



Experience with HEPData

Jae Hyeok Yoo (UCSB)

04/11/2017

CMS SUSY Workshop @ Ghent

Disclaimer

- I don't work for HEPData (I work for UCSB)
- I am sharing my experience with HEPData for [JHEP 08 \(2016\) 122](#) (SUS-15-007: RA4-MJ paper with 2015 data)
 - HEPData entry for SUS-15-007: <https://hepdata.net/record/ins1459054>

Why am I talking about it?

- Now all CMS SUSY papers are obligated to enter the results into HEPData
- There are two CMS SUSY analyses with 2015 data in HEPData
 - RA2/b: <https://hepdata.net/record/ins1422778>
 - RA4-MJ: <https://hepdata.net/record/ins1459054>
- I will walk you through the steps I followed

What is HEPData?

- According to its webpage (<https://hepdata.net/about>)

The Durham High Energy Physics Database (HEPData) has been built up over the past four decades as **a unique open-access repository for scattering data from experimental particle physics**. It currently comprises **the data points from plots and tables** related to several thousand publications including those from the Large Hadron Collider (LHC). HEPData is funded by a grant from the UK STFC and is based at the IPPP at Durham University.

Submission procedure

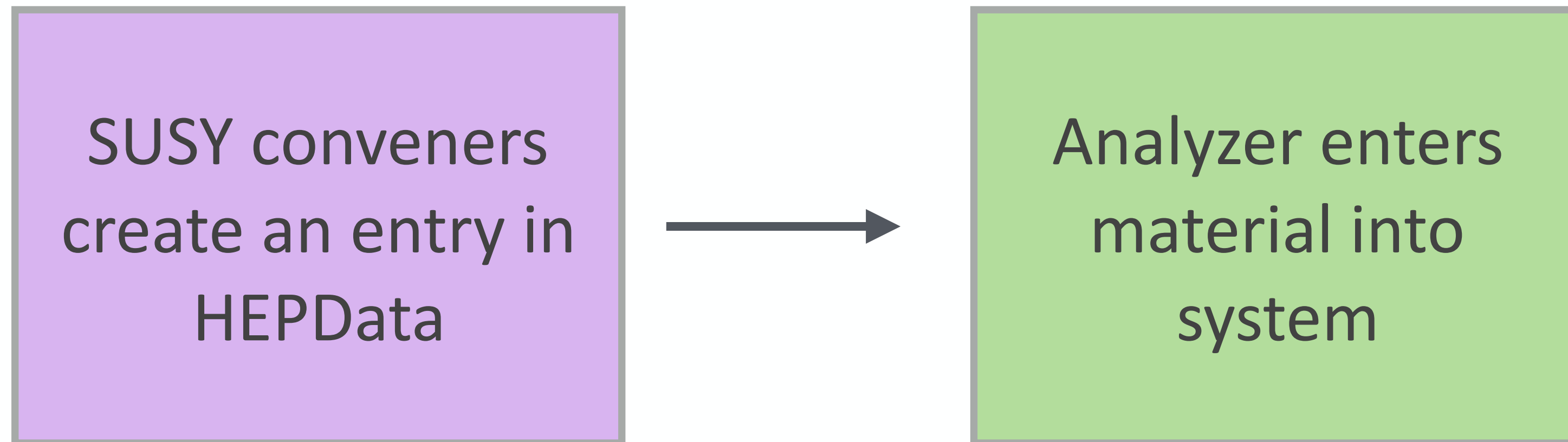
- NB: you need to have a publication to start with

Submission procedure

SUSY conveners
create an entry in
HEPData

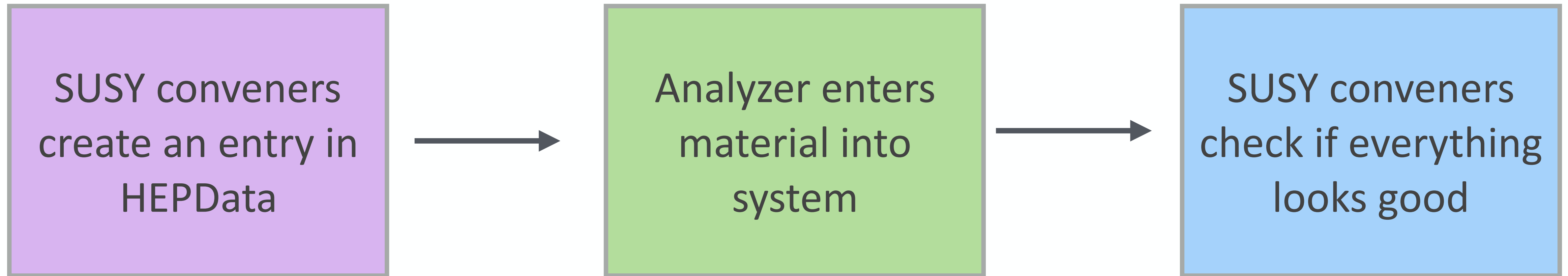
- * SUSY conveners: “coordinator” in HEPData nomenclature
- * Analyzer: “uploader” in HEPData nomenclature

Submission procedure



- * SUSY conveners: “coordinator” in HEPData nomenclature
- * Analyzer: “uploader” in HEPData nomenclature

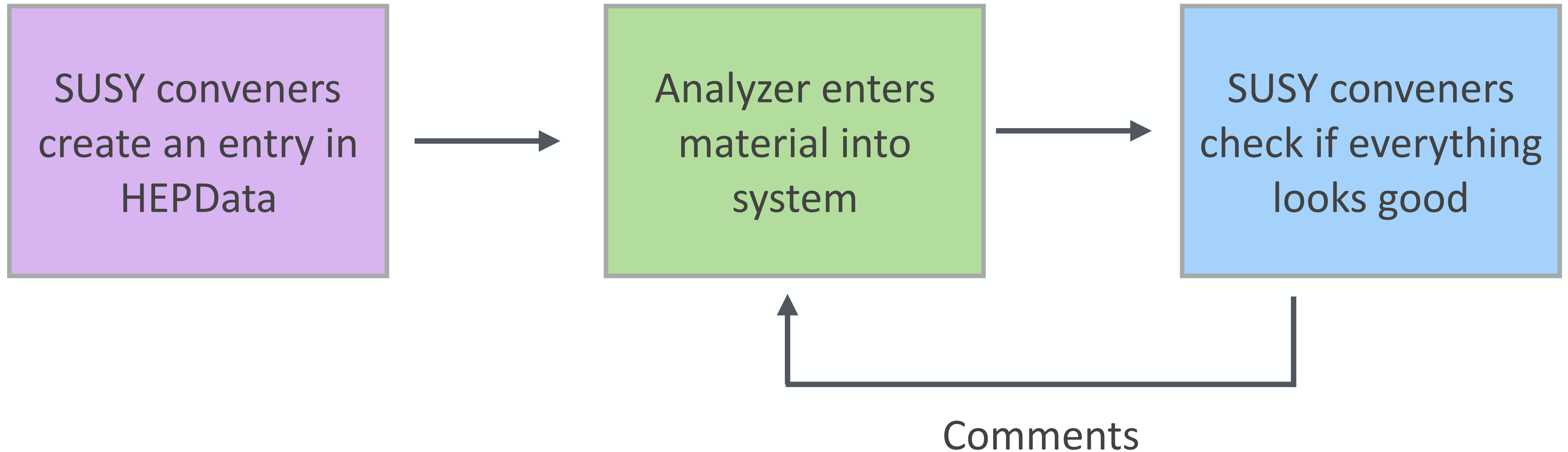
Submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

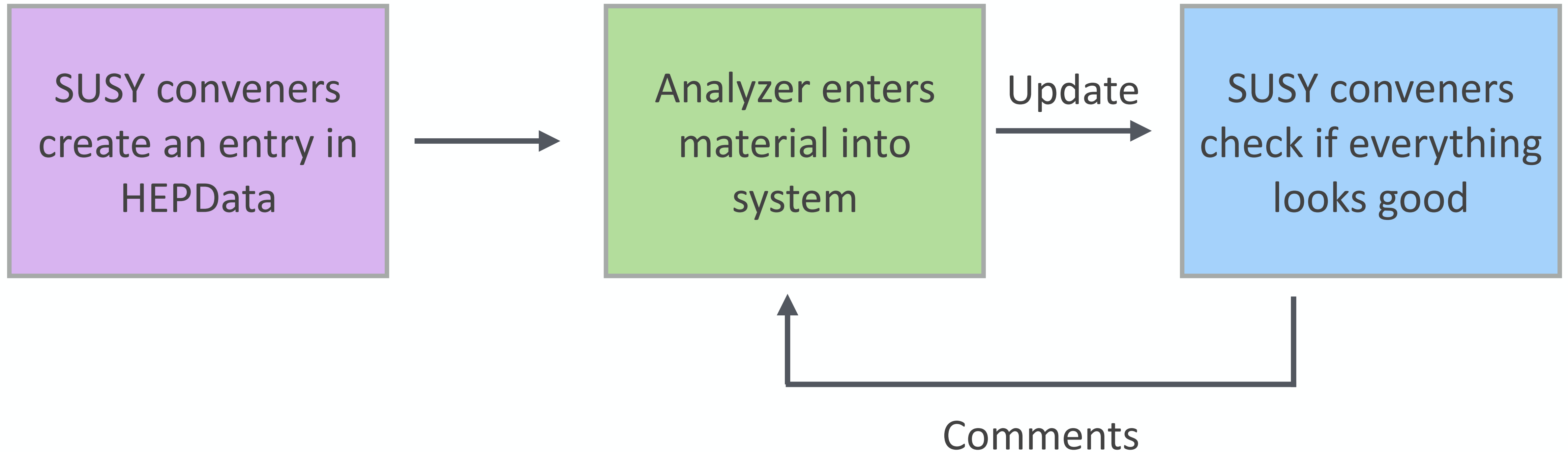
Submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

