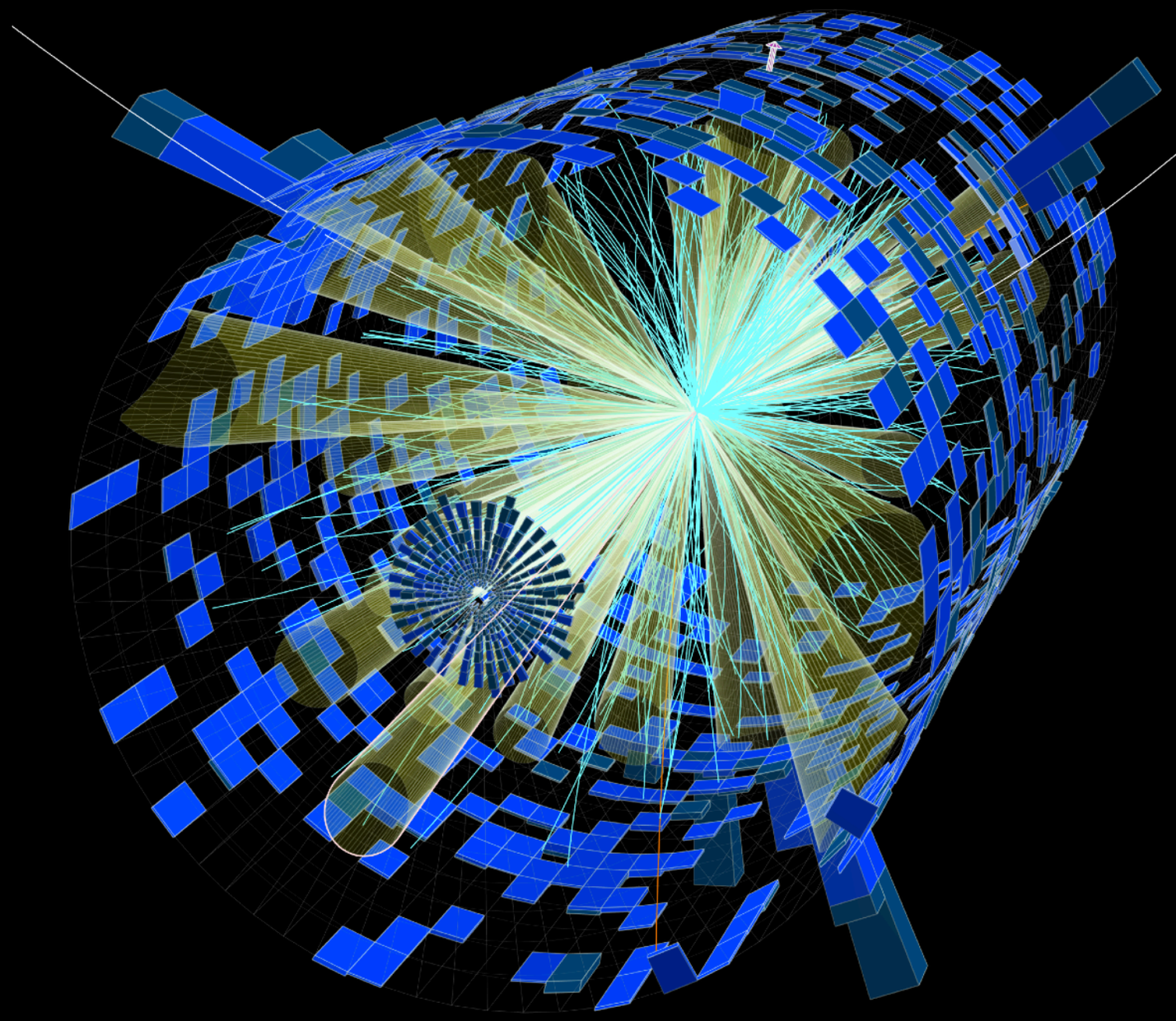
NYU Center for Data Science

Center for Cosmology and particle physics

HEPDATA & IRIS-HEP

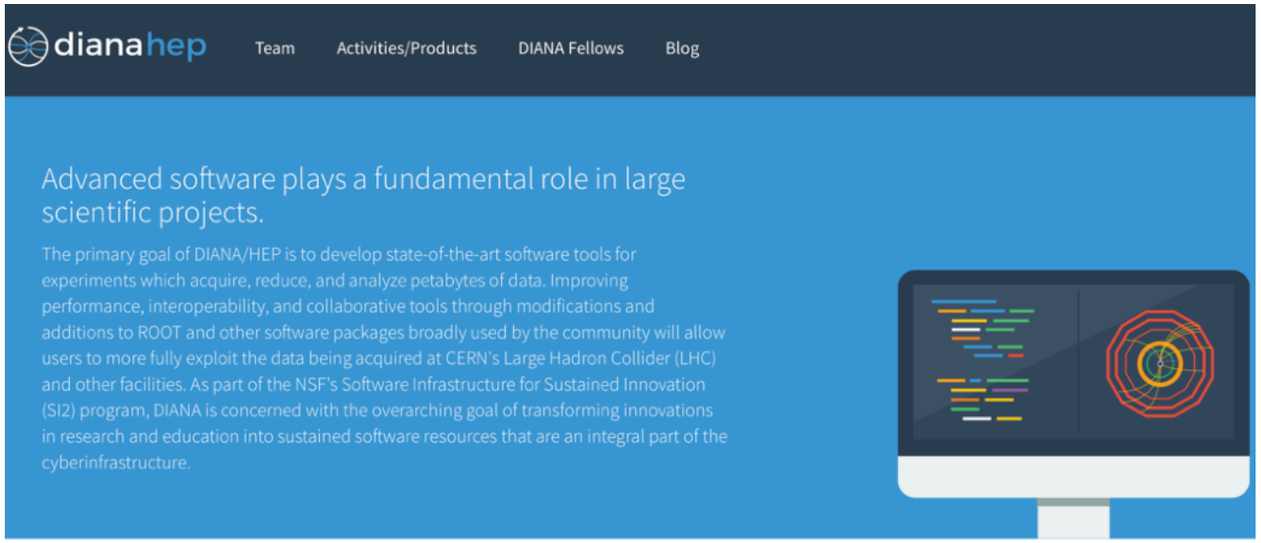


@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

NSF SUPPORTED HEP SOFTWARE EFFORTS

First DIANA-HEP (2014-2019)

Now IRIS-HEP Software Institute (2018-2023)



Collaborative Analyses

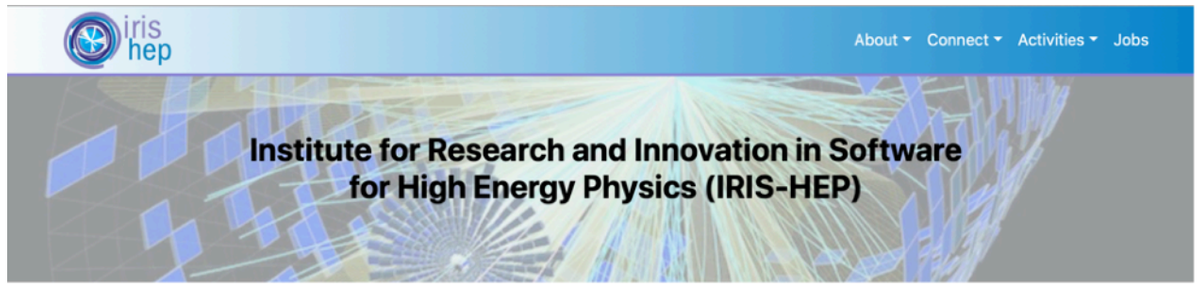
Establish infrastructure for a higher-level of collaborative analysis, building on the successful patterns used for the Higgs boson discovery and enabling a deeper communication between the theoretical community and the experimental community

Reproducible Analyses

Streamline efforts associated to reproducibility, analysis preservation, and data preservation by making these native concepts in the tools

Interoperability

Improve the interoperability of HEP tools with the larger scientific software ecosystem, incorporating best practices and algorithms from other disciplines into HEP



Computational and data science research to enable discoveries in fundamental physics

IRIS-HEP is a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the High Luminosity Large Hadron Collider (HL-LHC) at CERN, and other planned HEP experiments of the 2020's. These facilities are discovery machines which aim to understand the fundamental building blocks of nature and their interactions. Full Overview

Upcoming Events:

Jan 15–17, 2020 *New York University*
ML4Jets2020

Feb 17–19, 2020 *CERN*
Analysis Preservation Bootcamp

Apr 22–24, 2020 *Princeton University*
Connecting the Dots 2020

View all past events

News and Featured Stories:

First USATLAS Bootcamp held in coordination with Software Carpentries and IRIS-HEP/FIRST-HEP

CoDaS-HEP 2019 at Princeton University

For the third consecutive summer, high energy physics graduate students, postdocs and

Upcoming Topical Meetings:

