# HEPData status

**Graeme Watt** (Database Manager)

HEPData advisory board meeting
IPPP Durham, 16th October 2015

http://hepdata.cedar.ac.uk

Follow @HepData

- Experiment self-submissions one year on
- Progress towards new hepdata.net site

# Modes of data entry

1. Manually harvested from "`hep-ex`" papers. Extract tables from `.tex` source files, mostly by assistant. TO DO: more automation desired.

2. Data directly submitted by experiments.

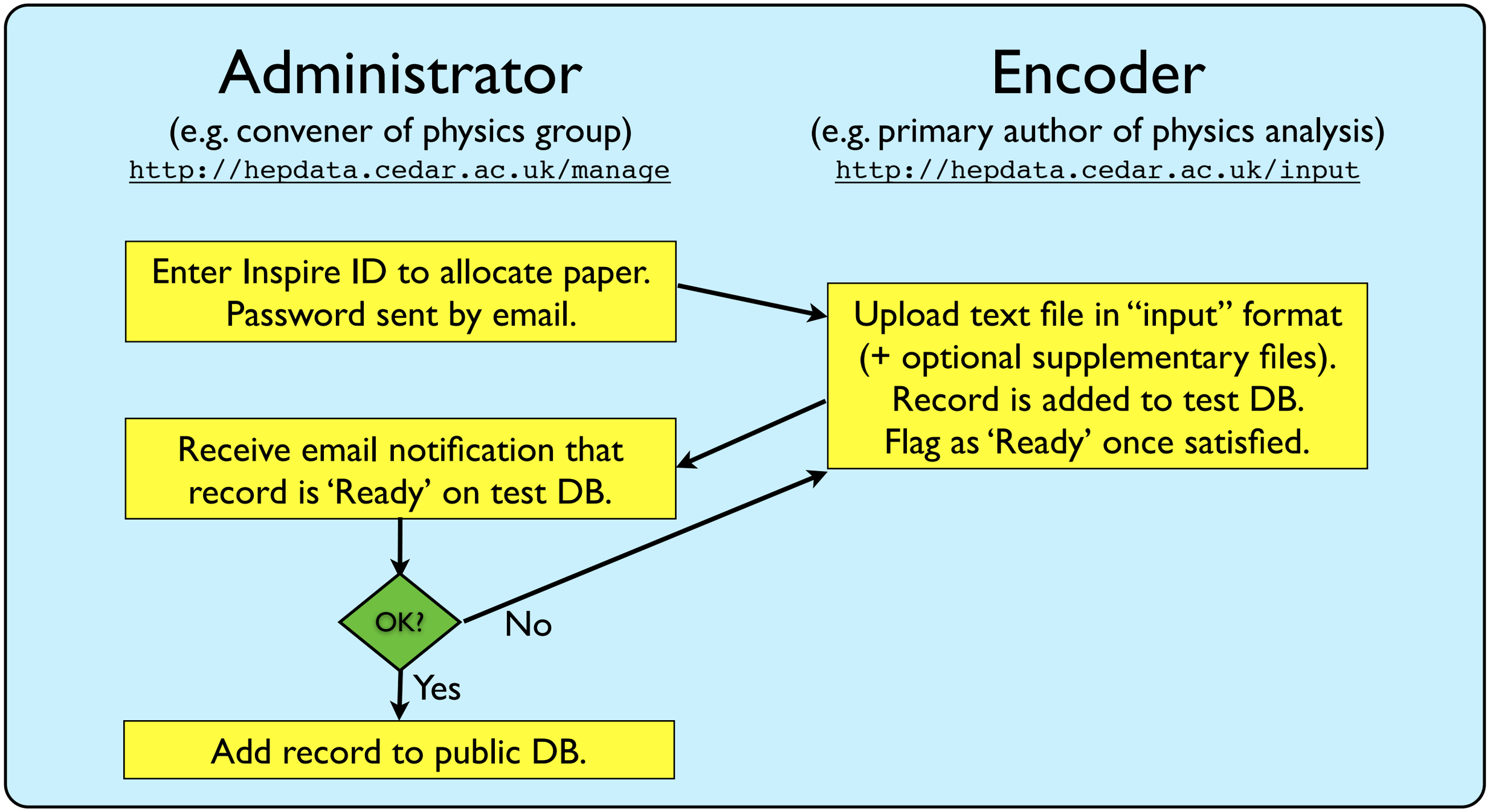
   - Past: no guidelines on preferred format, needed to manipulate into standard format.

   - Early 2014: encourage experiments to provide standard "input" text format.

Clearly need to move more papers from 1. to 2.

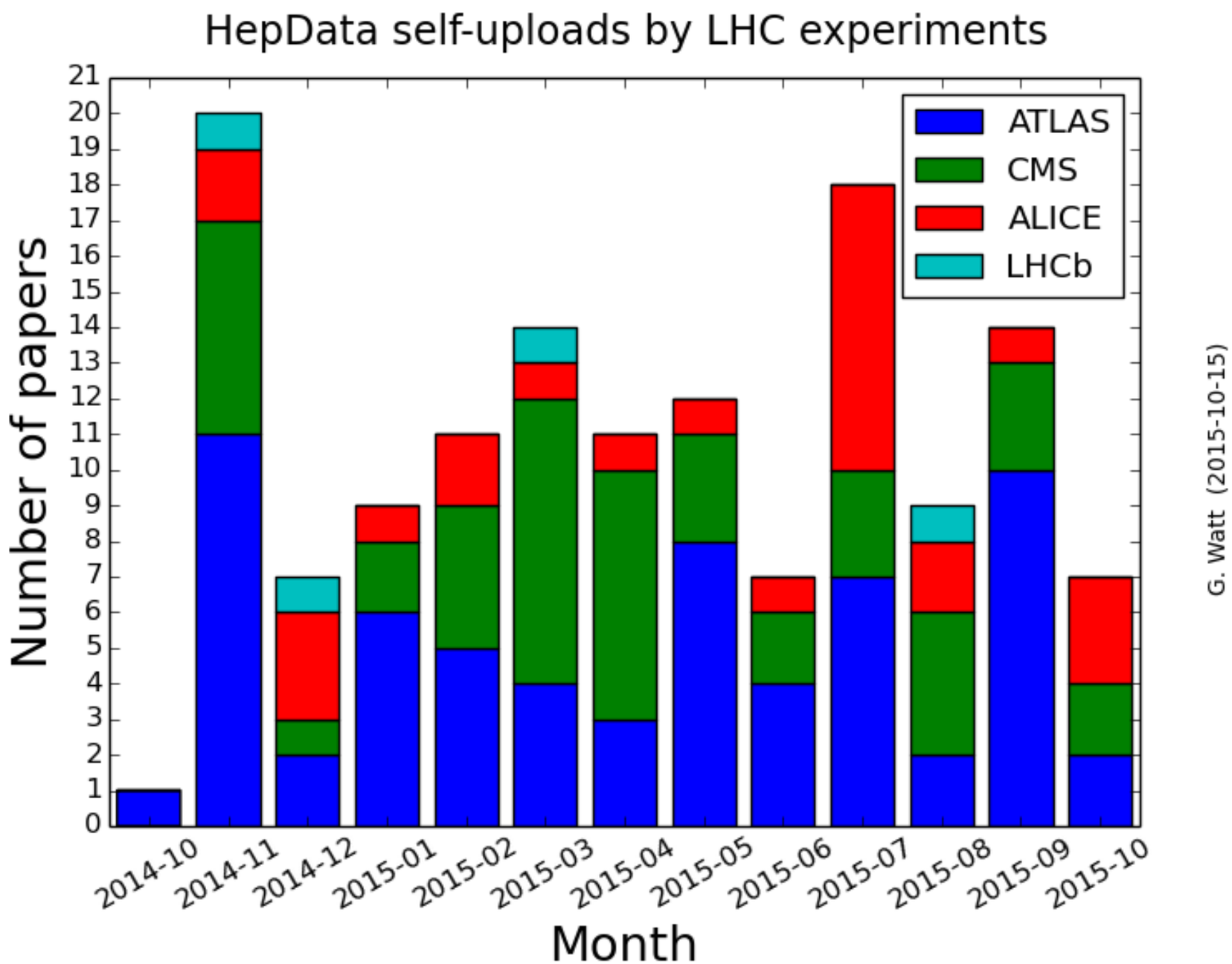
# Automated submission system

(Procedure agreed in HepData meeting at CERN on 26th June 2014)



