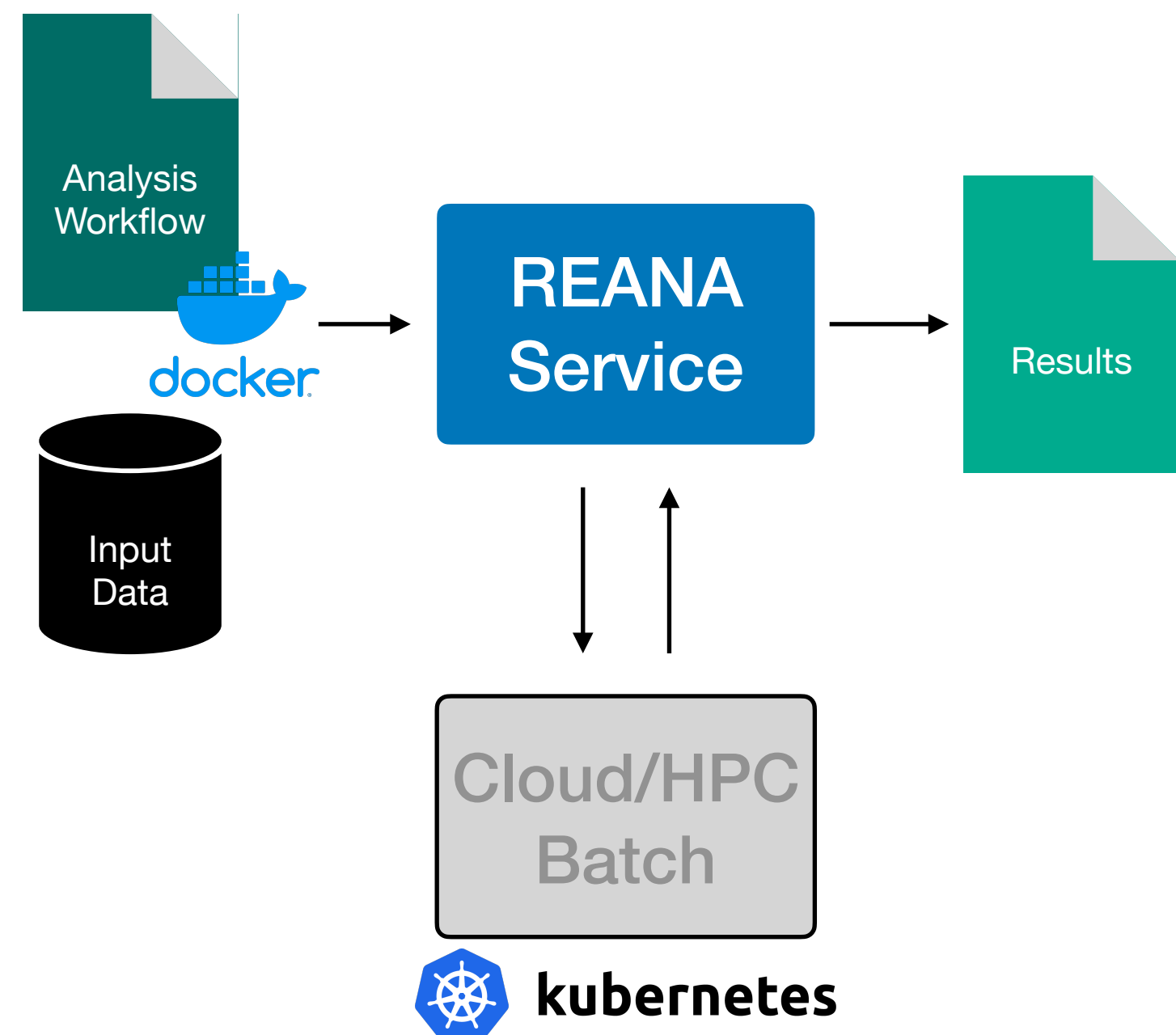
Example Analysis Workflow

Infrastructure for Reinterpretation

Experiments & Labs make preservation easy for analysis teams
Can the be reused on **new inputs**

REANA: workflows-as-a-service

- (Reusable of FAIR)



Infrastructure for Reinterpretation

CERN Analysis Preservation Portal (CAP)

- Archive for e.g. REANA workflows, of individual analyses,...
- links to other databases (**Findable** of FAIR) (INSPIRE, HepData,)