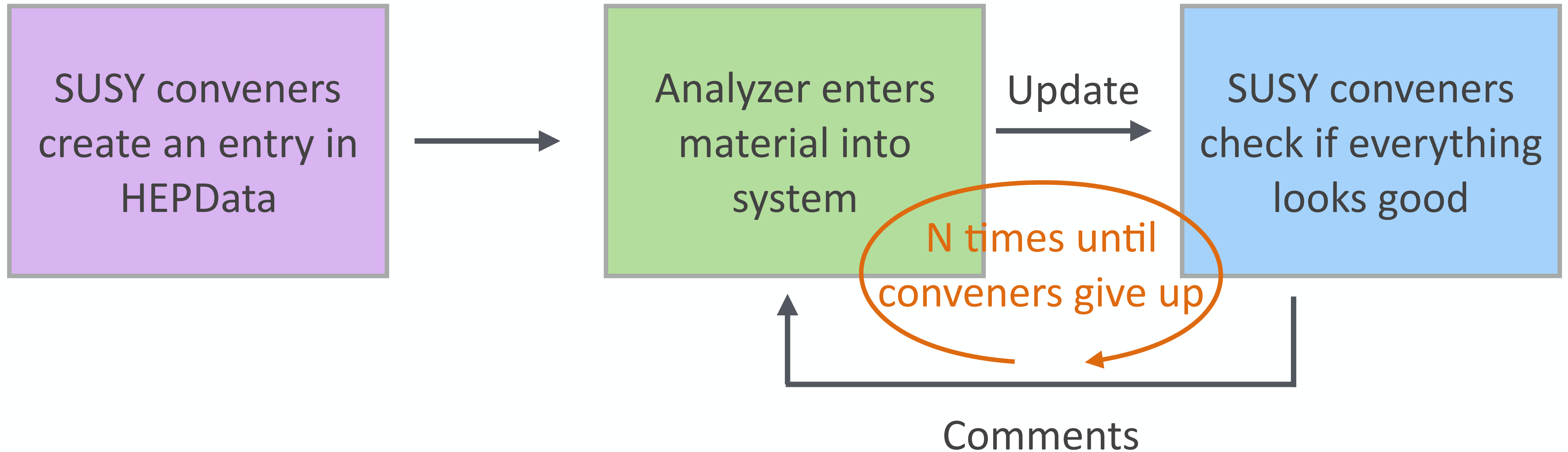
Submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

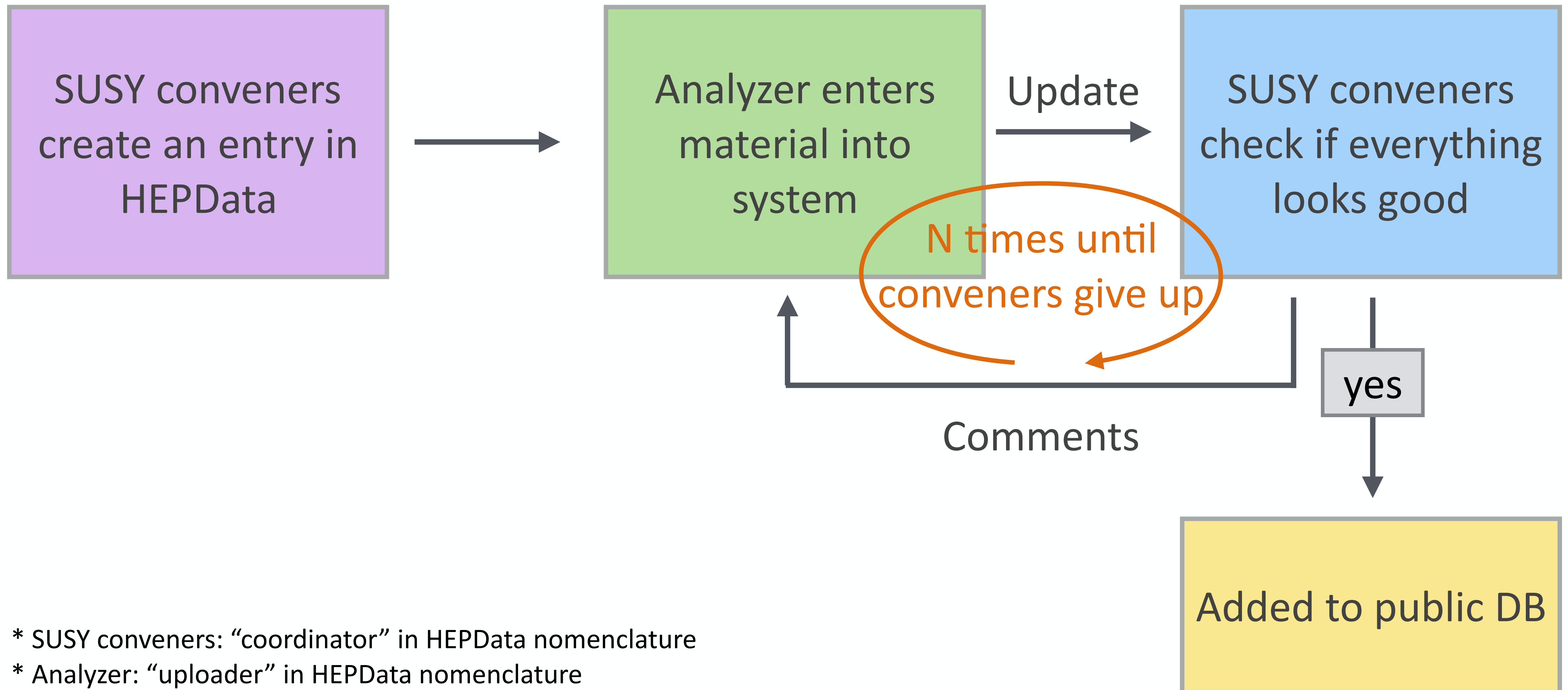
Submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

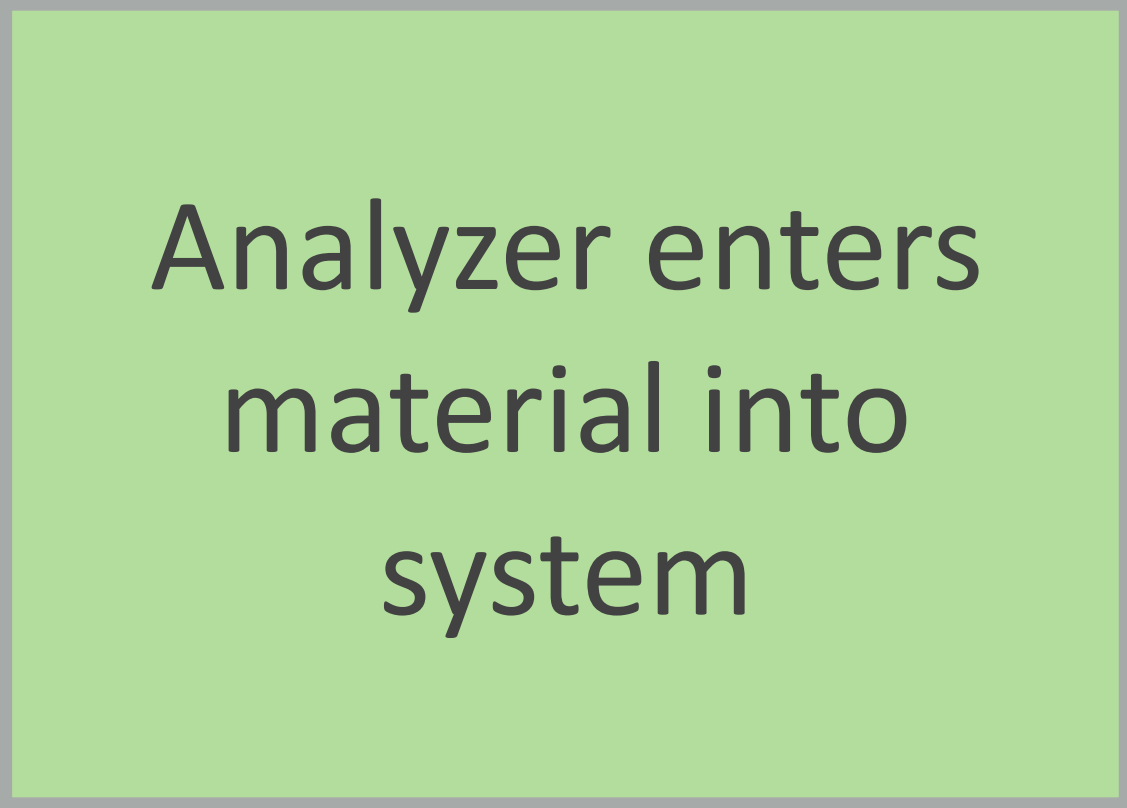
Submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

