Reinterpretation of LHC results for new physics

Sabine Kraml

Atelier “Physique théorique des deux infinis” — 7/8 June 2021

LPSC GRENOBLE | MODANE
Theorie @ IN2P3 Les deux infinis

LHC exploration of the TeV scale

Théorie @ IN2P3

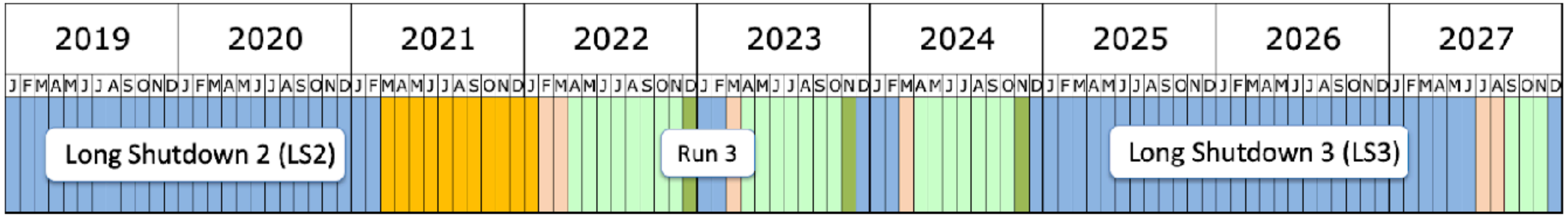
- The LHC is our energy-frontier machine for the foreseeable future —immense efforts and costs—
- To fully exploit the LHC physics potential, we need detailed phenomenological studies ***and*** we to make sure that the LHC results can be used by the whole HEP community now and in the future ⇒ close experiment-theory interaction

We are here; many Run 2 results still to come out

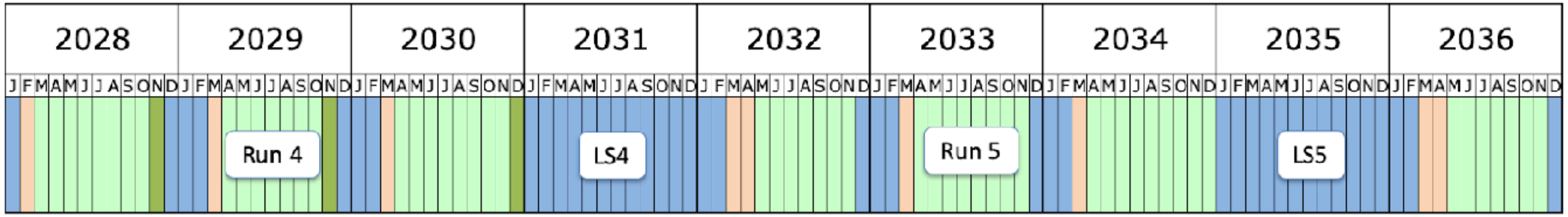
next ~6 years: Run 3 data taking and analysis; lots more interesting physics results