Analysis 1

Analysis Workflow

Internal Note

Preserved Likelihood

software

software

Analysis 2

Analysis Workflow

Internal Note

Preserved Likelihood

software

software

CERN Analysis Preservation

HomeWhat is CAP?Get StartedIntegrationsDocumentationLog in

CERN Analysis Preservation

capture, preserve and reuse physics analyses

CERN Analysis Preservation BETA

Hub

LbZLcDOK branching fraction

Files | Data | Source Code

LbZLcDOK.tar.gz (49 B)

BASIC INFORMATION

ANALYSIS NAME
MEASUREMENT
PROPOSERS
NAME
NAME
ORCID
NAME
NAME
STATUS
INSTITUTES INVOLVED
KEYWORDS
STRIPPING/TURBO SELECTIONS
TYPE OF DATASET
CUSTOM NAME
STRIPPING/TURBO LINE
BOOKKEEPING
LOCATIONS
TYPE OF DATASET
CUSTOM NAME
STRIPPING/TURBO LINE
BOOKKEEPING
LOCATIONS
NTUPLE/USERDST-PRODUCTION
CUSTOM NAME
INPUT DATASET
PLATFORM
DAVINCI VERSION
OUTPUT EOS LOCATION
USER ANALYSIS

LbZLcDOK branching fraction
BR(LbZLcD0barK) and BR(LbZLcD0*barK) with respect to LbZLcDs
Sebastian Neubert
Marian Stahl
0000-0001-8476-8188
Alessio Pucci
Nicola Skidmore
1 - in
16 results
B.O.
Penta
STATUS
draft
data
TYPE
XXLc
XXLc
/LHC
/LHC
PHYSICS_OBJECTS
jet
muon
PFMuon
GlobalMuon
TrackerMuon
electron
photon
MET
tau
track
vertex
2011
data
x56
v465
Jes01