Submission procedure



Analyzer enters
material into
system

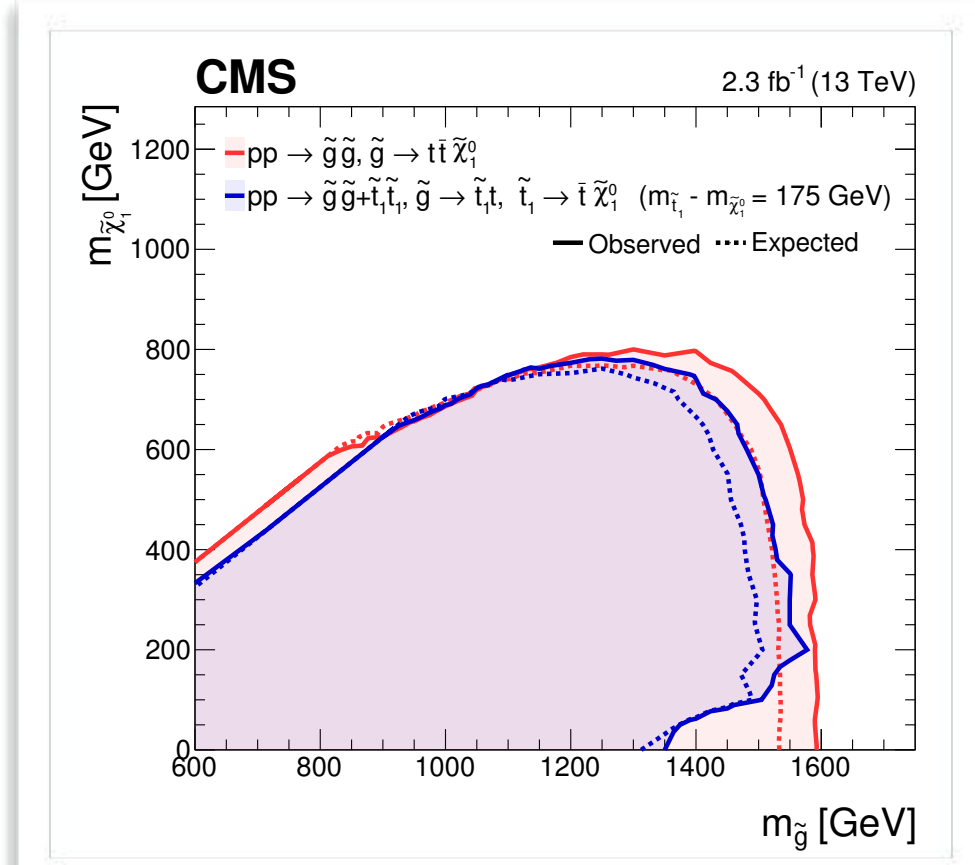
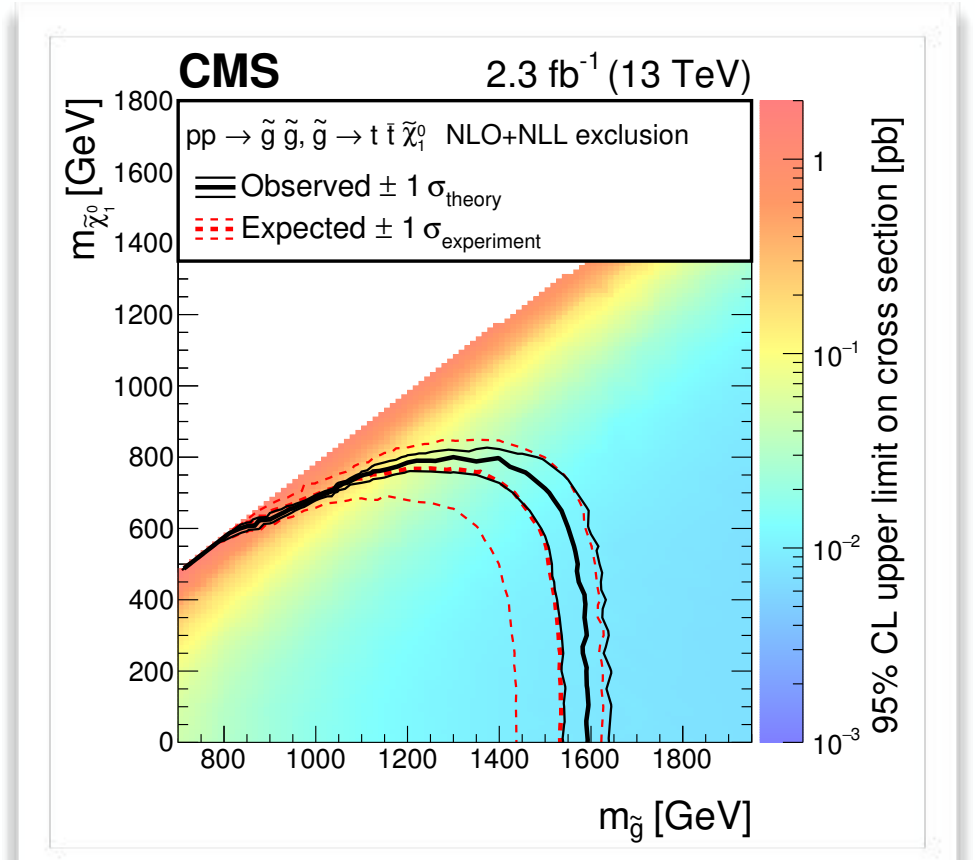
- Discuss this step further
 - What need to be entered?
 - How to enter them?
 - Check what you entered

What need to be entered?

- In principle, you need to enter all plots and tables in the paper
- But, some of them may not be very interesting to theorists
 - For example, MC-only plots
- For SUS-15-007, I've entered
 - Abstract
 - Table of predictions and observations
 - Exclusion limit plots
 - Cut-flow table