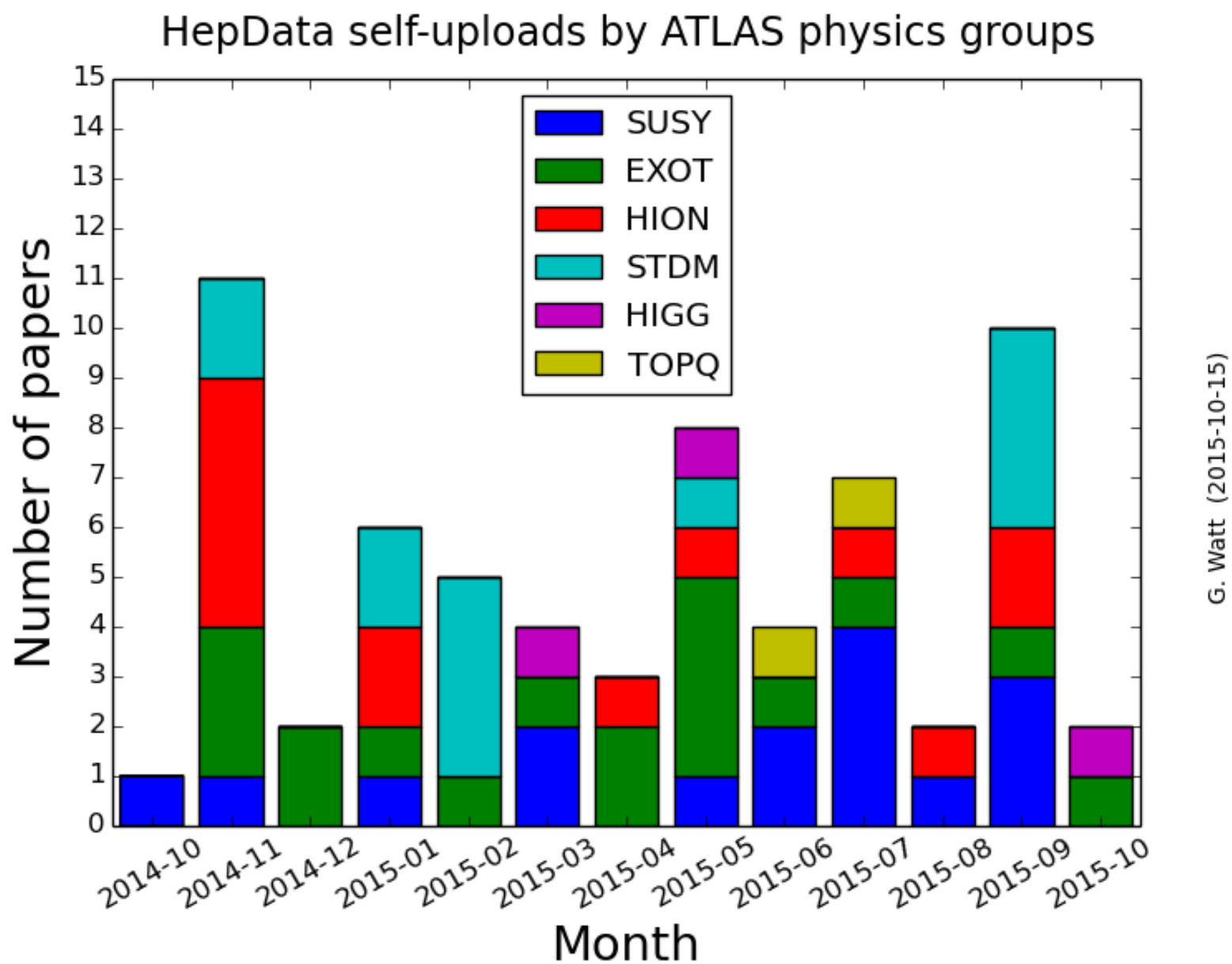

- First paper added by ATLAS SUSY group on 28th October 2014.

# Submission system usage



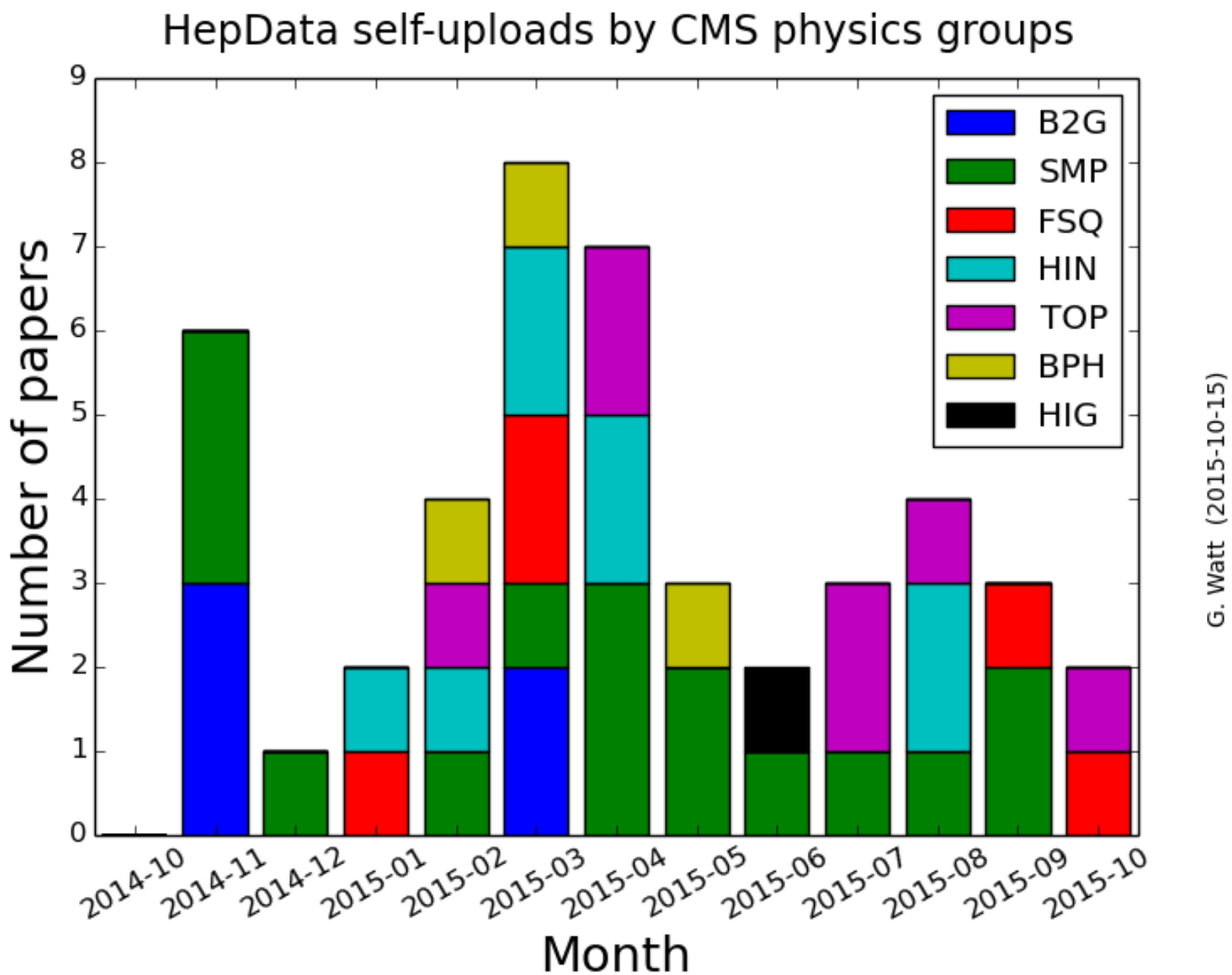

- Total of 140 LHC papers: ATLAS (65), CMS (45), ALICE (26), LHCb (4).

# Submission system usage by ATLAS




- Total of 65 ATLAS papers: SUSY (16), EXOT (18), HION (13), STDM (13), HIGG (3), TOPQ (2), BPHY (0).

# Submission system usage by CMS




- Total of 45 CMS papers: B2G (5), SMP (16), FSQ (5), HIN (8), TOP (7), BPH (3), HIG (1), SUS (0), EXO (0).

# Non-LHC experiments

- Submission system not yet used outside the four main LHC experiments.
- HERA papers often contain detailed tables of numbers: need to manually reformat.

  arXiv:1506.06042 (28 pages of data tables)
  http://hepdata.cedar.ac.uk/view/ins1377206

**ABRAMOWICZ 2015 — Combination of Measurements of Inclusive Deep Inelastic $e^{\pm}p$ Scattering Cross Sections and QCD Analysis of HERA Data**

- TO DO: engage non-LHC experiments more.