Jan 27, 2020
Allen project

Jan 29, 2020
Primary vertex finding at LHCb

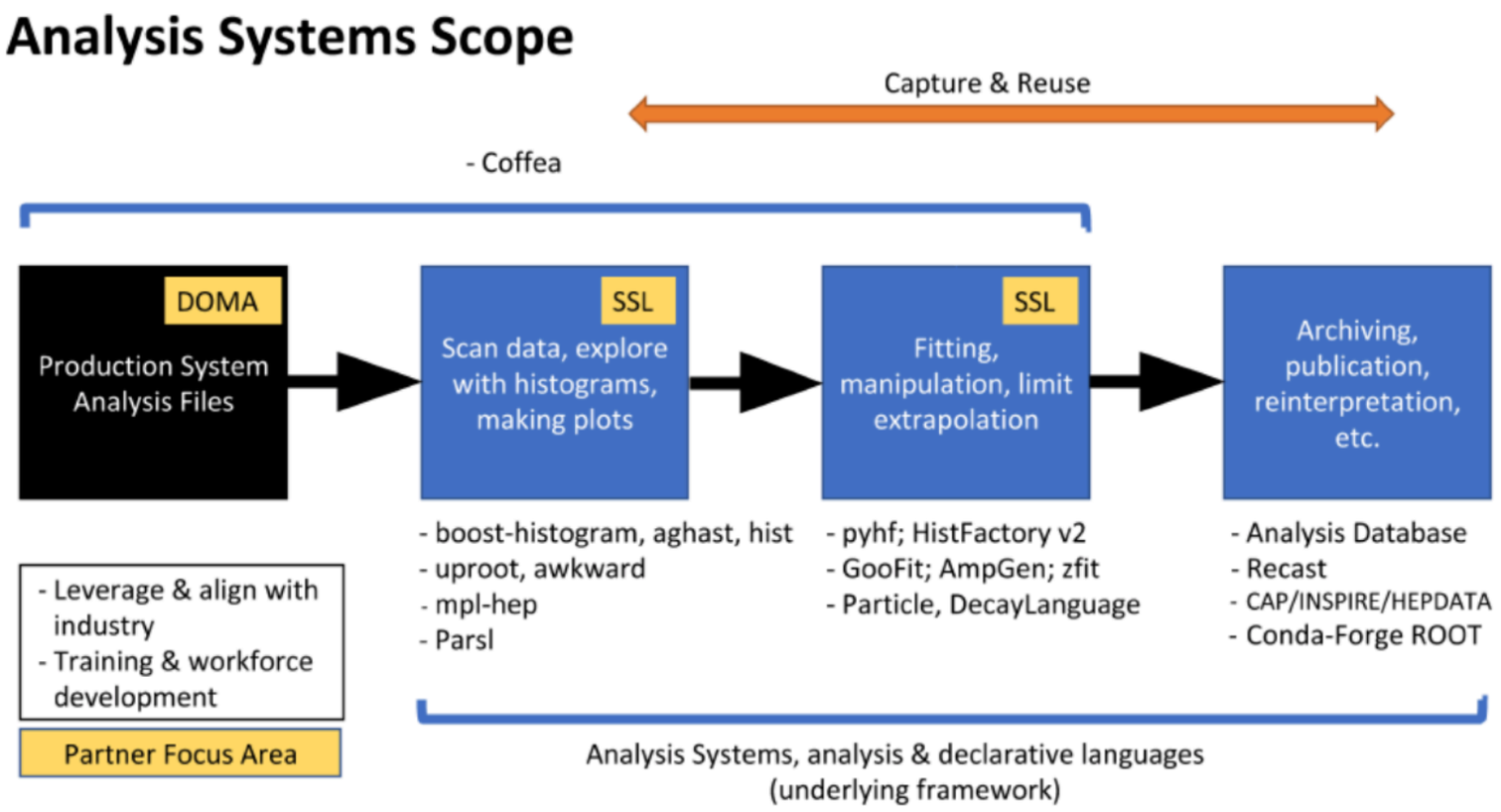
Feb 3, 2020
Modeling computing resource needs/costs

Feb 12, 2020
mkFit and CMS tracking

View all • Indico (recordings) • Vidyo room

Analysis Systems

The goal of the Analysis Systems focus area is to develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments by creating greater functionality, reducing time-to-insight, lowering the barriers for smaller teams, and streamlining analysis preservation, reproducibility, and reuse.



Focus Area Strategies:

- Establish declarative specifications for analysis tasks and workflows that will enable the technical development of analysis systems to be decoupled from the user- facing semantics of physics analysis.
- Leverage and align with developments from industry and the broader scientific software community to enhance sustainability of the analysis systems.
- Develop high-throughput, low-latency systems for analysis for HEP.
- Integrate analysis capture and reuse as first class concepts and capabilities into the analysis systems.

« Prioritizing the future

You Learn Something New Every Day »

KYLE CRANMER I USLHC I USA

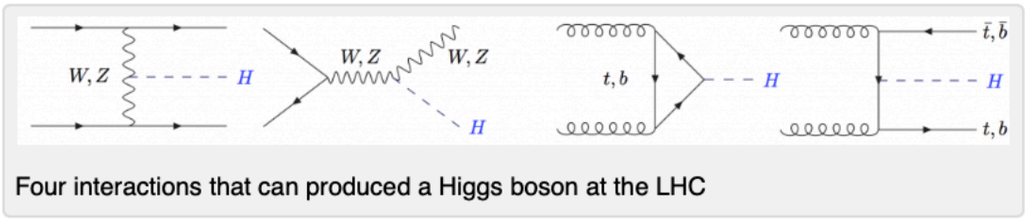
VIEW BLOG I READ BIO

Inspired by the Higgs, a step forward in open access

2013

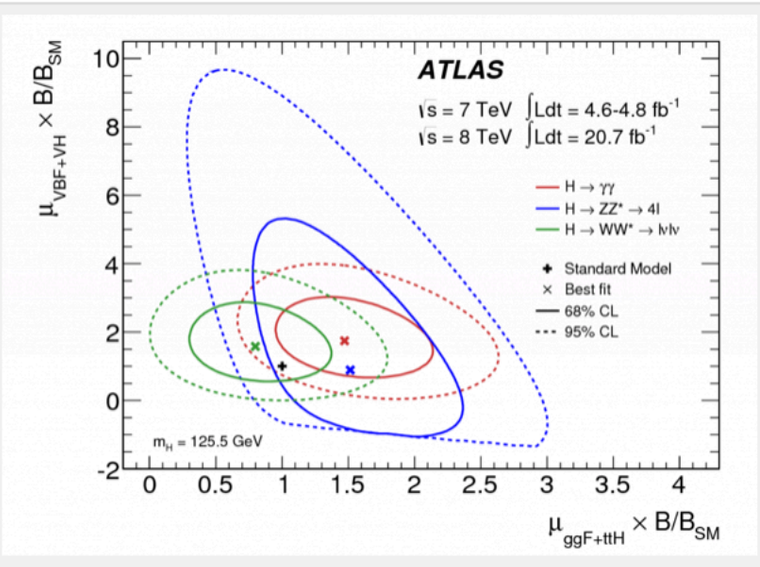
The discovery of the Higgs boson is a major step forward in our understanding of nature at the most fundamental levels. In addition to being the last piece of the standard model, it is also at the core of the fine tuning problem — one of the deepest mysteries in particle physics. So it is only natural that our scientific methodology rise to the occasion to provide the most powerful and complete analysis of this breakthrough discovery.