Region: bin	κ	T1tttt(NC)	T1tttt(C)	Fitted μ^{bkg} (PF)	Fitted μ^{bkg} (GF)	Obs.
$200 < E_T^{\text{miss}} \leq 400 \text{ GeV}$						
R1: all N_{jets}, N_b	—	0.1	3.2	336.0 ± 18.3	335.3 ± 18.2	336
R2: $6 \leq N_{\text{jets}} \leq 8, N_b = 1$	—	0.1	0.2	47.1 ± 6.9	49.5 ± 6.9	47
R2: $N_{\text{jets}} \geq 9, N_b = 1$	—	0.1	0.3	7.0 ± 2.6	7.5 ± 2.7	7
R2: $6 \leq N_{\text{jets}} \leq 8, N_b = 2$	—	0.1	0.3	42.0 ± 6.5	41.1 ± 6.2	42
R2: $N_{\text{jets}} \geq 9, N_b = 2$	—	0.1	0.5	7.0 ± 2.6	6.6 ± 2.5	7
R2: $6 \leq N_{\text{jets}} \leq 8, N_b \geq 3$	—	0.1	0.2	12.0 ± 3.5	11.1 ± 3.2	12
R2: $N_{\text{jets}} \geq 9, N_b \geq 3$	—	0.2	0.6	1.0 ± 1.0	0.9 ± 0.9	1
R3: all N_{jets}, N_b	—	0.2	3.8	21.0 ± 4.6	21.6 ± 4.2	21
R4: $6 \leq N_{\text{jets}} \leq 8, N_b = 1$	$1.12 \pm 0.09 \pm 0.43$	0.2	0.2	3.3 ± 1.4	3.6 ± 1.0	6
R4: $N_{\text{jets}} \geq 9, N_b = 1$	$0.91 \pm 0.06 \pm 0.81$	0.2	0.4	0.4 ± 0.3	0.4 ± 0.2	1
R4: $6 \leq N_{\text{jets}} \leq 8, N_b = 2$	$1.11 \pm 0.06 \pm 0.42$	0.3	0.4	2.9 ± 1.2	2.9 ± 0.8	2
R4: $N_{\text{jets}} \geq 9, N_b = 2$	$1.05 \pm 0.11 \pm 0.94$	0.3	0.6	0.5 ± 0.3	0.4 ± 0.2	0
R4: $6 \leq N_{\text{jets}} \leq 8, N_b \geq 3$	$1.25 \pm 0.11 \pm 0.47$	0.3	0.3	0.9 ± 0.4	0.9 ± 0.3	0
R4: $N_{\text{jets}} \geq 9, N_b \geq 3$	$1.05 \pm 0.10 \pm 0.93$	0.3	0.7	0.1 ± 0.1	0.1 ± 0.1	0
$E_T^{\text{miss}} > 400 \text{ GeV}$						
R1: all N_{jets}, N_b	—	0.1	0.5	16.0 ± 4.0	17.1 ± 4.0	16
R2: $6 \leq N_{\text{jets}} \leq 8, N_b = 1$	—	0.2	0.1	8.0 ± 2.8	6.8 ± 2.5	8
R2: $N_{\text{jets}} \geq 9, N_b = 1$	—	0.1	0.2	1.0 ± 1.0	1.7 ± 1.2	1
R2: $6 \leq N_{\text{jets}} \leq 8, N_b \geq 2$	—	0.5	0.3	3.0 ± 1.7	2.5 ± 1.4	3
R2: $N_{\text{jets}} \geq 9, N_b \geq 2$	—	0.4	0.6	1.0 ± 1.0	0.9 ± 0.9	1
R3: all N_{jets}, N_b	—	0.4	0.9	4.0 ± 2.0	2.9 ± 1.4	4
R4: $6 \leq N_{\text{jets}} \leq 8, N_b = 1$	$1.09 \pm 0.16 \pm 0.42$	0.7	0.2	2.2 ± 1.7	1.2 ± 0.7	0
R4: $N_{\text{jets}} \geq 9, N_b = 1$	$0.98 \pm 0.16 \pm 0.87$	0.4	0.3	0.2 ± 0.3	0.3 ± 0.2	1
R4: $6 \leq N_{\text{jets}} \leq 8, N_b \geq 2$	$1.29 \pm 0.22 \pm 0.50$	1.9	0.5	1.0 ± 0.8	0.5 ± 0.4	0
R4: $N_{\text{jets}} \geq 9, N_b \geq 2$	$0.90 \pm 0.14 \pm 0.80$	1.6	1.0	0.2 ± 0.3	0.1 ± 0.1	0

$\mathcal{L} = 2.3 \text{ fb}^{-1}$	Other	QCD	t \bar{t} V	Single t	W+jets	t \bar{t} (1 ℓ)	t \bar{t} (2 ℓ)	SM bkg.	T1tttt(NC)	T1tttt(C)
No selection	—	—	—	—	—	—	—	—	31.3	190.0
$1\ell, p_T > 20 \text{ GeV}$	—	—	—	—	—	—	—	—	11.9	68.7
$H_T > 500 \text{ GeV}$	4131.9	31831.5	721.9	2926.6	31885.1	27628.7	3357.8	102483.4	11.9	44.9
$E_T^{\text{miss}} > 200 \text{ GeV}$	310.6	154.7	89.1	457.2	4343.1	2183.6	584.0	8122.3	10.5	21.5
$N_{\text{jets}} \geq 6, p_T > 30 \text{ GeV}$	27.3	8.0	36.8	82.8	278.7	792.3	171.4	1397.4	9.6	20.4
$N_b \geq 1$	9.4	2.7	29.6	63.9	66.3	632.2	137.4	941.4	9.1	19.1
$M_J > 250 \text{ GeV}$	6.7	2.6	22.6	43.8	46.1	455.2	87.2	664.2	9.0	16.5
$m_T > 140 \text{ GeV}$	0.7	1.4	3.0	3.5	1.2	5.5	32.5	47.9	7.0	9.2
$M_J > 400 \text{ GeV}$	0.4	0.8	1.1	1.4	0.6	2.8	9.7	16.7	6.4	4.5
$N_b \geq 2$	0.16	0.04	0.55	0.68	0.00	1.29	4.52	7.24	4.87	3.47
$E_T^{\text{miss}} > 400 \text{ GeV}$	0.02	0.00	0.12	0.31	0.00	0.07	0.72	1.24	3.60	1.48
$N_{\text{jets}} \geq 9, p_T > 30 \text{ GeV}$	0.01	0.00	0.03	0.00	0.00	0.01	0.11	0.16	1.64	1.00



How to enter the info?: “input format”

Input format for the observed limit line

```
*dataset:
*location: Figure 10
*dscomment: Interpretation of results in the T1tttt model. The
colored regions show the upper limits (95\% CL) on the
production cross section for  $pp \rightarrow \tilde{g}\tilde{g}, \tilde{g} \rightarrow t\bar{t}\tilde{\chi}^0_1$  in the
 $m_{\tilde{g}}-m_{\tilde{\chi}^0_1}$  plane. The curve shows
the observed limit on the corresponding SUSY particle masses
obtained by comparing the excluded cross section with
theoretical cross sections.
```

```
*reackey: P P --> GLUINO GLUINO X
*obskey: CLS
```

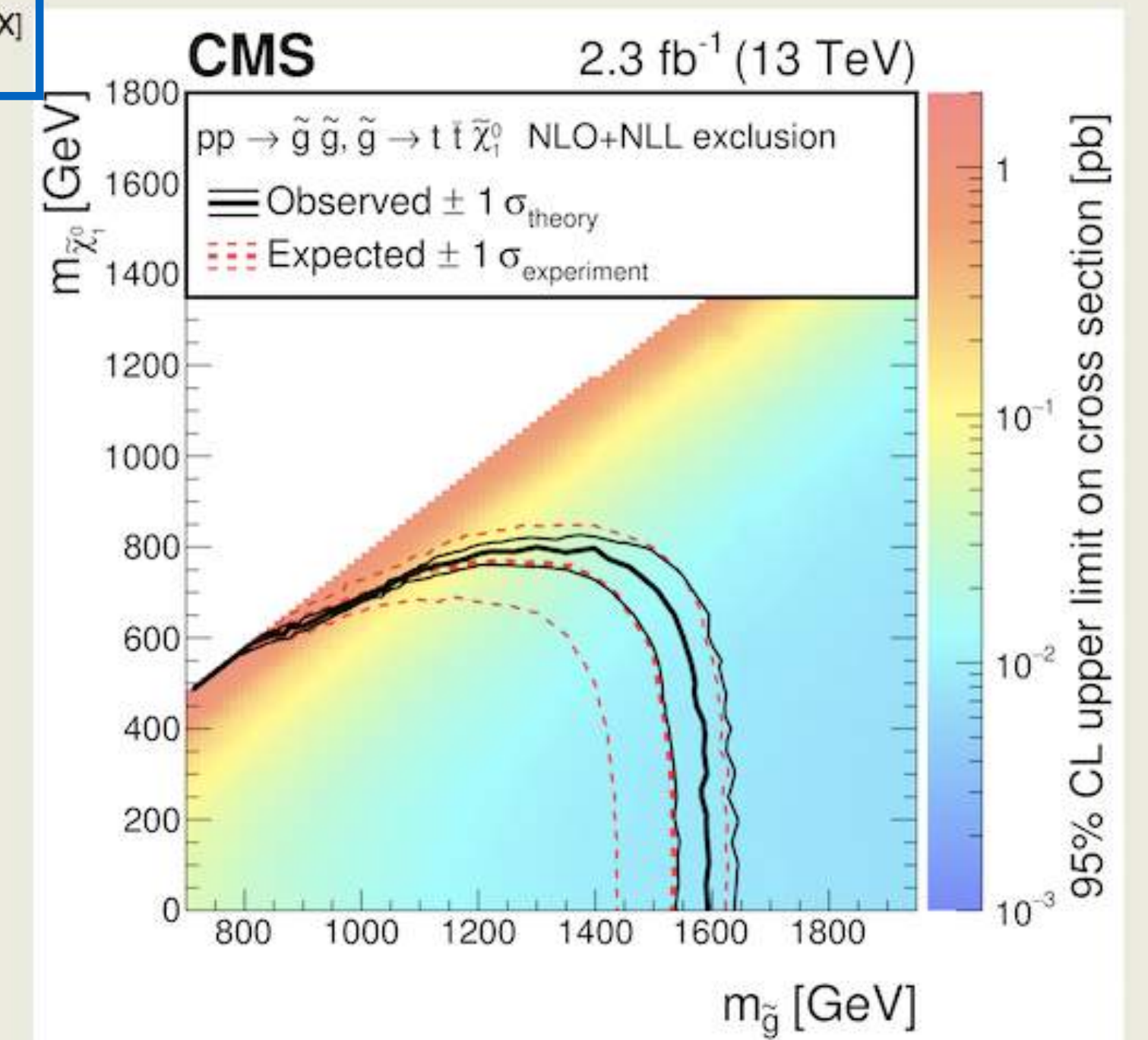
```
*qual: RE : P P --> GLUINO GLUINO X
*qual: SQRT(S) IN GEV : 13000.0
*yheader:  $m_{\tilde{\chi}^0_1}$ 
*xheader:  $m_{\tilde{g}}$ 
*data: x : y
712.5; 487.5;
797.1; 572.1;
.
.
```