# Broadening the scope

- Tau decays from ALEPH (data read from plots):

**BUSKULIC 1994 — One prong tau decays into charged kaons**

**BUSKULIC 1994 — $K^0$ production in one-prong $\tau$ decays**

**BUSKULIC 1997 — A study of $\tau$ decays involving $\eta$ and $\omega$ mesons**

**BARATE 1997 — Measurement of the spectral functions of vector current hadronic tau decays**

**BARATE 1998 — Measurement of the axial-vector $\tau$ spectral functions and determination of $\alpha_s(M_\tau^2)$ from hadronic $\tau$ decays**

- Data for tuning detector simulations (GEANT):

**ACHASOV 2015 — Measurement of the $K_L$ nuclear interaction length in the NaI(Tl) calorimeter**

# HepData ➡ SOON HEPData

- Meeting with Inspire in December 2014 to start working jointly towards a new HEPData web site.
  - Overlay on Invenio v2 digital library software.
  - Modern interface and improved submission process.
  - Better searching, allocation of DOIs, versioning capabilities, closer integration of auxiliary files.
  - New input data format based on ~~ROOT~~ / YAML.

```
YAML: YAML Ain't Markup Language

What It Is: YAML is a human friendly data serialization
  standard for all programming languages.
```

# Old data "input" format

```
*dataset:
*location: Page 20 of preprint
*dscomment: The measured total cross sections.
The first systematic uncertainty is the combined
systematic uncertainty excluding luminosity, the
second is the luminosity
*reackey: P P --> Z0 Z0 X
*obskey: SIG
*qual: RE : P P --> Z0 Z0 X
*yheader: SIG(total) IN FB
*xheader: SQRT(S) IN GEV
*data: x : y
 7000; 6.7 +- 0.7 (DSYS=+0.4,-0.3,DSYS=0.3:lumi);
*dataend:
```

http://hepdata.cedar.ac.uk/view/ins1203852/d2/input

Table 2 ( Page 20 of preprint. ) HIDE DATA or as: input, plain text, AIDA, PyROOT, YODA, ROOT, mpl, ScaVis or MarcXML

The measured total cross sections. The first systematic uncertainty is the combined systematic uncertainty excluding luminosity, the second is the luminosity.

reaction keywords: [P P --> Z0 Z0 X]
observable keywords: [SIG]

| RE | P P --> Z0 Z0 X |
|---|---|
| SQRT(S) IN GEV | SIG(total) IN FB |
| 7000. | 6.7 ± 0.7 (stat) +0.4,-0.3 (sys) ± 0.3 (sys,lumi) |
| | Plot<br>SelectPlot |

- All data needs to be transformed into a standard "input" text format.
- Metadata for each table followed by data points in a structured format.
- `*reackey` and `*obskey` keywords can be used for searching database.

# New YAML text format

http://hepdata.cedar.ac.uk/view/ins1203852/d2/yaml

```
---
name: 'Table 2'
label: 'Data from Page 20 of preprint'
description: |
    The measured total cross sections.  The first systematic uncertainty is the
combined systematic uncertainty excluding luminosity, the second is the luminosity.
keywords:
  - {name: reactions, values: ['P P --> Z0 Z0 X']}
  - {name: observables, values: ['SIG']}
  - {name: cmenergies, values: [7000.0]}
additional_resources:
independent_variables:
  - header: {name: 'SQRT(S)', units: 'GEV'}
    values:
      - {value: 7000}
dependent_variables:
  - header: {name: 'SIG(total)', units: 'FB'}
    qualifiers:
      - {name: 'RE', value: 'P P --> Z0 Z0 X'}
    values:
      - value: 6.7
        errors:
          - {symerror: 0.7, label: 'stat'}
          - {asymerror: {plus: 0.4, minus: -0.3}, label: 'sys'}
          - {symerror: 0.3, label: 'sys,lumi'}
```

http://hepdata.net/record/60079

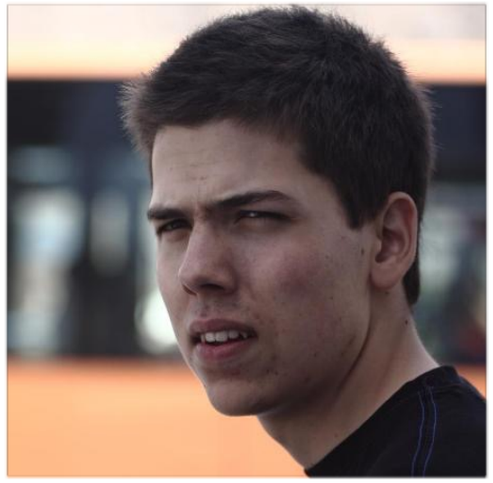
| RE | P P --> Z0 Z0 X |
|---|---|
| SQRT(S) [GEV] | SIG(total) [FB] |
| 7000 | 6.7 ±0.7 stat -0.3, 0.4 sys ±0.3 sys,lumi |

- Data from existing HepData database was exported to the new YAML format for migration to the new system.

# Work on HEPData by CERN team

Eamonn Maguire

Laura Rueda Garcia



Jan Stypka

+ Michal Szostak (CERN summer student)

+ others

Code: `https://github.com/HEPData`

- Amazing work from CERN team in only a few months!

# New site at hepdata.net

hepdata.net/record/60079

Apple HepData search Amazon News

HEPData Search HEP Data Search Help Sign in

Browse all | Aad, Georges et al. | Accessed 2 times (2.0/day) | Download Submission as

Hide Publication Information

## Measurement of $ZZ$ production in $pp$ collisions at $\sqrt{s} = 7$ TeV and limits on anomalous $ZZZ$ and $ZZ\gamma$ couplings with the ATLAS detector

Aad, Georges , Abajyan, Tatevik , Abbott, Brad , Abdallah, Jalal , Abdel Khalek, Samah , Abdelalim, Ahmed Ali , Abdinov, Ovsat , Aben, Rosemarie , Abi, Babak , Abolins, Maris

JHEP 1303 (2013) 128

ATLAS

DOI | View paper in Inspire

Additional Resources

**Abstract (data abstract)**

CERN-LHC. Measurements of the cross section for ZZ production using the 4l and 2l2nu decay channels in proton-proton collisions at a centre-of-mass energy of 7 TeV

Table 2

Data from Page 20 of preprint

The measured total cross sections. The first systematic uncertainty is the combined systematic uncertainty excluding luminosity, the second is the...

Table 3

Data from Figure 8A

Normalized ZZ fiducial cross section (multiplied by 10^6 for readability) in bins of the leading reconstructed dilepton pT for the...

Table 4

Data from Figure 8B

Normalized ZZ fiducial cross section (multiplied by 10^6 for readability) in bins of the dilepton pT for the 2l2nu channel....

Table 5

Data from Figure 9A

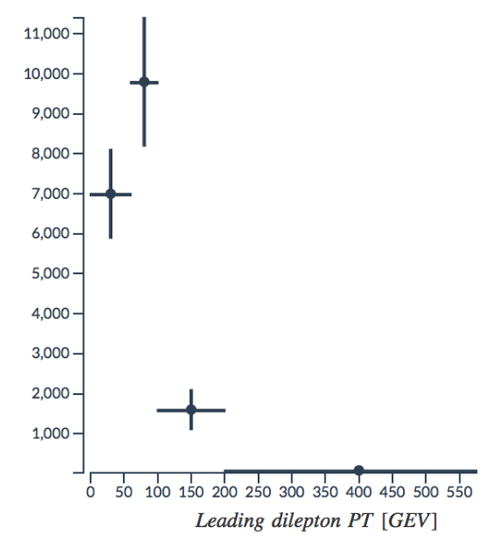
Normalized ZZ fiducial cross section (multiplied by 10^6 for readability) in bins of the leading reconstructed dilepton pT for the 4 lepton channel. The first systematic uncertainty is detector systematics, the second is background systematic uncertainties.

cmenergies 7000 | observables DSIG/DPT | reactions P P --> Z0 Z0 X

### Data

| SQRT(S) | 7000.0 GeV |
|---|---|
| RE | P P --> Z0 < LEPTON+ LEPTON- > Z0 < LEPTON+ LEPTON- > X |
| Leading dilepton PT [GEV] | 10**6 * 1/SIG(fiducial) * D(SIG(fiducial))/DPT [GEV**-1] |
| 0 - 60 | 7000 ±1100 stat ±79 sys,detector ±15 sys,background |
| 60 - 100 | 9800 ±1600 stat ±75 sys,detector ±15 sys,background |
| 100 - 200 | 1600 ±490 stat ±41 sys,detector ±2 sys,background |
| 200 - 600 | 80 ±60 stat ±2 sys,detector |