JME-10-004

MUON

FWD-10-005

MUON

AN-2011/103

ELECTRON MUON

AN-2011/062

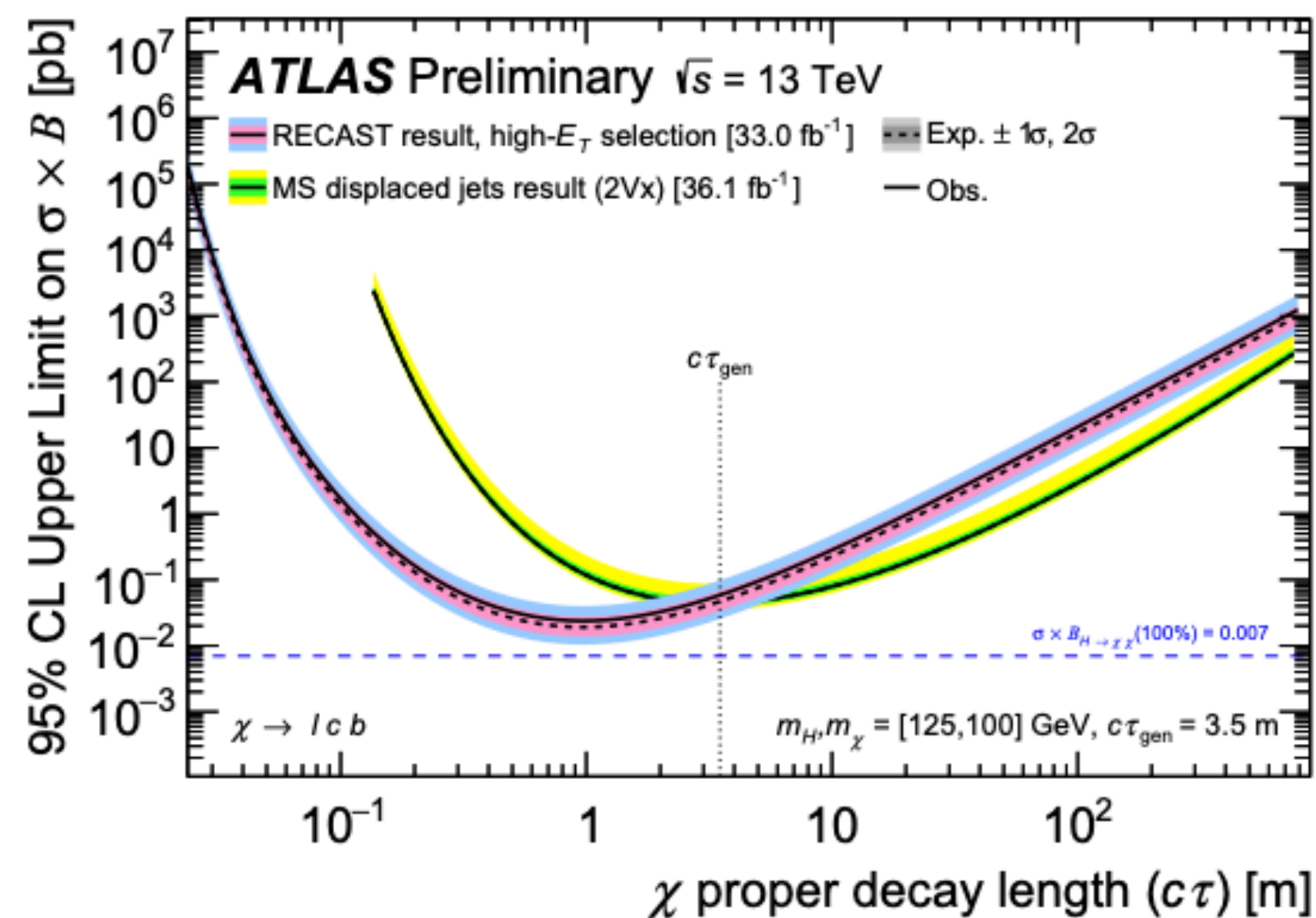
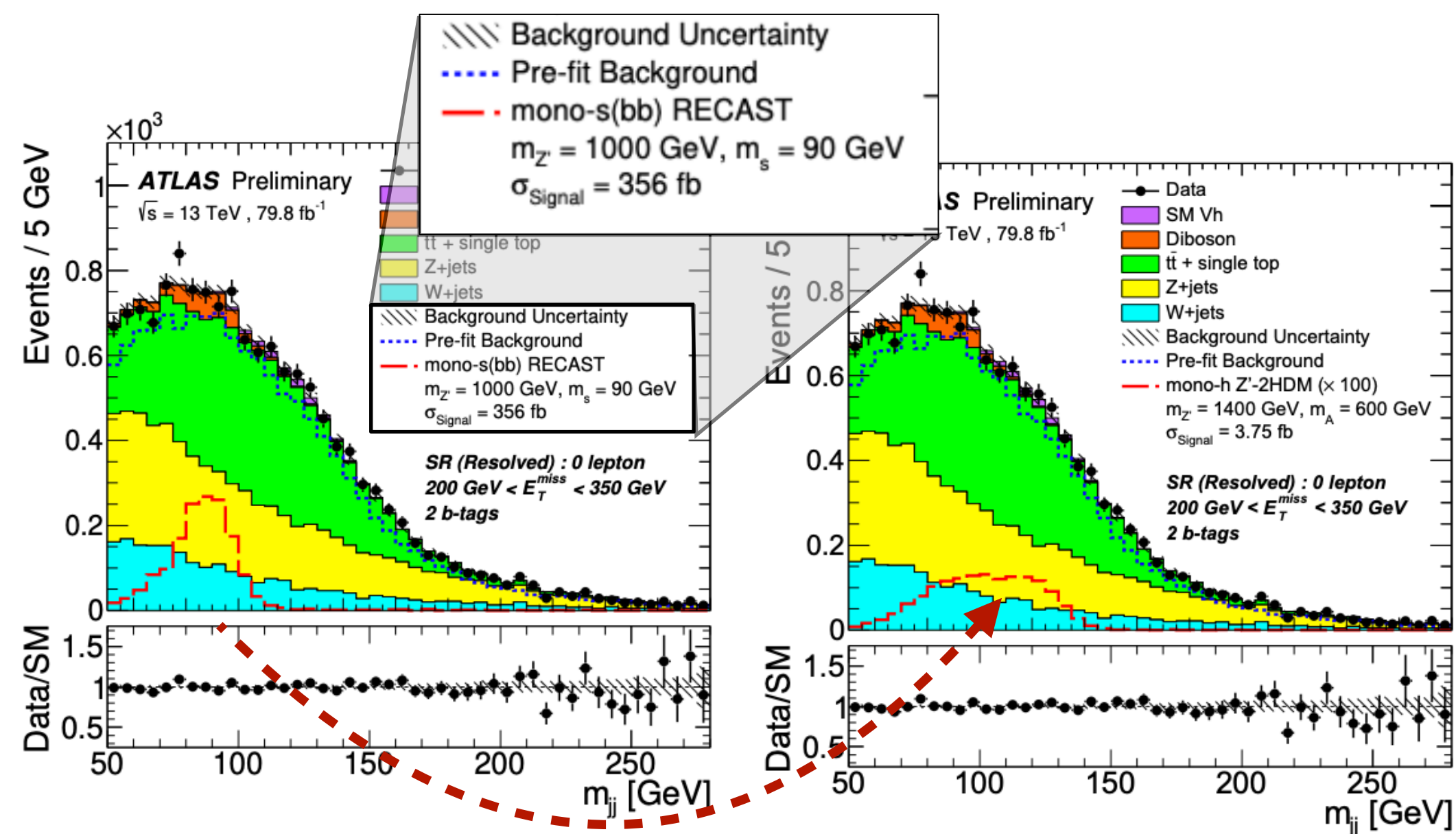
MUON