Output on the webpage

Table 5 (Figure 10.) [HIDE DATA](#) or as: [input](#), [plain text](#), [AIDA](#), [PyROOT](#), [YODA](#), [ROOT](#), [mpl](#), [DMelt](#), [MarcXML](#) or [YAML](#)
Interpretation of results in the T1tttt model. The colored regions show the upper limits (95\% CL) on the production cross section for $pp \rightarrow \tilde{g}\tilde{g}, \tilde{g} \rightarrow t\bar{t}\tilde{\chi}^0_1$ in the $m_{\tilde{g}}-m_{\tilde{\chi}^0_1}$ plane. The curve shows the observed limit on the corresponding SUSY particle masses obtained by comparing the excluded cross section with theoretical cross sections.

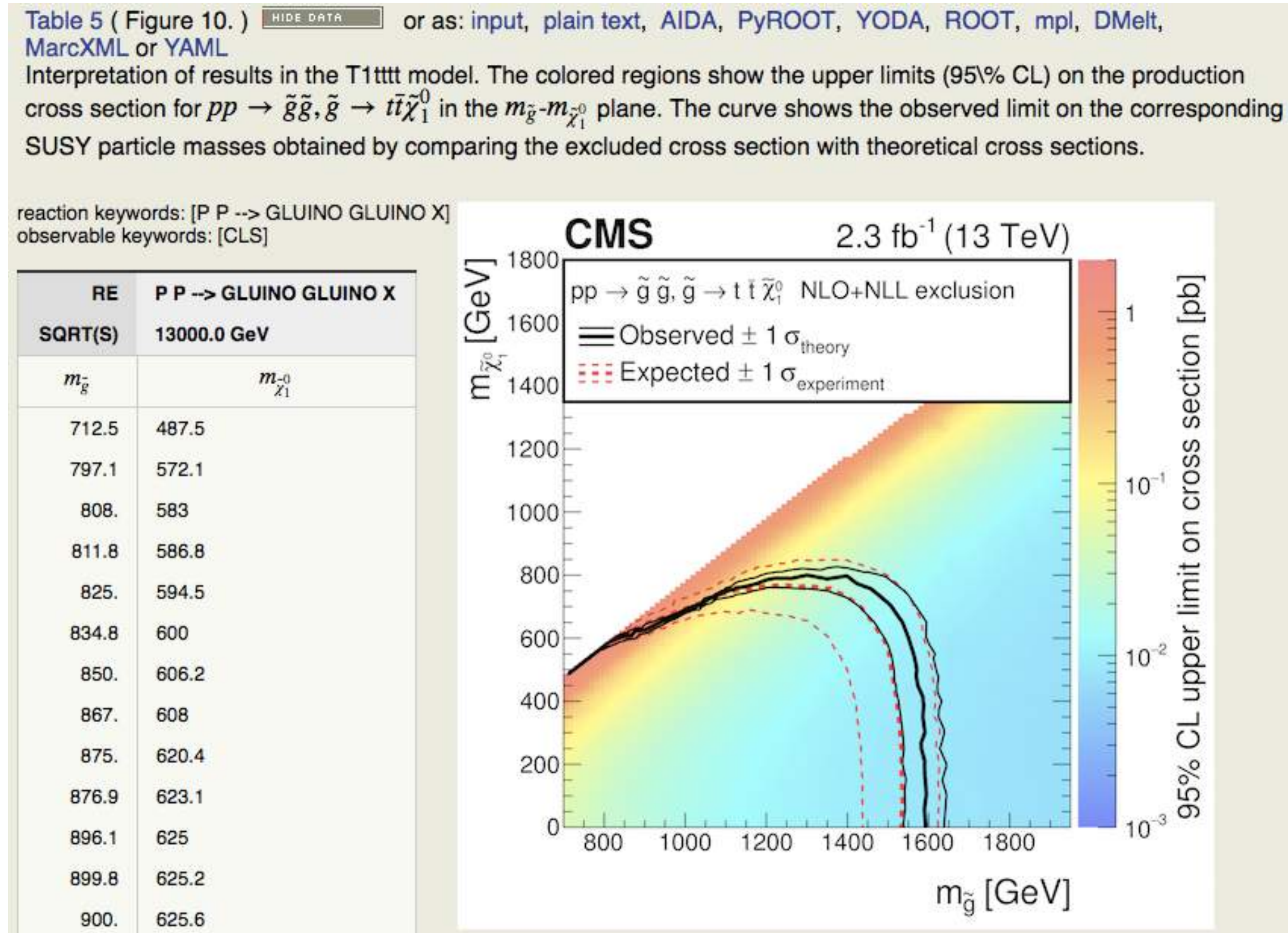
reaction keywords: [P P --> GLUINO GLUINO X]
observable keywords: [CLS]

RE	P P --> GLUINO GLUINO X
SQRT(S)	13000.0 GeV
$m_{\tilde{g}}$	$m_{\tilde{\chi}^0_1}$
712.5	487.5
797.1	572.1
808.	583
811.8	586.8
825.	594.5
834.8	600
850.	606.2
867.	608
875.	620.4
876.9	623.1
896.1	625
899.8	625.2
900.	625.6

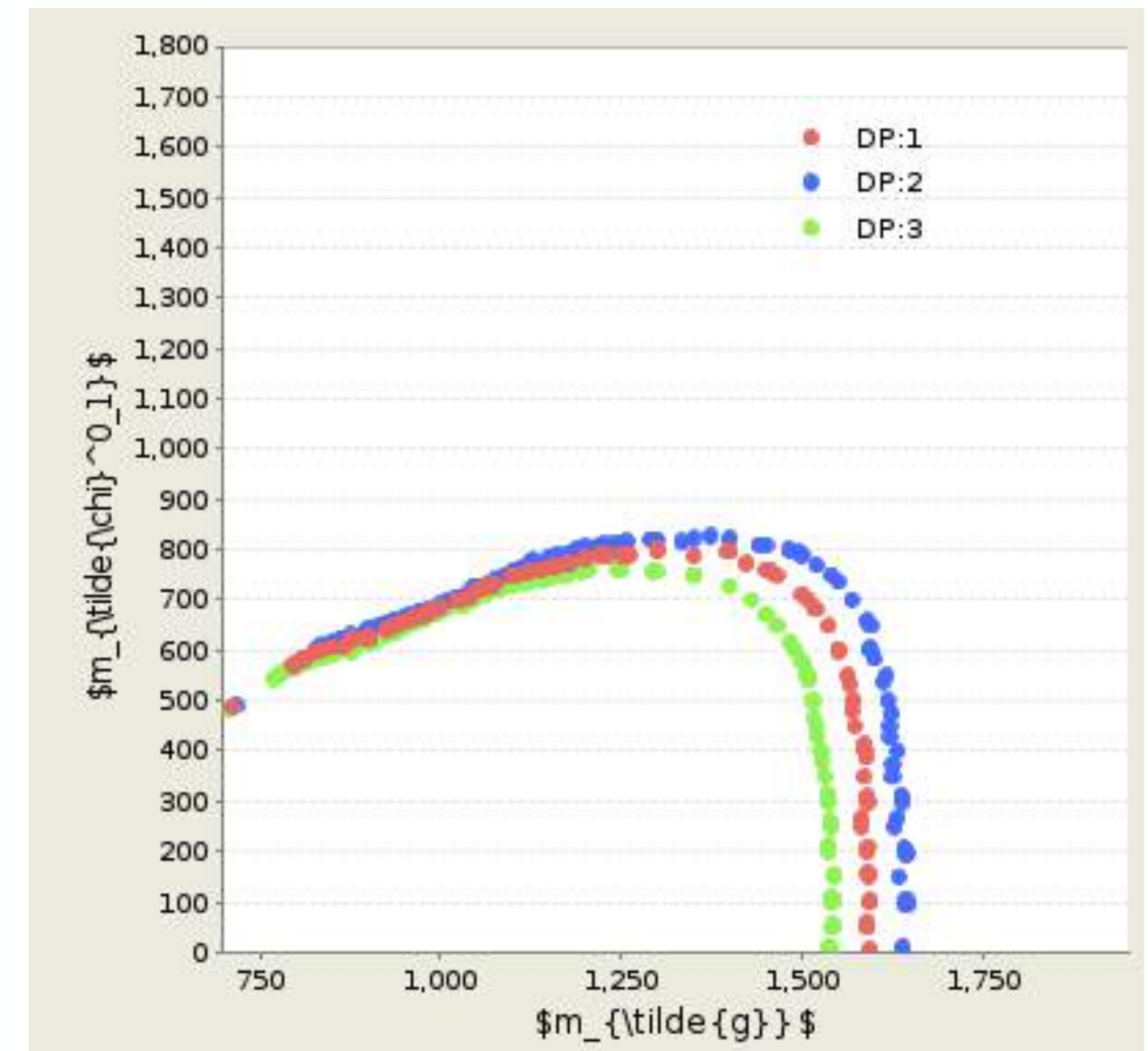


How to make sure that you did it correctly?