Visualize

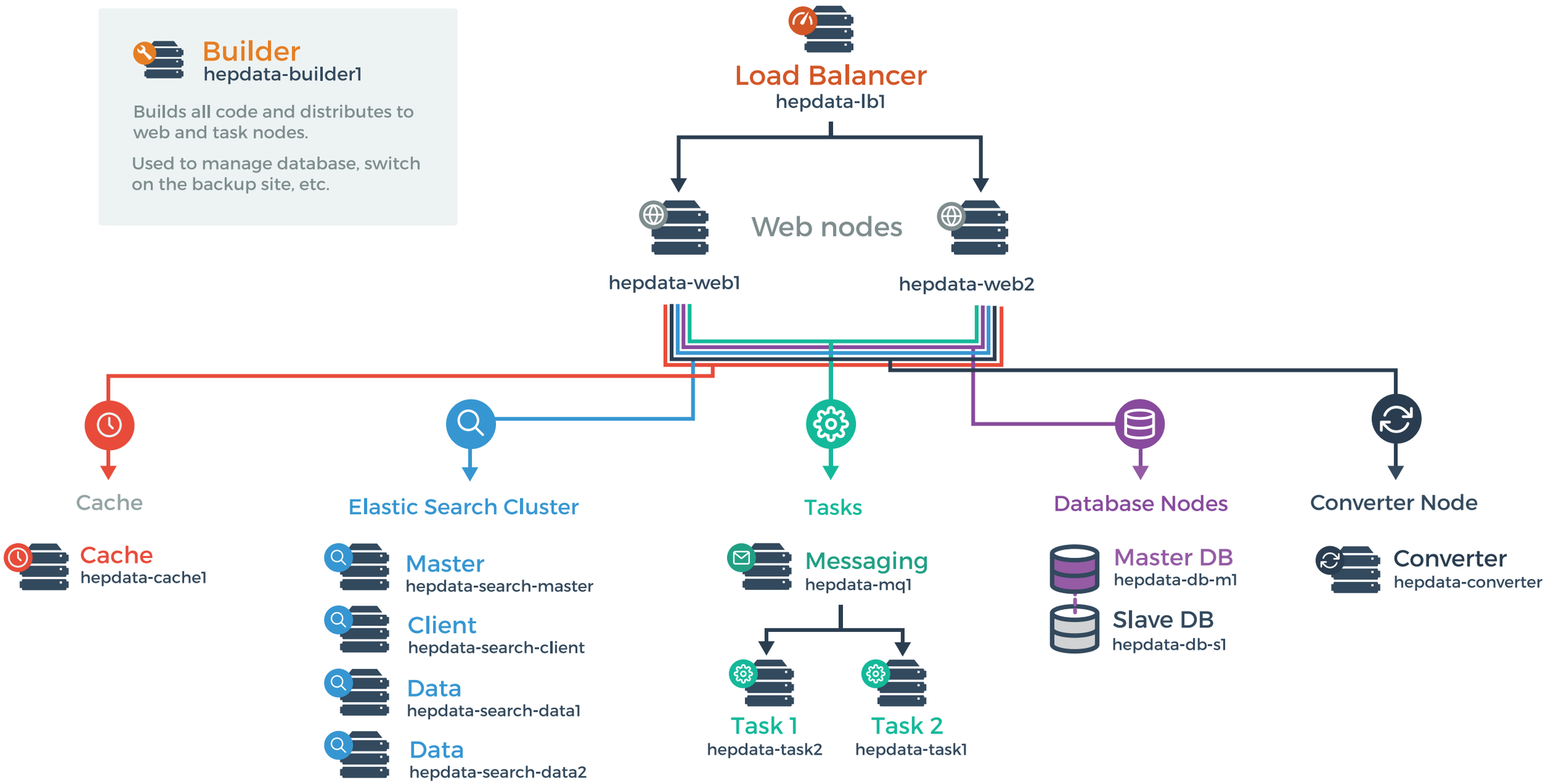



Sum errors: ☑

Copyright ~1975-Present, HEPData | Powered by Invenio, funded by STFC, UK, supported by IPPP Durham.

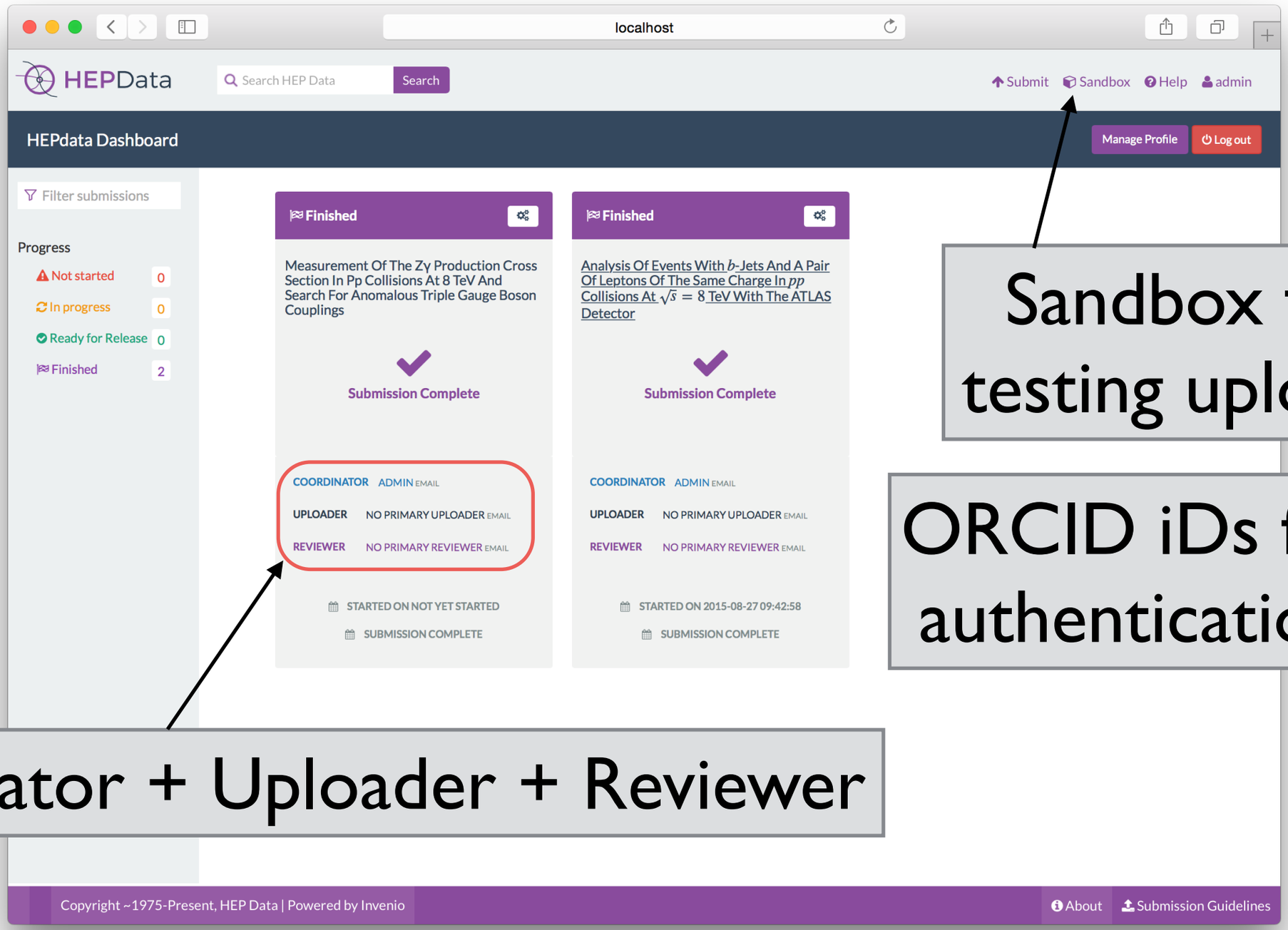
About | Submission Guidelines

# HEPData Openstack Architecture




- Environment comprised of 15 virtual machines.