This week the ATLAS collaboration has taken an important step forward by making the likelihood function for three key measurements about the Higgs available to the world digitally. Furthermore, this data is being shared in a way that represents a template for how particle physics operates in the fast-evolving world of open access to data. These steps are a culmination of decades of work, so allow me to elaborate.



Four interactions that can produced a Higgs boson at the LHC

First of all, what are the three key measurements, and why are they important? The three results were presented by ATLAS in this recent paper. Essentially, they are measurements for how often the Higgs is produced at the LHC through different types of interactions (shown above) and how often it decays into three different force carrying particles (photons, W, and Z bosons). In this plot, the black + sign at (1,1) represents the standard model prediction and the three sets of contours represent the measurements performed by ATLAS. These measurements are fundamental tests of the standard model and any deviation could be a sign of new physics like supersymmetry!



Higgs production and decay measured by ATLAS.

AAHEP7 Information Provider Summit

1-3 April 2014
Stony Brook University
US/Eastern timezone

INVENIO

Common Data Analysis Preservation Platform for LHC Experiments?

Tibor Šimko & Sünje Dallmeier-Tiessen
CERN

AAHEP7
Stony Brook University, New York
2 April 2014

The Durham HepData Project

Reaction Database · Data Reviews · PDF Plotter — About HepData · Submitting Data

HepData status+data

Contact us at: hepdata@projects.hepforge.org

HepData is funded by the UK STFC and hosted at the Durham IPPP.
HepData also maintains the UK mirror of the PDG.

Mike Whalley
IPPP Durham U., UK

AAHEP7 Information Provider Summit
Stony Brook - 2nd/3rd April 2014

INVENIO » ATLAS Data: Citations

INSPIRE

Welcome to INSPIRE, the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

Information | Citations (3) | Files

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

ATLAS Collaboration (Aad, Georges (Freiburg U.) [...]) *Show all 2923 authors*

Cite as: ATLAS Collaboration (2013) HepData, http://doi.org/10.7484/INSPIREHEP.DATA.RF5P.6M3K

Description: -2 log Likelihood for the H→ ZZ*→ 4l channel in the (μ_ggF+ttH * B/BSM, μ_VBF+VH * B/BSM) plane for a Higgs boson mass mH = 125.5 GeV.

Preview not available

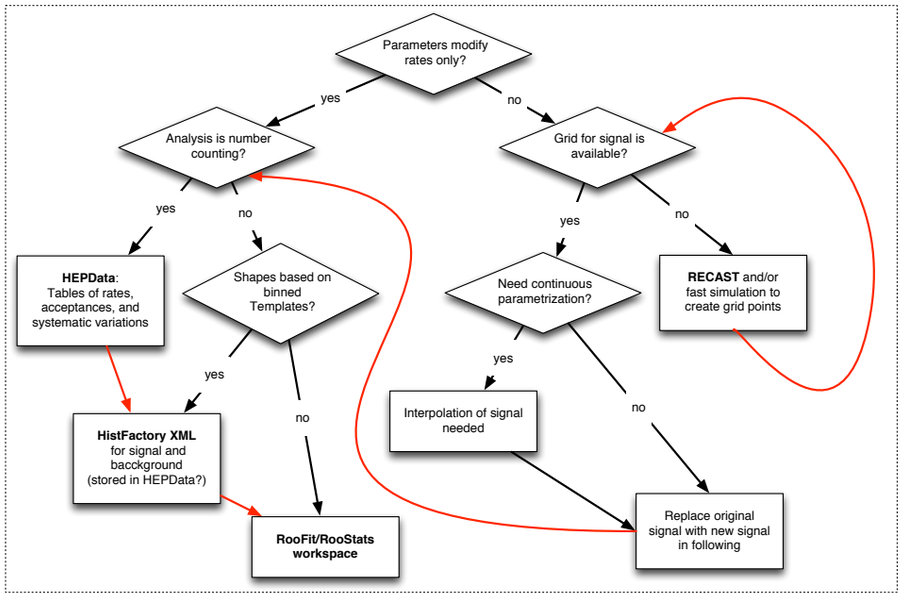
Note: * Temporary entry *

This dataset complements the following publication:
Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

Record created 2013-09-11, last modified 2013-12-16

Export
BibTeX, EndNote, LaTeX(US), LaTeX(EU), Harvmac, MARC, MARCXML, NLM, DC

ROADMAP



ATL-CB-PUB-2015-001
17 March 2015

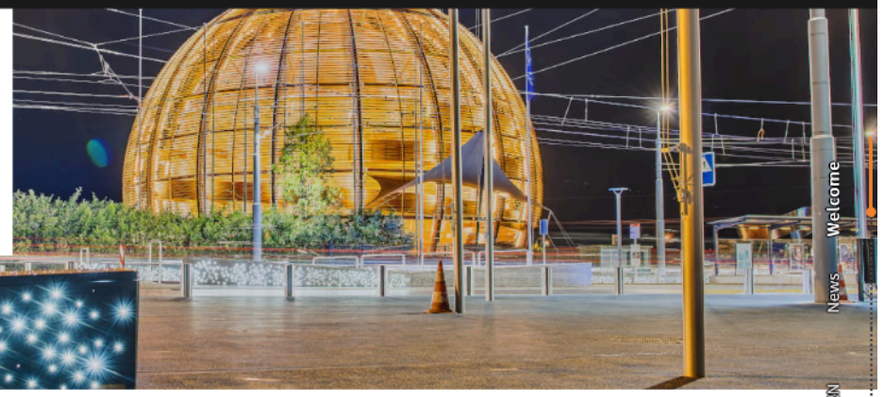
Level-1. Published results

All scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN.