2009–2013
Run 1 @ 7-8 TeV

2015–2018
Run 2 @ 13 TeV



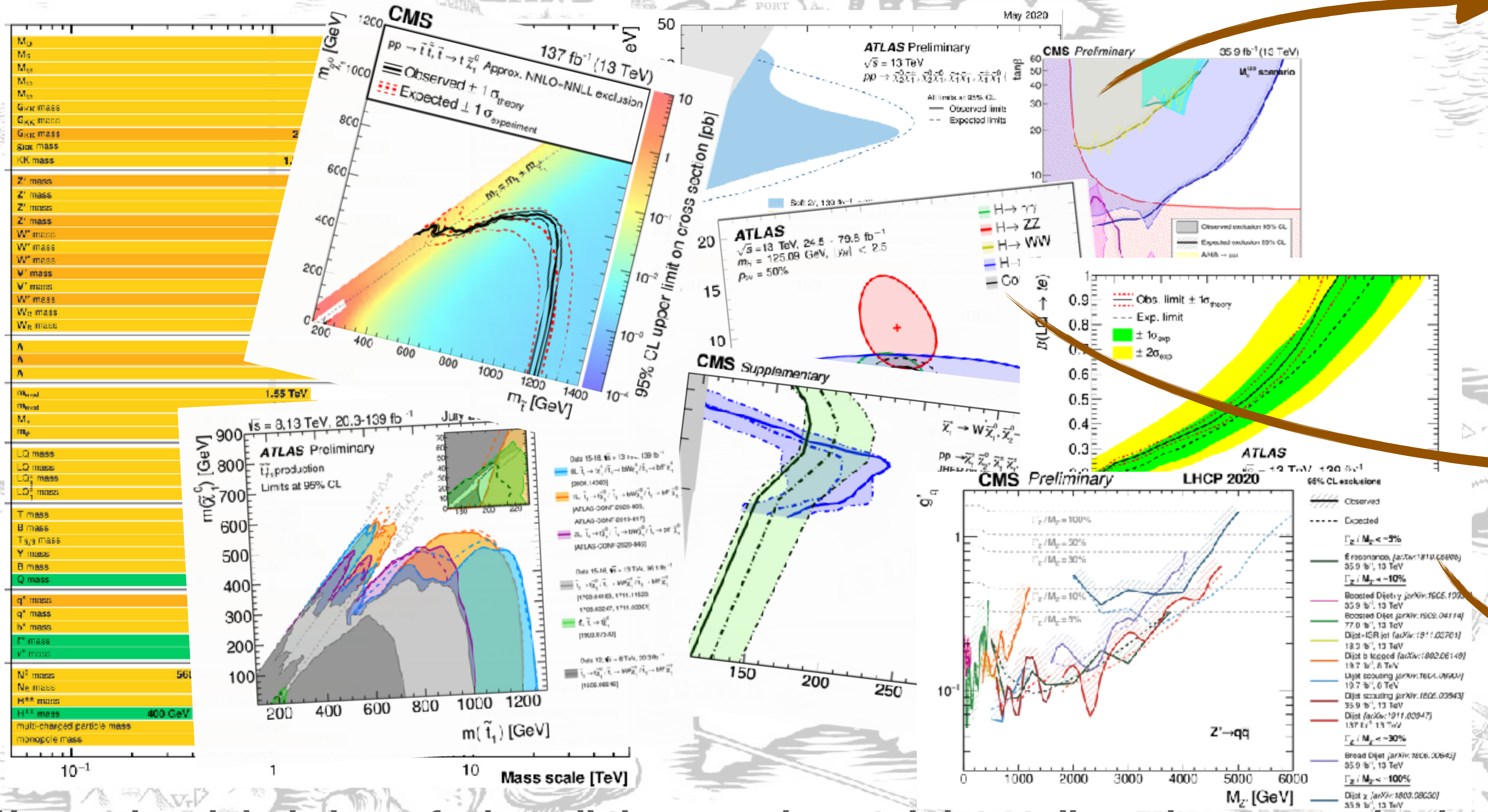
beyond: LHC Phase II (high-luminosity)



The challenge

Théorie @ IN2P3

Precise measurements of SM processes and direct searches for new physics in a vast variety of channels. Despite the multitude of BSM scenarios tested this way by the LHC experiments, it still constitutes only a small subset of the possible theories and parameter combinations to which the experiments are sensitive.