# Submission dashboard

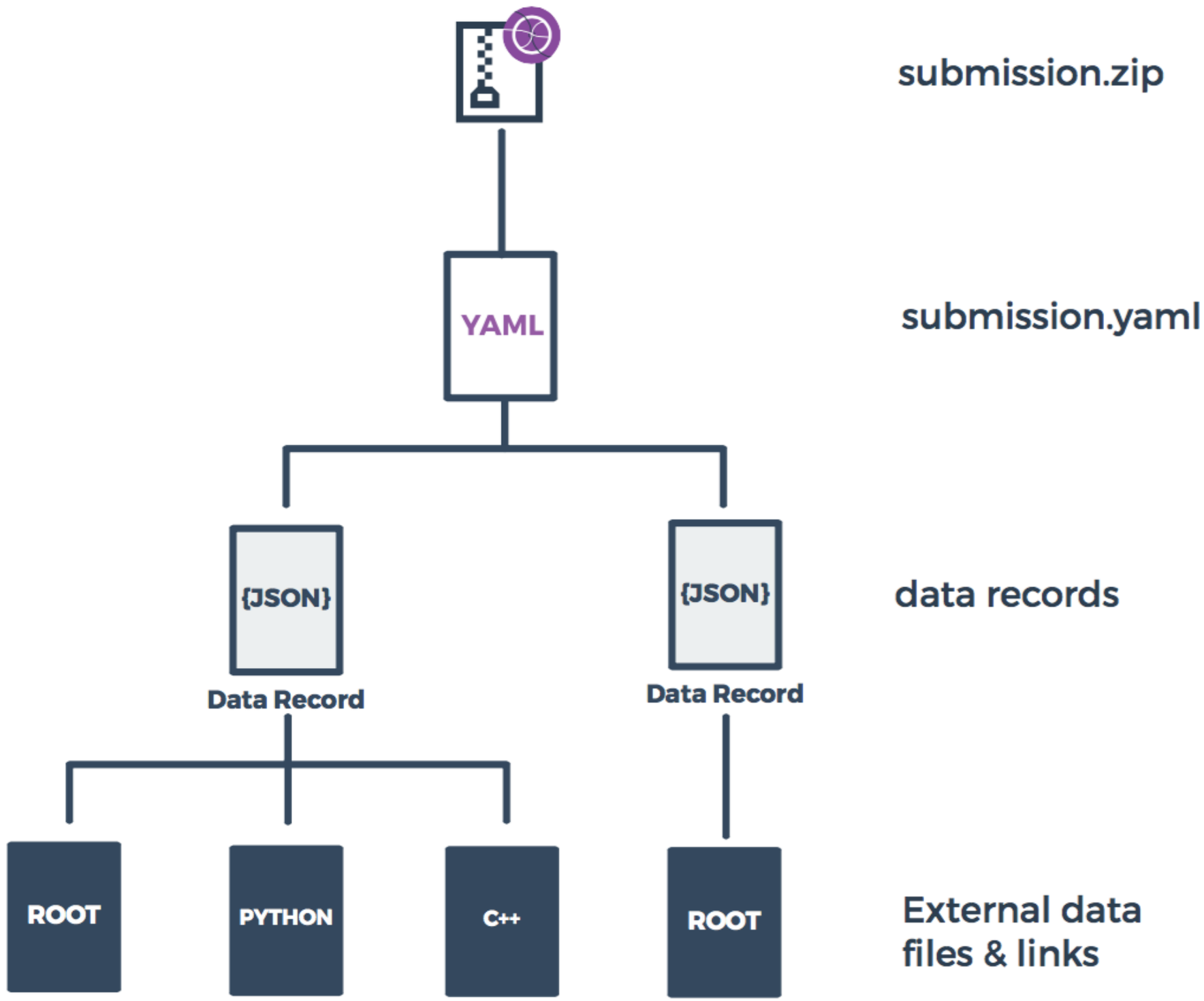



Sandbox for testing uploads

ORCID iDs for authentication

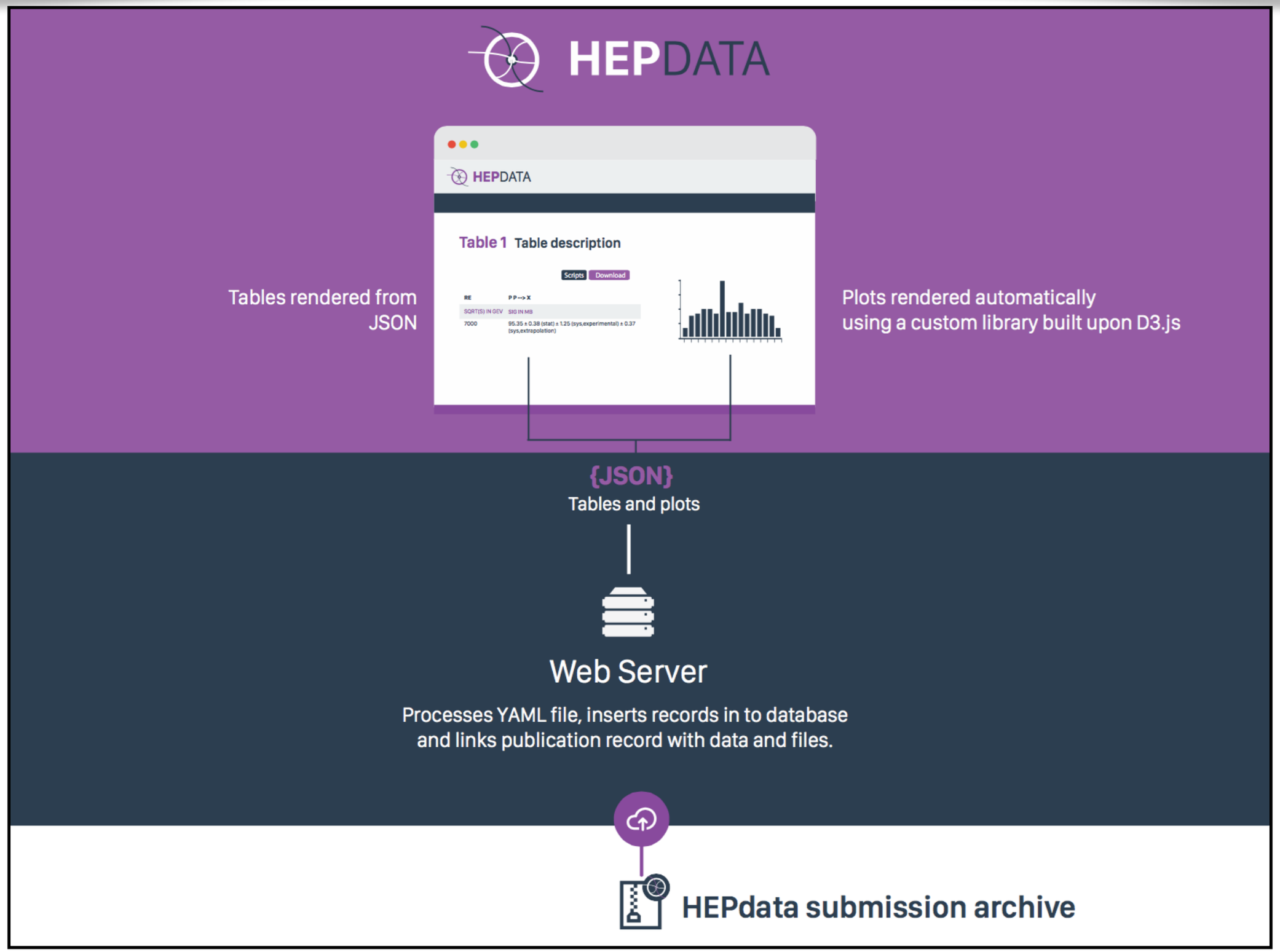
Coordinator + Uploader + Reviewer

# HEPData submission archive




http://github.com/HEPData/hepdata-submission

# Upload of submission archive



# Review system

Versioning

HEPData

Search HEPData | Search

Submit | Sandbox | Help | admin

Browse all | Mikkelsen, R.E. et al. | Accessed 5 times (5.0/day) | Download Submission as

Hide Publication Information

## Elastic photonuclear cross sections for bremsstrahlung from relativistic ions

Mikkelsen, R.E. , Sørensen, A.H. , Uggerhøj, U.I.

View paper in Inspire

**Abstract (data abstract)**
CERN-LHC. Measurements of the cross section for ZZ production using the 4l and 2l2nu decay channels in proton-proton collisions at a centre-of-mass energy of 7 TeV with 4.6 fb^-1 of data collected in 2011. The final states used are 4 electrons, 4 muons, 2 electrons and 2 muons, 2 electrons and missing transverse momentum, and 2 muons and missing transverse momentum (MET). The cross section values reported in the tables should be multiplied by a factor of 1.0141 to take

Upload New Files

Viewing version 2

Filter 9 data tables

Table 1

Page 17 of preprint

The measured fiducial cross sections. The first systematic uncertainty is the combined systematic uncertainty excluding luminosity, the second is the...

passed review

Table 2

Auxiliary Figure 9b.

Signal acceptance for the GGM model with tan(β)=30 in the combined electron and muon SR-Z.

passed review

Table 3

Figure 8A

Normalized ZZ fiducial cross section (multiplied by 10^6 for readability) in values of the leading reconstructed dilepton pT for the

Signal acceptance for the GGM model with tan(β)=30 in the combined electron and muon SR-Z.

SQRT(S) 8000.0 GeV

Data

| SQRT(S) | | 8000.0 GeV |
|---|---|---|
| MU [GEV] | M(GLUINO) [GEV] | ACCEPTANCE |
| 120 | 400 | 0.002229 |
| 150 | 400 | 0.004794 |
| 300 | 400 | 0.008519 |
| 390 | 400 | 0.005903 |
| 150 | 500 | 0.005636 |
| 200 | 500 | 0.01111 |
| 300 | 500 | 0.01587 |
| 400 | 500 | 0.01666 |
| 490 | 500 | 0.01149 |
| 120 | 600 | 0.00119 |
| 200 | 600 | 0.00873 |
| 300 | 600 | 0.0189 |
| 400 | 600 | 0.02529 |

Review Summary

Todo | Attention Required | Passed

Conversation

2015-10-13 09:47:08 admin: This table is missing some values.