Data associated with journal publications are also made available: tables and data from plots (e.g. cross section values, likelihood profiles, selection efficiencies, cross section limits, …) are stored in appropriate repositories such as HEPDATA[2]. ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET [3]. For searches information on signal acceptances is also made available to allow reinterpretation of these searches in the context of models developed by theorists after the publication. ATLAS is also exploring how to provide the capability for reinterpretation of searches in the future via a service such as RECAST [4]. RECAST allows theorists to evaluate the sensitivity of a published analysis to a new model they have developed by submitting their model to ATLAS.

CERN Council appoints Fabiola Gianotti for second term of office as CERN Director General

Press release | 6 November, 2019



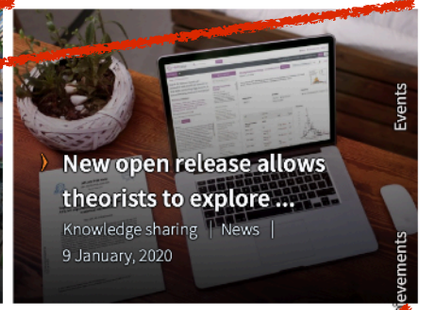
LATEST NEWS

Particle physicists formulate the f...

ring | News |
20

LHCb explores the beauty of lepton universali...

Physics | News | 15 January, 2020



New open release allows theorists to explore ...

Knowledge sharing | News |
9 January, 2020

View all news

New open release allows theorists to explore LHC data in a new way

The ATLAS collaboration releases full analysis likelihoods, a first for an LHC experiment

9 JANUARY, 2020 | By Katarina Anthony

Explore ATLAS open likelihoods on the HEPData platform (Image: CERN)

What if you could test a new theory against LHC data? Better yet, what if the expert knowledge needed to do this was captured in a convenient format? This tall order is now on y from the ATLAS collaboration, with the first open release of full analysis likelihoods

Related Articles

Open Data: pushing back the frontiers togethe...

Knowledge sharing | News | 22 October, 2018

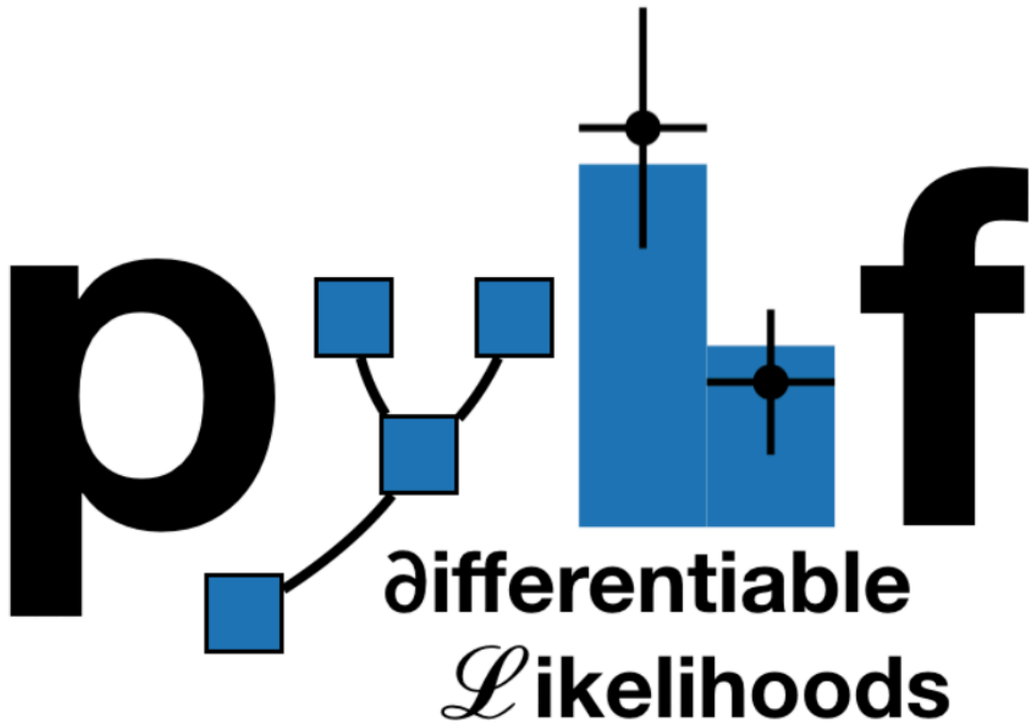
The invisible structure providing Open Acces...

Knowledge sharing | News | 30 July, 2018

Sit down for coffee with the Standard Model

At CERN | News |
7 April, 2017

View all news



pyhf is a pure-python implementation of the widely-used HistFactory p.d.f. template described in [CERN-OPEN-2012-016]. It also includes interval estimation is based on the asymptotic formulas of "Asymptotic formulae for likelihood-based tests of new physics" [arxiv:1007.1727]. The aim is also to support modern computational graph libraries such as PyTorch and TensorFlow in order to make use of features such as autodifferentiation and GPU acceleration.