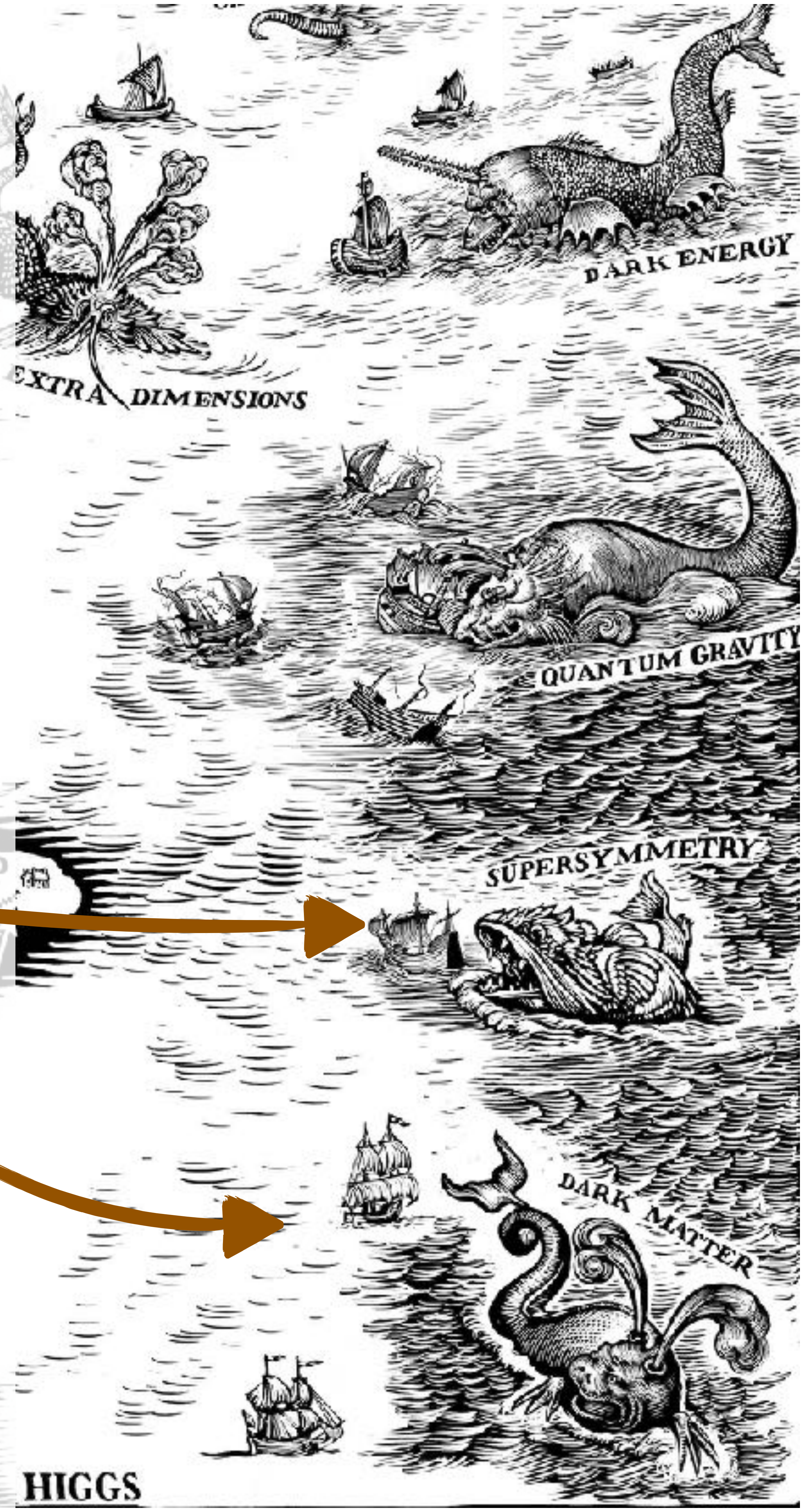


Illustration (c) C.Wormell from "A Map of the Invisible" by J.Butterworth

We need a global view of what all the experimental data tell us about new physics.

Reinterpretation methods and public tools

Théorie @ IN2P3

Use simplified-model results to constrain full BSM models

Public analysis database

Reproduce exp. analyses in Monte Carlo event simulation

Combined likelihood analysis from Higgs measurements

CONTUR method

Use SM precision measurements (fiducial differential distributions) to constrain new physics