No messages yet...

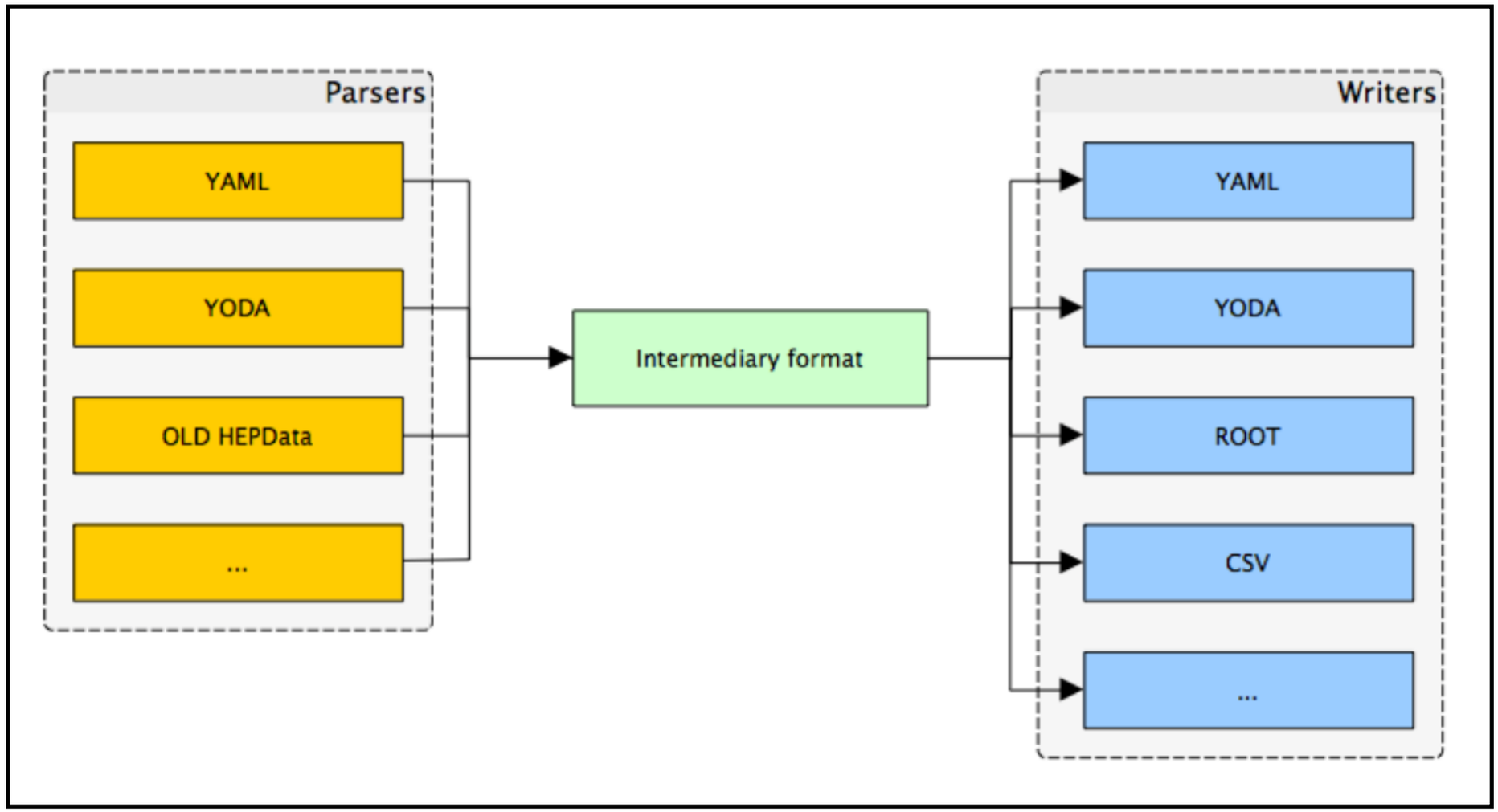
Send feedback on this data...

Send Feedback

Y Axis MU [GEV]

Copyright ~1975-Present, HEPData | Powered by Invenio, funded by STFC, UK, supported by IPPP Durham. | About | Submission Guidelines

# Conversion tools




- Work by Michal Szostak (CERN summer student)
  http://cds.cern.ch/record/2055193
  http://github.com/HEPData/hepdata-converter

# ROOT Import

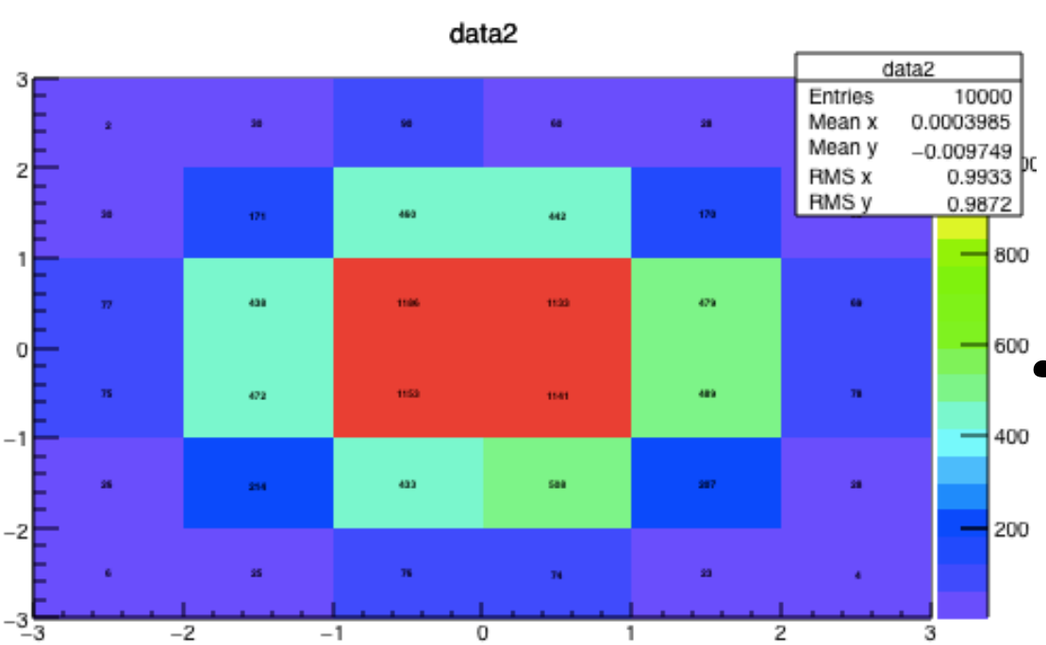
Slide from Lukas Heinrich (MCnet student)