A talk on pyhf by Matthew Feickert

Team

- Kyle Cranmer
- Lukas Heinrich
- Matthew Feickert
- Giordon Stark

nature physics

PERSPECTIVE

https://doi.org/10.1038/s41567-018-0342-2

Corrected: Publisher Correction

OPEN

Open is not enough

Xiaoli Chen[1,2], Sünje Dallmeier-Tiessen[1]*, Robin Dasler[1,11], Sebastian Feger[1,3], Pamfilos Fokianos[1], Jose Benito Gonzalez[1], Harri Hirvonsalo[1,4,12], Dinos Kousidis[1], Artemis Lavasa[1], Salvatore Mele[1], Diego Rodriguez Rodriguez[1], Tibor Šimko[1]*, Tim Smith[1], Ana Trisovic[1,5]*, Anna Trzcinska[1], Ioannis Tsanaktsidis[1], Markus Zimmermann[1], Kyle Cranmer[6], Lukas Heinrich[6], Gordon Watts[7], Michael Hildreth[8], Lara Lloret Iglesias[9], Kati Lassila-Perini[4] and Sebastian Neubert[10]

The solutions adopted by the high-energy physics community to foster reproducible research are examples of best practices that could be embraced more widely. This first experience suggests that reproducibility requires going beyond openness.

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

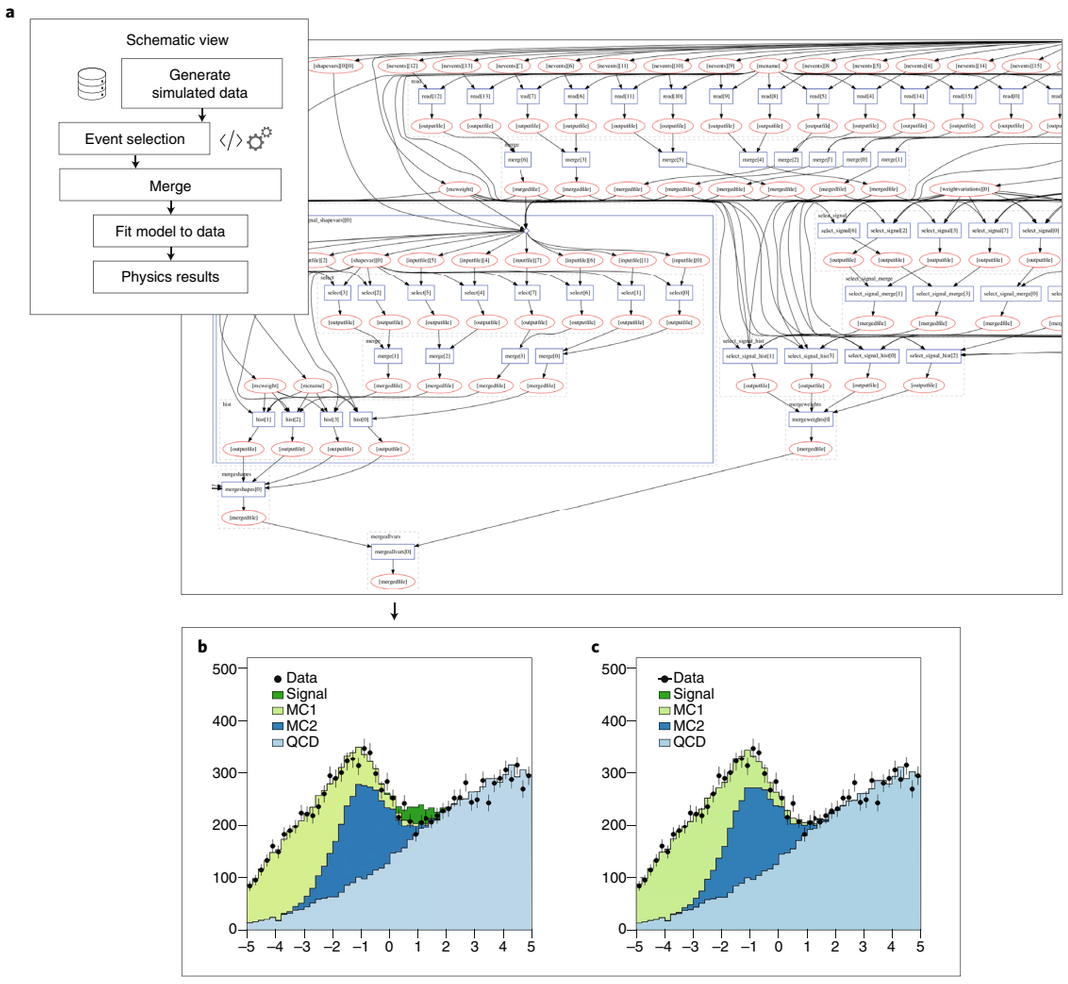
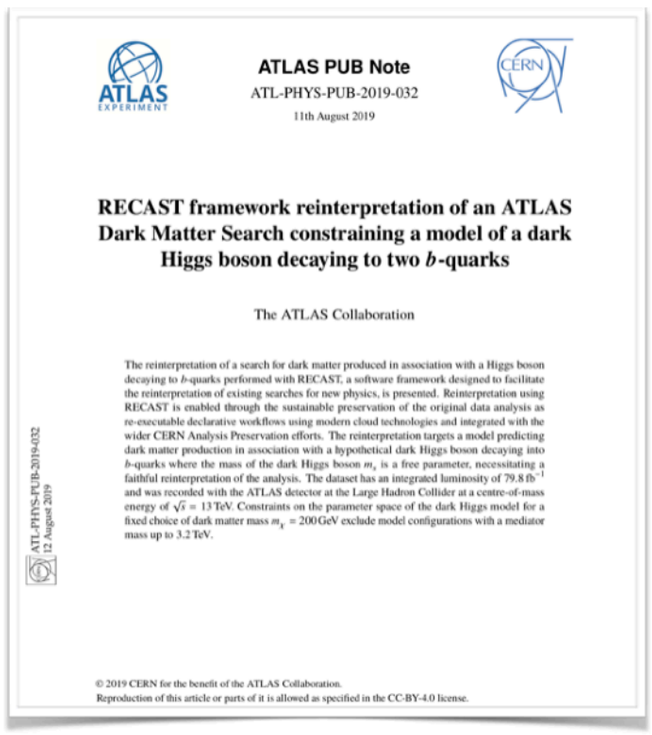