- Can convert table back to plot to find mistakes

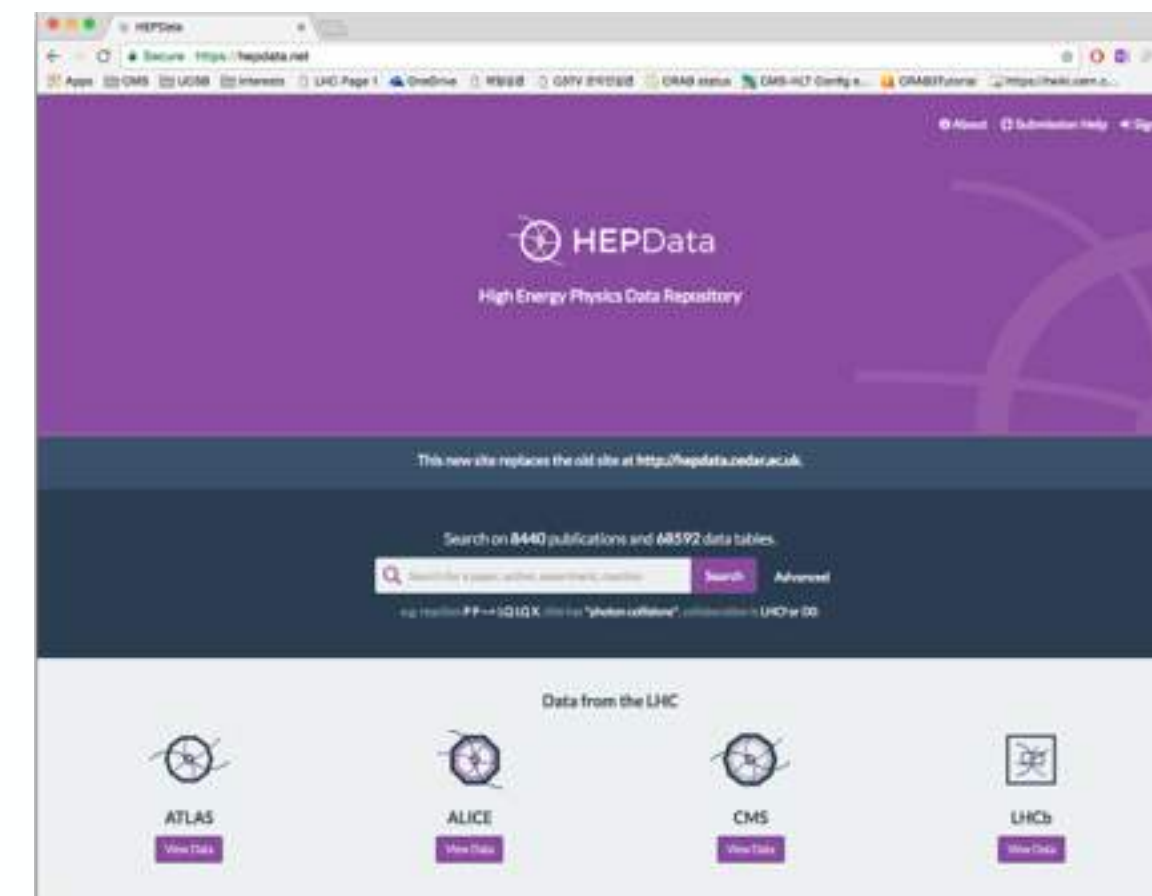
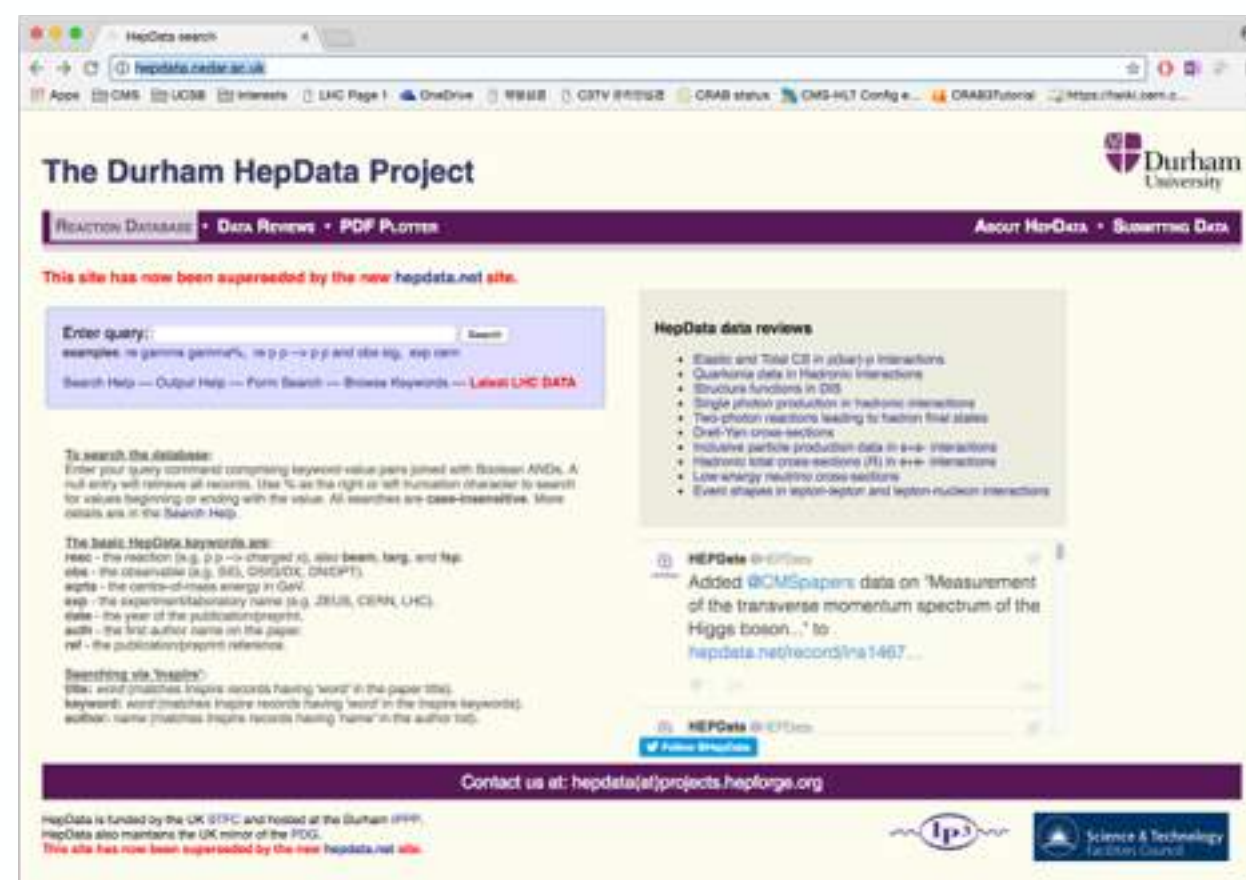


Observed limit curves



Things have changed

- The website changed since my experience
 - old page (<http://hepdata.cedar.ac.uk/>) to new page (<https://hepdata.net/>)



- Format of the input file changed from “Input format” to “YAML”
- The submission procedure changed

YAML (*/ˈjæməɪl/*, rhymes with *camel*) is a **human-readable data serialization language**. It stands for Yet Another Markup Language or YAML Ain't Markup Language (Wikipedia)

Change in input format: “input format” to “YAML”