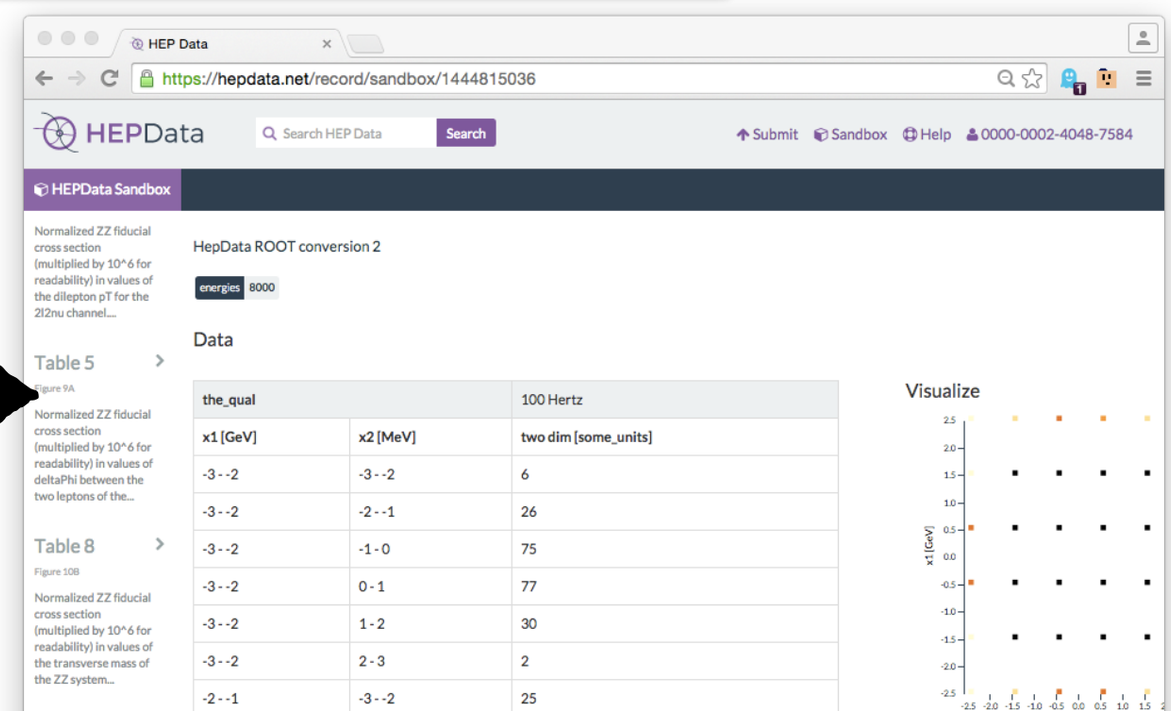
- popular demand by experiments.
- extension to native HepData YAML format to extract data from ROOT
- meta-data description still easily editable in YAML
- flexible choice of value extraction (multiple histograms for errors, native bin errors, user-defined python method)

```yaml
- name: MyName
  dependent_variables:
    - header: {name: Asymmetry, units: 'Bla'}
      conversion: {
        inputs: {nominal: 'input.root:nominal_histo', up: 'input.root:syst_up', down: 'input.root:syst_down'},
        formatter: nominal_with_variations_formatter,
        formatter_args: {label: syst}
      }
    - header: {name: Another Var, units: 'A Unit'}
      conversion: {
        inputs: {nominal: 'input.root:data1'},
        formatter: standard_format,
        formatter_args: {error_config: asymmetric, label: lumi}
      }
  independent_variables:
    - header: {name: x1, units: GeV}
```

form value with error in separate histos (systematics)

take bin error



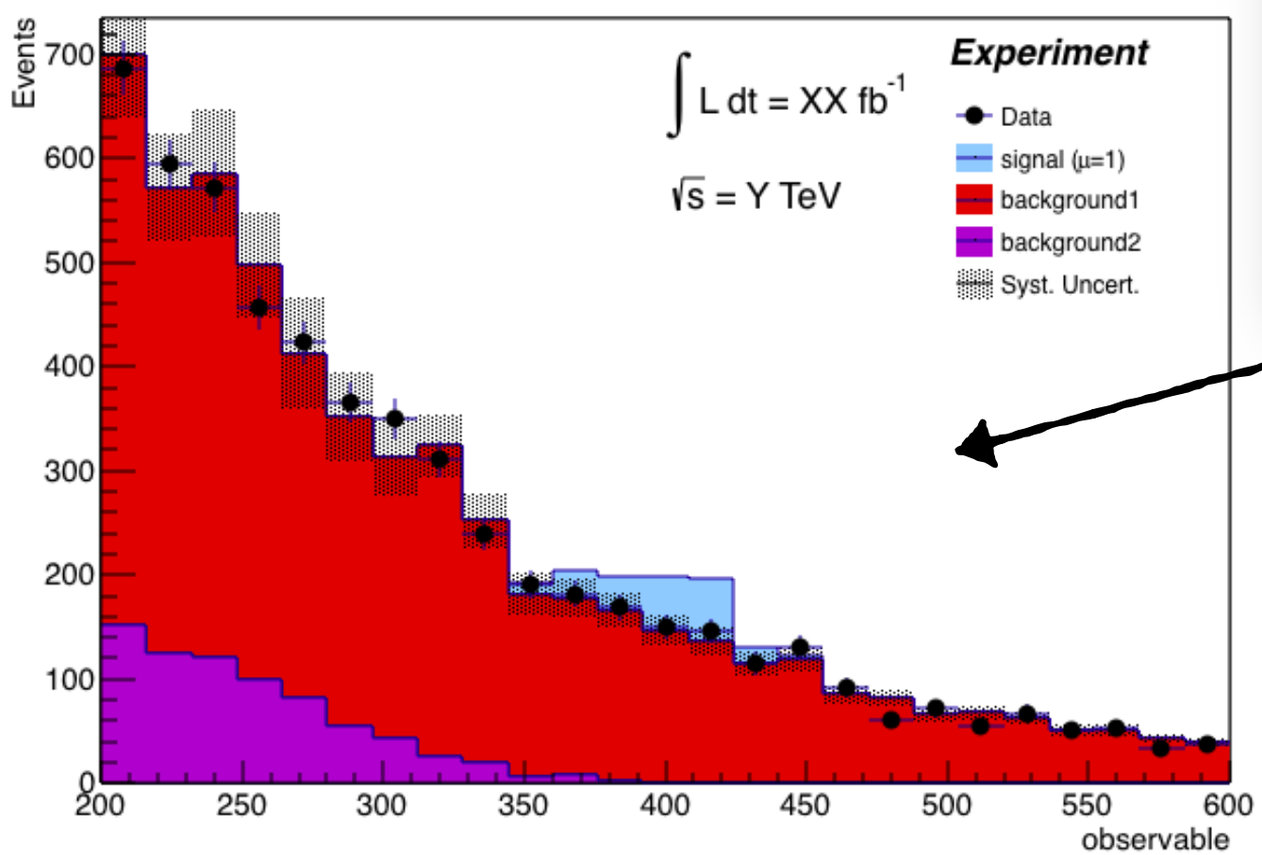



Slide from Lukas Heinrich (MCnet student)

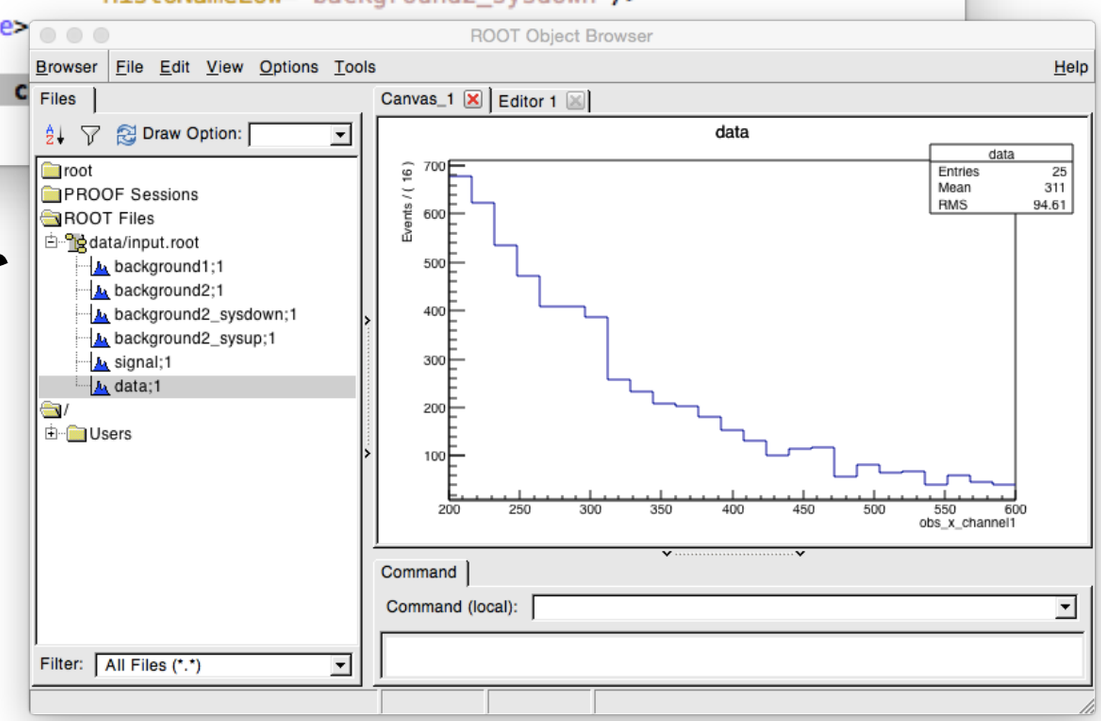
- HistFactory: model description framework
  - widely used in experiments, e.g. Higgs discovery, ~all SUSY analyses in ATLAS use HistFactory.
  - detailed description of distributions: resolved by component (signal, backgrounds) including systematic variations and correlations in well-formed schema.
  - format: (XML + ROOT)
  - underlying structure for many publication plots

Terminal — emacs — 75×19

```
<!DOCTYPE Channel  SYSTEM 'HistFactorySchema.dtd'>
<Channel Name="channel1" InputFile="./data/input.root" >
    <Data HistoName="data" HistoPath=""/>
    <Sample Name="signal" HistoPath="" HistoName="signal">
        <NormFactor Name="mu" Val="1s" High="10" Low="0"/>
    </Sample>
    <Sample Name="background1" HistoPath="" HistoName="background1">
        <OverallSys Name="OverallSyst1"
                    High="1.1"
                    Low="0.9"/>
    </Sample>
    <Sample Name="background2" HistoPath="" HistoName="background2">
        <HistoSys Name="HistoSys1"
                  HistoNameHigh="background2_sysup"
                  HistoNameLow="background2_sysdown"/>
    </Sample>
</Channel>
-uu-:---F1  c
```