Fig. 2 | Example of a complex computational workflow on REANA mimicking a beyond the standard model (BSM) analysis . This figure shows an example where the experimental data is compared to the predictions of the standard model with an additional hypothesized signal component. The example permits one to study the complex computational workflows used in typical particle physics analyses. **a**–**c**, The computational workflow (**a**) may consist of several tens of thousands of computational steps that are massively parallelizable and run in a cascading 'map-reduce' style of computations on distributed compute clusters. The workflow definition is modelled using the Yadage workflow specification and produces an upper limit on the signal strength of the BSM process. A typical search for BSM physics consists of simulating a hypothetical signal process (**c**), as well as the background processes predicted by the standard model with properties consistent with the hypothetical signal (marked dark green in (**b**)). The background often consists of simulated background estimates (dark blue and light green histograms) and data-driven background estimates (light blue histogram). A statistical model involving both signal (dark green histogram) and background components is built and fit to the observed experimental data (black markers). **b**, Results of the model in its pre-fit configuration at nominal signal strength. We can see the excess of the signal over data, meaning that the nominal setting does not describe the data well. The post-fit distribution would scale down the signal in order to fit the data. This REANA example is publicly available at ref. [35]. For icon credits, see Fig. 1.

recast

RECAST is a framework for extending the impact of existing analyses performed by high-energy physics experiments.

- *Request*: Upload alternative signals in the LHE format and request that any given analysis is "recast" for an alternative model. Note: this is a request, there is no obligation for the experiments to respond.
- *Subscribe*: Anyone can subscribe to an analysis to be informed of activity associated with the analysis
- *Recast*: Experimentalists can accept the request, process these alternative signals with the full simulation, reconstruction, and analysis selection. If authorized by their collaboration, they can respond with an authoritative result for the selection efficiency and cross-section limits for the alternative signal. Note, anyone can provide a non-authoritative result, for instance one based on a phenomenological recasting tool.



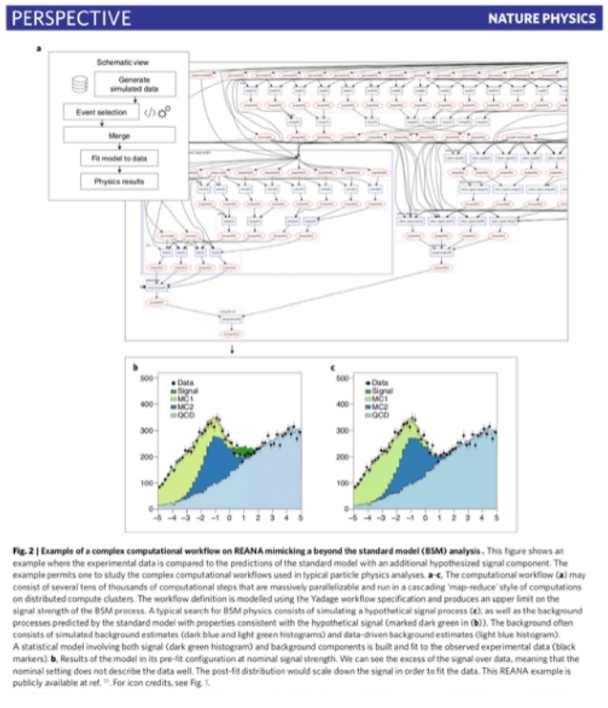
ATLAS PUB Note
ATL-PHYS-PUB-2019-032
11th August 2019

RECAST framework reinterpretation of an ATLAS Dark Matter Search constraining a model of a dark Higgs boson decaying to two *b*-quarks

The ATLAS Collaboration

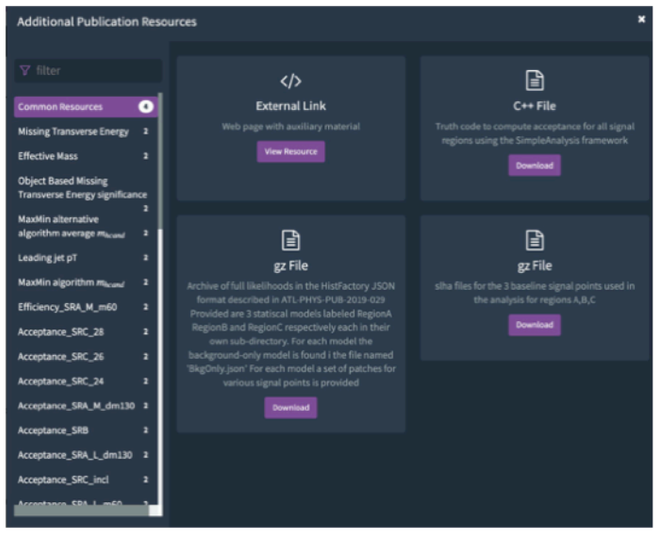
The reinterpretation of a search for dark matter produced in association with a Higgs boson decaying to *b*-quarks performed with RECAST, a software framework designed to facilitate the reinterpretation of existing searches for new physics, is presented. Reinterpretation using RECAST is enabled through the sustainable preservation of the original data analysis as re-executable declarative workflows using modern cloud technologies and integrated with the wider CERN Analysis Preservation efforts. The reinterpretation targets a model predicting dark matter production in association with a hypothetical dark Higgs boson decaying into *b*-quarks where the mass of the dark Higgs boson m_s is a free parameter, necessitating a faithful reinterpretation of the analysis. The dataset has an integrated luminosity of 79.8 fb^{-1} and was recorded with the ATLAS detector at the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Constraints on the parameter space of the dark Higgs model for a fixed choice of dark matter mass $m_\chi = 200$ GeV exclude model configurations with a mediator mass up to 3.2 TeV.