Precision calculations → MC simulation, SM-BSM interference effects

LHC Reinterpretation Forum

280 participants at 6th workshop in Feb 2021

Global analysis based on statistical learning

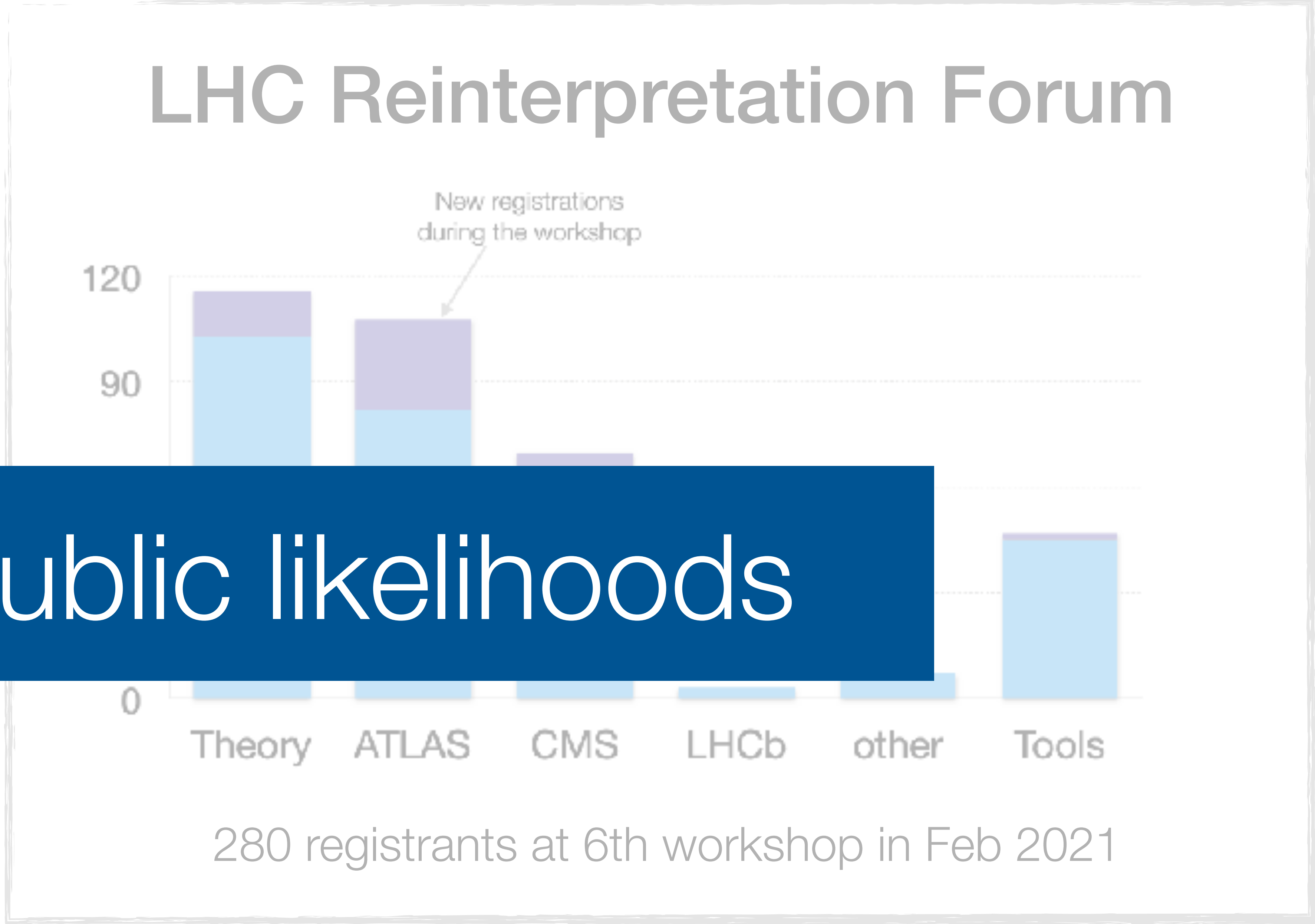
Here: highlighting just few aspects

Use simplified-model results to constrain full BSM models — SModelS

MadAnalysis 5 — Public analysis database — Reproduce exp. analyses in Monte Carlo event simulation

Combined likelihood analysis from Higgs measurements — Lilith

CONTUR method — Use SM precision measurements (fiducial differential distributions) to constrain new physics