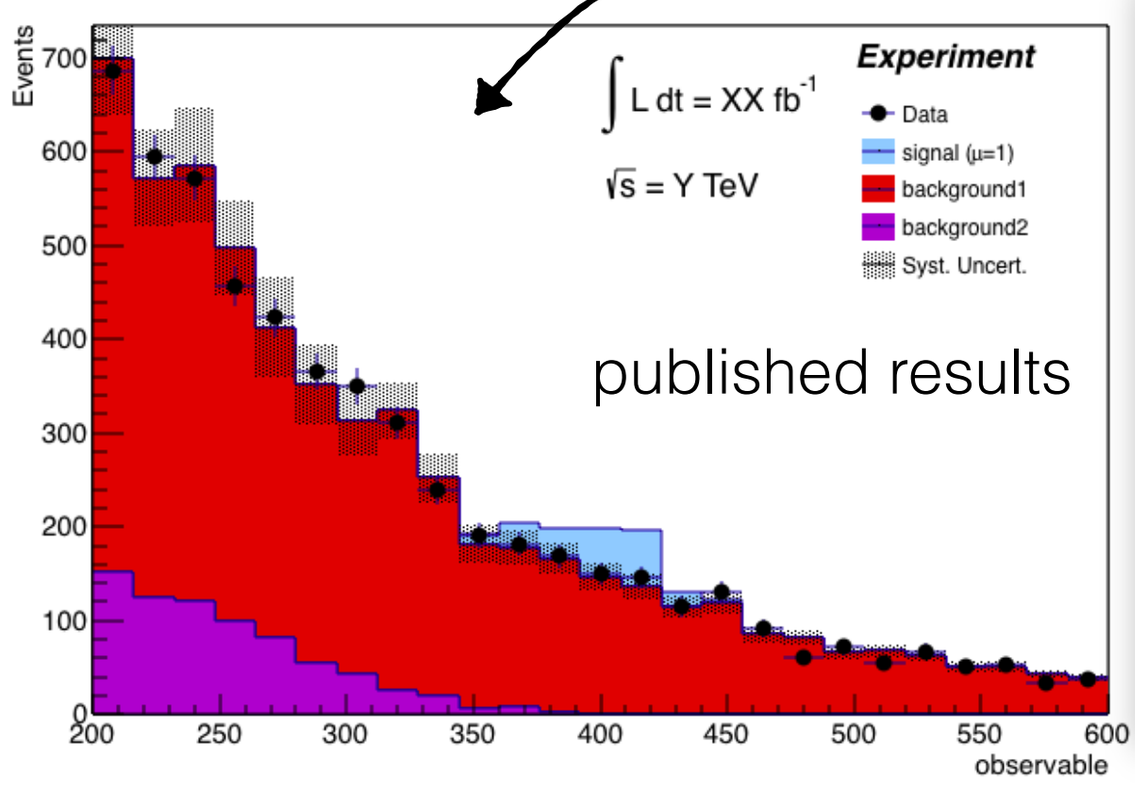

+

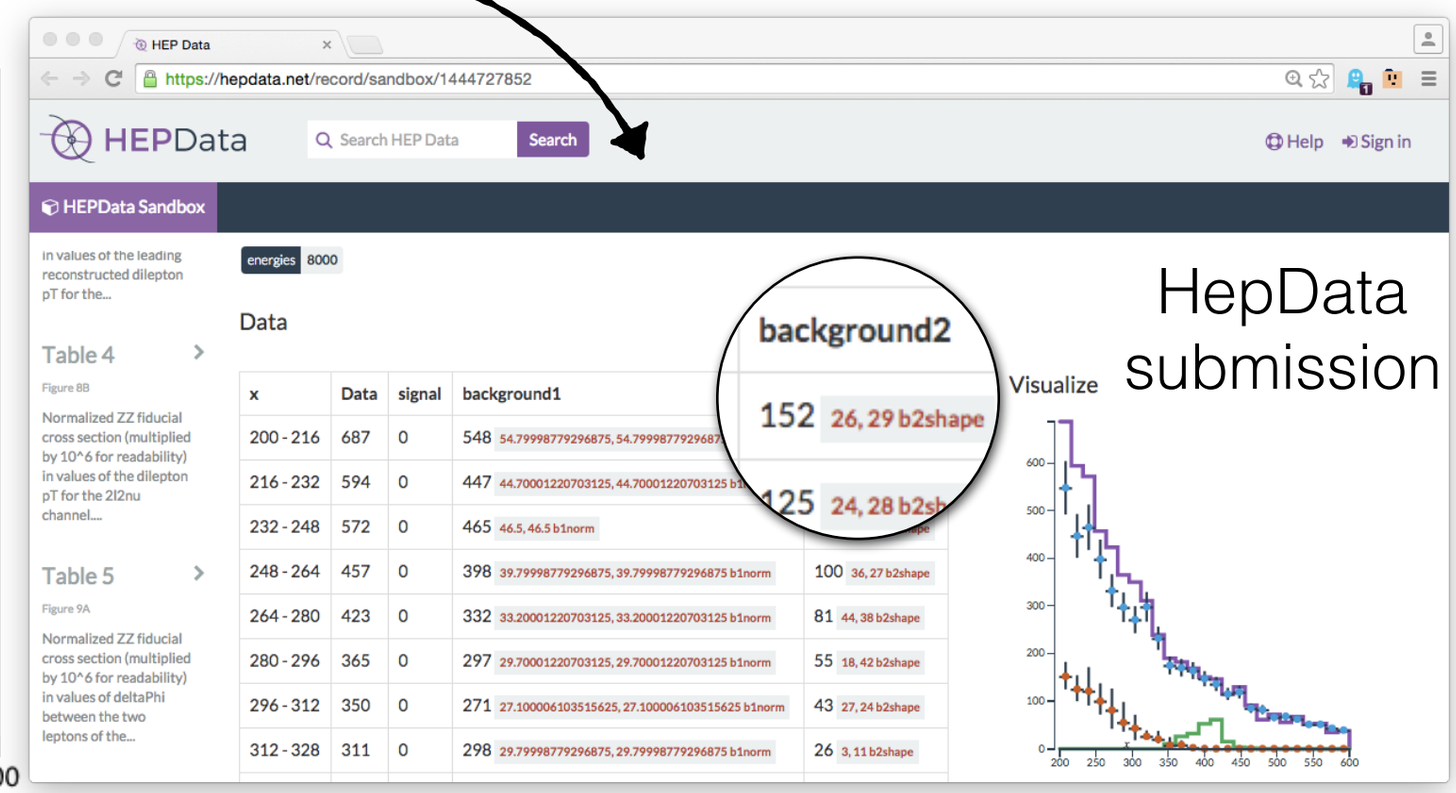


Slide from Lukas Heinrich (MCnet student)

- HepData import
  - full HistFactory configuration (XML+ROOT) as “additional resource”
    - complete, machine-readable, exportable
  - developed set of tools to extract HepData tables from HistFactory configuration
    - result, uncertainties, correlations directly from measurement instead of manual compilation by experimenters.
  - increased coherence between publication and HepData record.

HistFactory



# Summary

- Around 140 self-uploads from LHC experiments so far. Need to encourage use also by non-LHC experiments.
- New **HEPData** site developed by Inspire team at CERN nearing completion. Need to take over development.
- Experiments should soon make the transition to the new submission system and the associated new YAML format.

http://hepdata.cedar.ac.uk

http://hepdata.net

[Credit: some material taken from talk by Eamonn Maguire.]

Slide from Frank Krauss

# Deliverables for HEPData (this year)

- integration with Inspire/Invenio (until end of 2015) ✓
  - definition of data model (done) ✓
  - migration of data sets (under way, some glitches) ✓
  - re-implementation of submission system ✓
    (last issues being ironed out)
  - versioning (last issues being ironed out) ✓
  - algorithms for cross-discovery and cross-display 😐
    (not yet tested in new framework)

7802/8195 records

Project for Juan Luis Boya

Slide from Frank Krauss

# Deliverables for HEPData 2.0

- re-implementation of search system
  - some manual labour to be done (→ Mike Whalley)
- better methods to integrate supporting material
  - additional plots, correlation matrices, (analysis) software, etc.
- direct upload in ROOT format (first steps) ☺
- new process types: distribution in decays, particle interactions with matter ☺
- maintaining service → distribution of tasks with Invenio