ATL-PHYS-PUB-2019-032
12 August 2019



PERSPECTIVE NATURE PHYSICS

Fig. 2 | Example of a complex computational workflow on REANA mimicking a beyond the standard model (BSM) analysis . This figure shows an example where the experimental data is compared to the predictions of the standard model with an additional hypothesized signal component. The example permits one to study the complex computational workflows used in typical particle physics analyses. **a**–**c**, The computational workflow (**a**) may consist of several tens of thousands of computational steps that are massively parallelizable and run in a cascading 'map-reduce' style of computations on distributed compute clusters. The workflow definition is modelled using the Yadage workflow specification and produces an upper limit on the signal strength of the BSM process. A typical search for BSM physics consists of simulating a hypothetical signal process (**c**), as well as the background processes predicted by the standard model with properties consistent with the hypothetical signal (marked dark green in (**b**)). The background often consists of simulated background estimates (dark blue and light green histograms) and data-driven background estimates (light blue histogram). A statistical model involving both signal (dark green histogram) and background components is built and fit to the observed experimental data (black markers). **b**, Results of the model in its pre-fit configuration at nominal signal strength. We can see the excess of the signal over data, meaning that the nominal setting does not describe the data well. The post-fit distribution would scale down the signal in order to fit the data. This REANA example is publicly available at ref. [31]. For icon credits, see Fig. 1.



Website GitHub Paper

Sinclert Pérez joined NYU

- Research Software Engineer
- Previously worked at CERN team on REANA
- 50% IRIS-HEP

Sinclert Pérez
New York University

Research Software Engineer

Focus is to improve interoperability of cyberinfrastructure components (eg. RECAST, CAP, HEPData, INSPIRE, …)

VISUALIZATION

https://www.youtube.com/watch?time_continue=5&v=9egt9ZTm7T0&feature=emb_title

pyhf likelihoods

Kyle Cranmer
Lukas Heinrich
Matthew Feickert
Giordon Stark

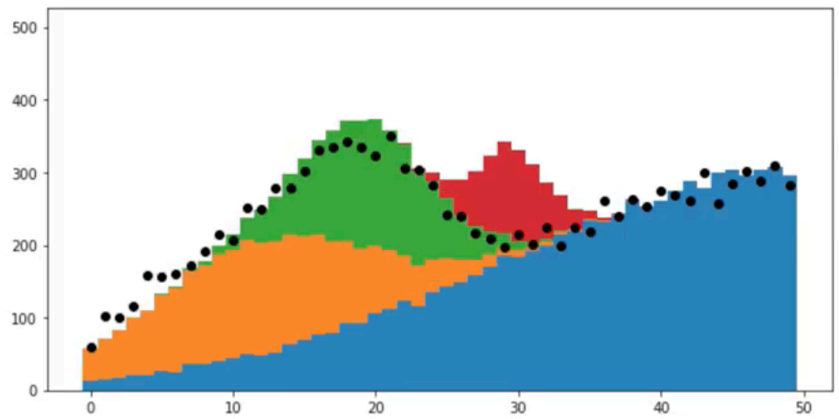
- Recent Python implementation (pyhf) of HistFactory.
- ROOT/XML workspace replaced by plain-text pyhf JSON.
- Two HEPData records released by ATLAS with pyhf JSON:
  https://www.hepdata.net/record/ins1748602 (2019-10-21)
  https://www.hepdata.net/record/ins1765529 (2019-12-06)
- Need to mint DOIs for local resource files and identify pyhf.
- **Native** support of pyhf JSON (see HEPData/hepdata#164):
  - ✦ Replace usual HEPData YAML data files with pyhf JSON.
  - ✦ Validate against JSON schema distributed with pyhf.
  - ✦ Develop appropriate *visualisation* and *conversion* tools.

Interactive Exploration of a HistFactory Model

One advantage of a pure-python implementation of Histfactory is the ability to explore the pdf interactively within the setting of a notebook. Try moving the sliders and oberserve the effect on the samples. For example changing the parameter of interest *SigXsecOverSM* (or μ) controls the overall normalization of the (BSM) signal sample (μ=0 for background-only and μ=1 for the nominal signal-plus-background hypothesis)

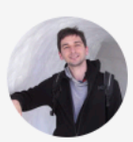
```
interact(plot, order = fixed([1,2,3,0]), ax = fixed(None), **{n[0]: tuple(m) for n,m in zip(sorted(reversed(list(parnam
```

SigXsecOv... 4
mc1_weigh... 0
mc1_shap... 1
mc2_weigh... 0
mc2_shap... 0



pyhf

130 views • Sep 19, 2018 1 0 SHARE SAVE ...



Lukas Heinrich
3 subscribers

SUBSCRIBE

Category People & Blogs

API: AUTOMATING UPLOADS

EXAMPLE RECAST → ZENODO

If experiments do adopt something like this, would be nice to have API connection to upload result.