AN-2011/103

ELECTRON MUON

AN-2010/411

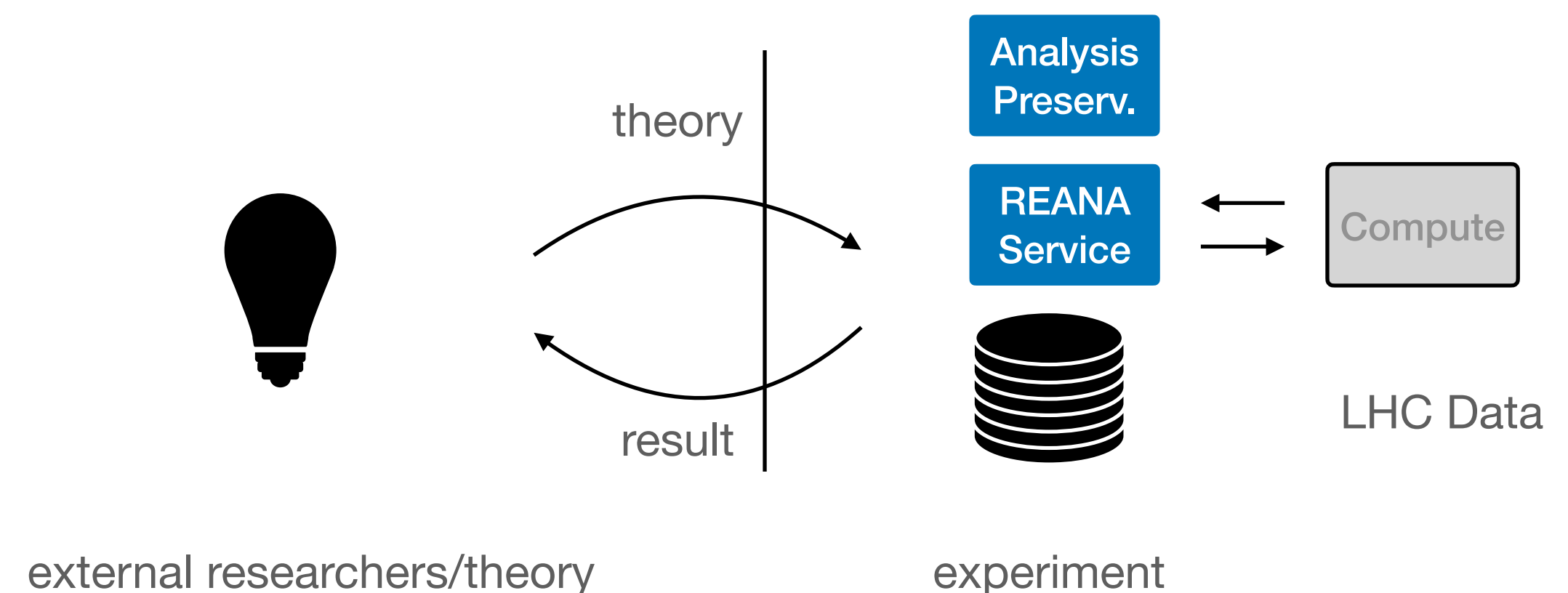
ELECTRON MUON MET



First Fruits of systematic Analysis
Preservation & Reuse effort: RECAST

New Physics results at fraction
of cost compared to "Raw Open Data"

Alternative type of access to data at
higher abstraction level "data analysis"



New perspective: Software Pipeline itself
as a FAIR "data product"

Outlook

LHC data can be made available to external researchers in many ways

- Open and FAIR interpreted in nuanced & different ways

Significant Developments in last years:

- Systematic Analysis Preservation & Reuse (RECAST)
 - support data/research/computing infrastructure in development (REANA, CAP)
- Likelihoods established as new open data products for precision HEP.
- Common Open Data Policy for LHC
- Concepts (scientific workflows, likelihood preservation) applicable beyond HEP
fruitful area for PUNCH4NFDI