YAML

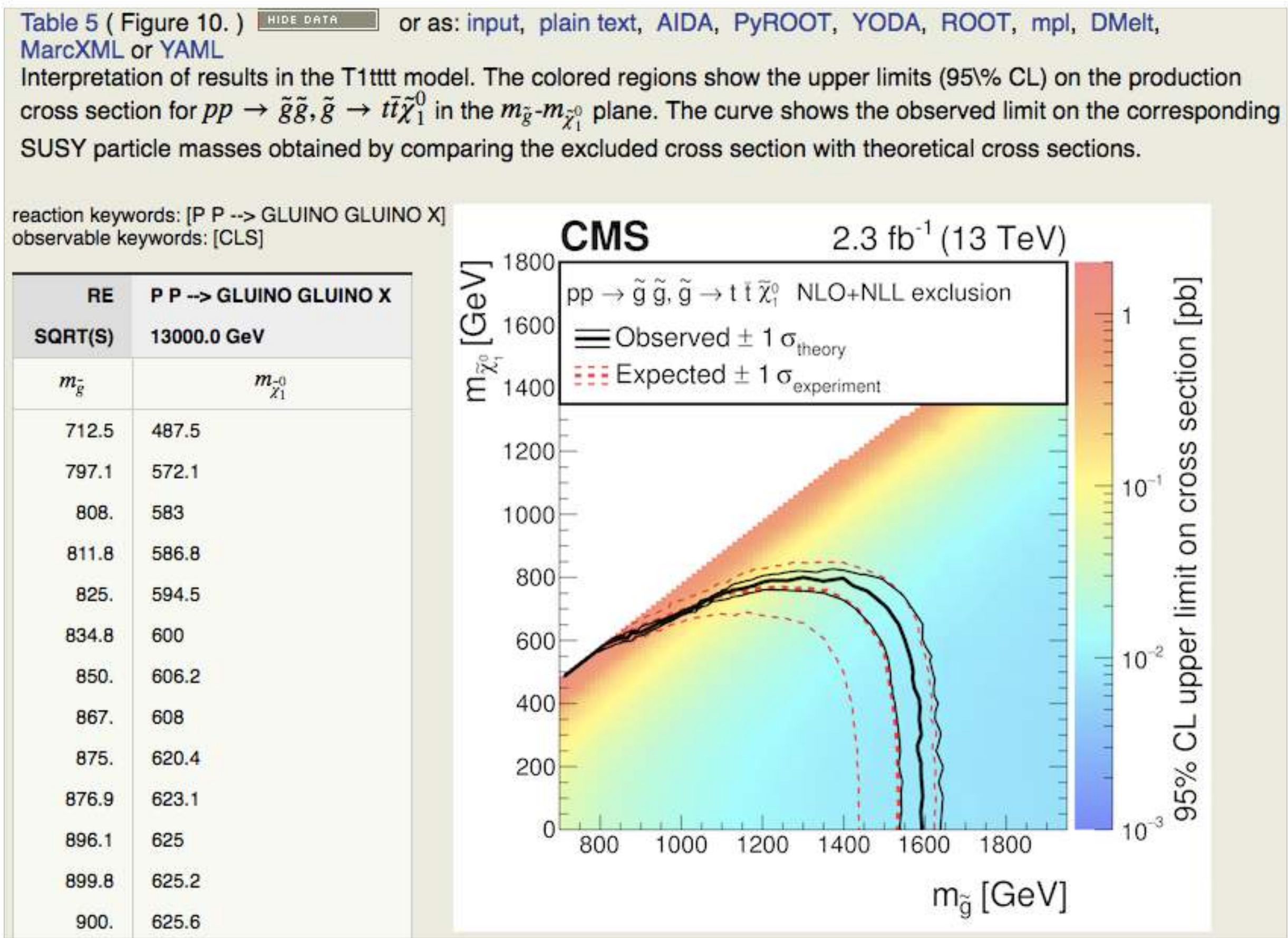
Input format

```
*dataset:
*location: Figure 10
*dscomment: Interpretation of results in the T1tttt model. The colored regions
show the upper limits (95\% CL) on the production cross section for $pp
\rightarrow \tilde{g}\tilde{g}, \tilde{g}\rightarrow t\bar{t}\tilde{\chi}^0_1$
in the $m_{\tilde{g}}$-$m_{\tilde{\chi}^0_1}$ plane. The curve shows the
observed limit on the corresponding SUSY particle masses obtained by comparing
the excluded cross section with theoretical cross sections.
*reackey: P P --> GLUINO GLUINO X
*obskey: CLS
*qual: RE : P P --> GLUINO GLUINO X
*qual: Sqrt(S) IN GeV : 13000.0
*yheader: $m_{\tilde{\chi}^0_1}$
*xheader: $m_{\tilde{g}}$
*data: x : y
712.5; 487.5;
797.1; 572.1;
.
.
```

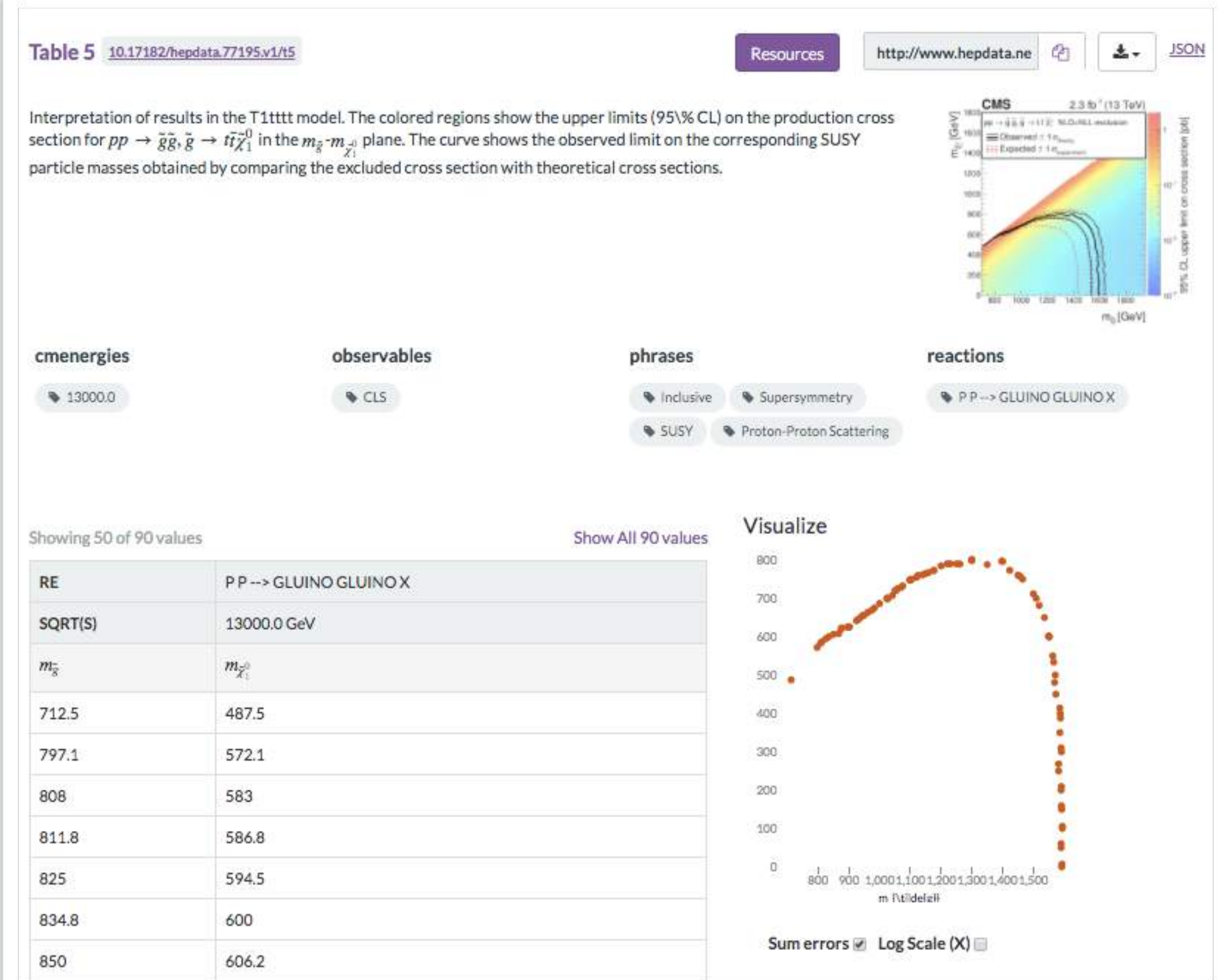
```
---
name: 'Table 5'
label: 'Data from Figure 10'
description: |
    Interpretation of results in the T1tttt model. The colored regions show the upper
    limits (95\% CL) on the production cross section for $pp\rightarrow \tilde{g}\tilde{g},
    \tilde{g}\rightarrow t\bar{t}\tilde{\chi}^0_1$ in the $m_{\tilde{g}}$-$m_{\tilde{\chi}^0_1}$
    plane. The curve shows the observed limit on the corresponding SUSY particle
    masses obtained by comparing the excluded cross section with theoretical cross
    sections.
keywords:
  - {name: reactions, values: ['P P --> GLUINO GLUINO X']}
  - {name: observables, values: ['CLS']}
  - {name: phrases, values: ['Inclusive', 'Supersymmetry', 'SUSY', 'Proton-Proton
    Scattering', 'Supersymmetry']}
  - {name: cmenergies, values: [13000.0]}
additional_resources:
  - {location: '/resource/1459054/figFigure10.png', description: 'Image file'}
  - {location: '/resource/1459054/thumb_figFigure10.png', description: 'Thumbnail image
    file'}
independent_variables:
  - header: {name: '$m_{\tilde{g}}$'}
    values:
      - {value: 712.5}
      - {value: 797.1}
      .
      .
dependent_variables:
  - header: {name: '$m_{\tilde{\chi}^0_1}$'}
    qualifiers:
      - {name: 'RE', value: 'P P --> GLUINO GLUINO X'}
      - {name: 'Sqrt(S)', value: '13000.0', units: 'GeV'}
    values:
      - value: 487.5
      - value: 572.1
```


Change of visualization (website)

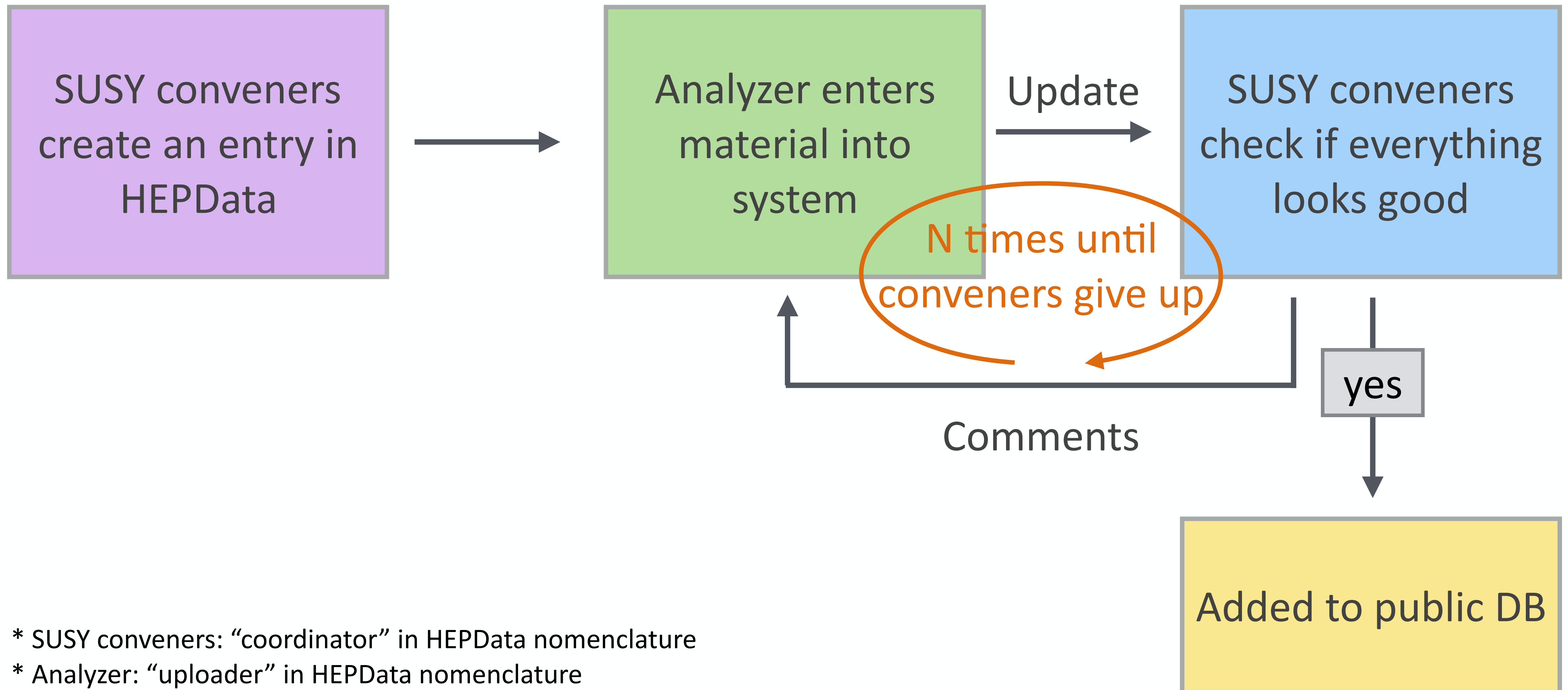
Old



New



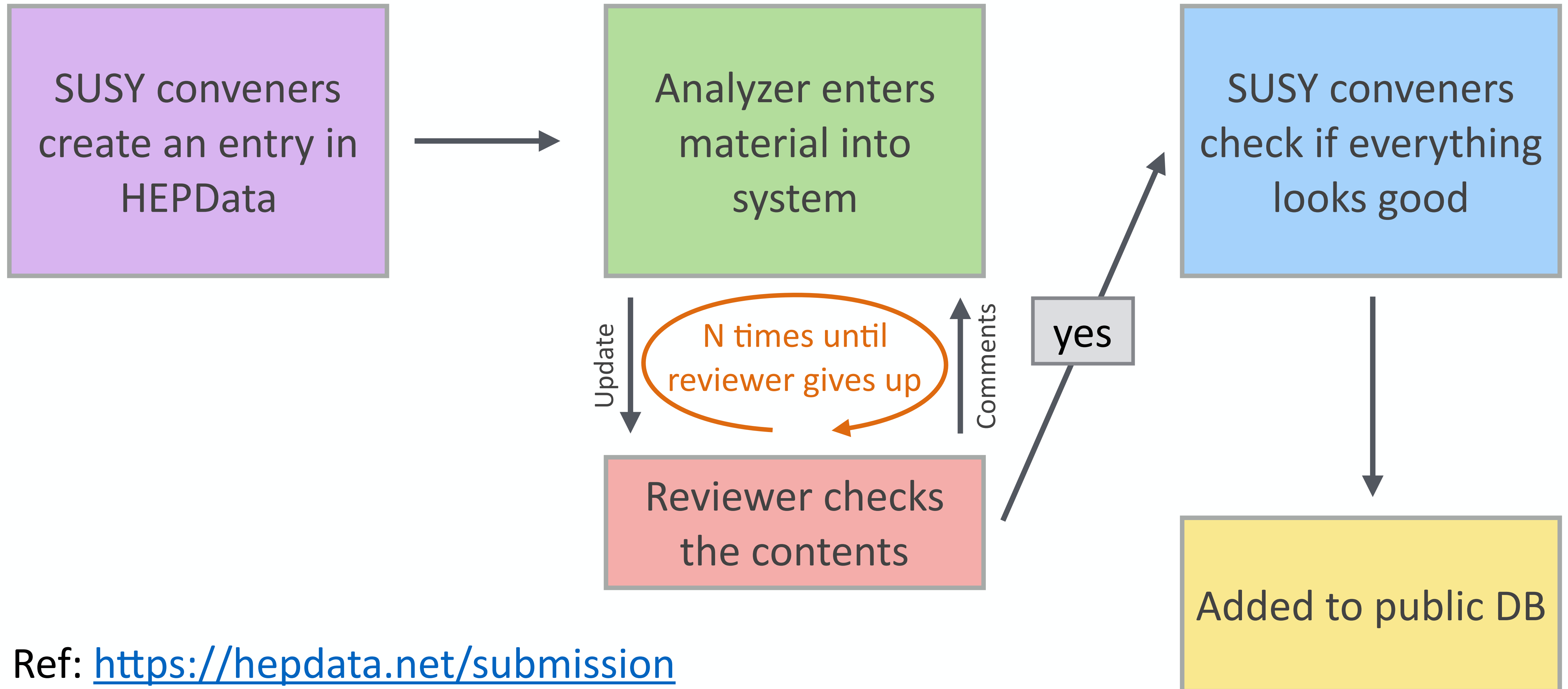
Old submission procedure



* SUSY conveners: “coordinator” in HEPData nomenclature

* Analyzer: “uploader” in HEPData nomenclature

New submission procedure



Ref: <https://hepdata.net/submission>

Summary

- I shared my experience with HEPData for SUS-15-007
- Things have changed since then, but I hope you've got some idea
- One advice
 - When you enter material for your analysis, look at existing entries to learn details: preferentially SUSY results from CMS or ATLAS
- Useful links
 - [HEPData webpage](#)
 - [Submission steps](#)
 - [Jim Olsen's talk on HEPData \(p3-5\)](#)