Public likelihoods

PROTOMODEL BUILDER — Global analysis based on statistical learning

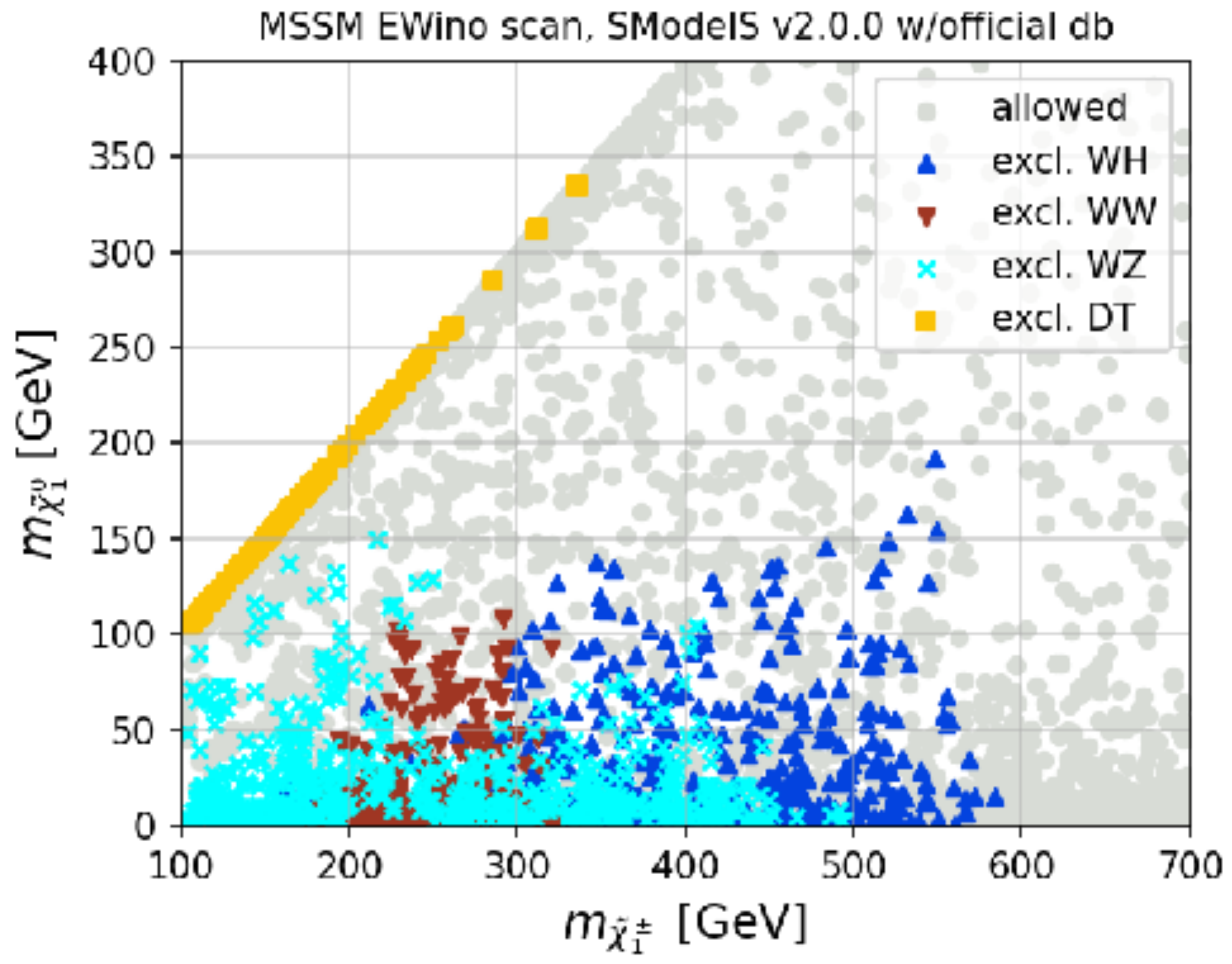
SModelS

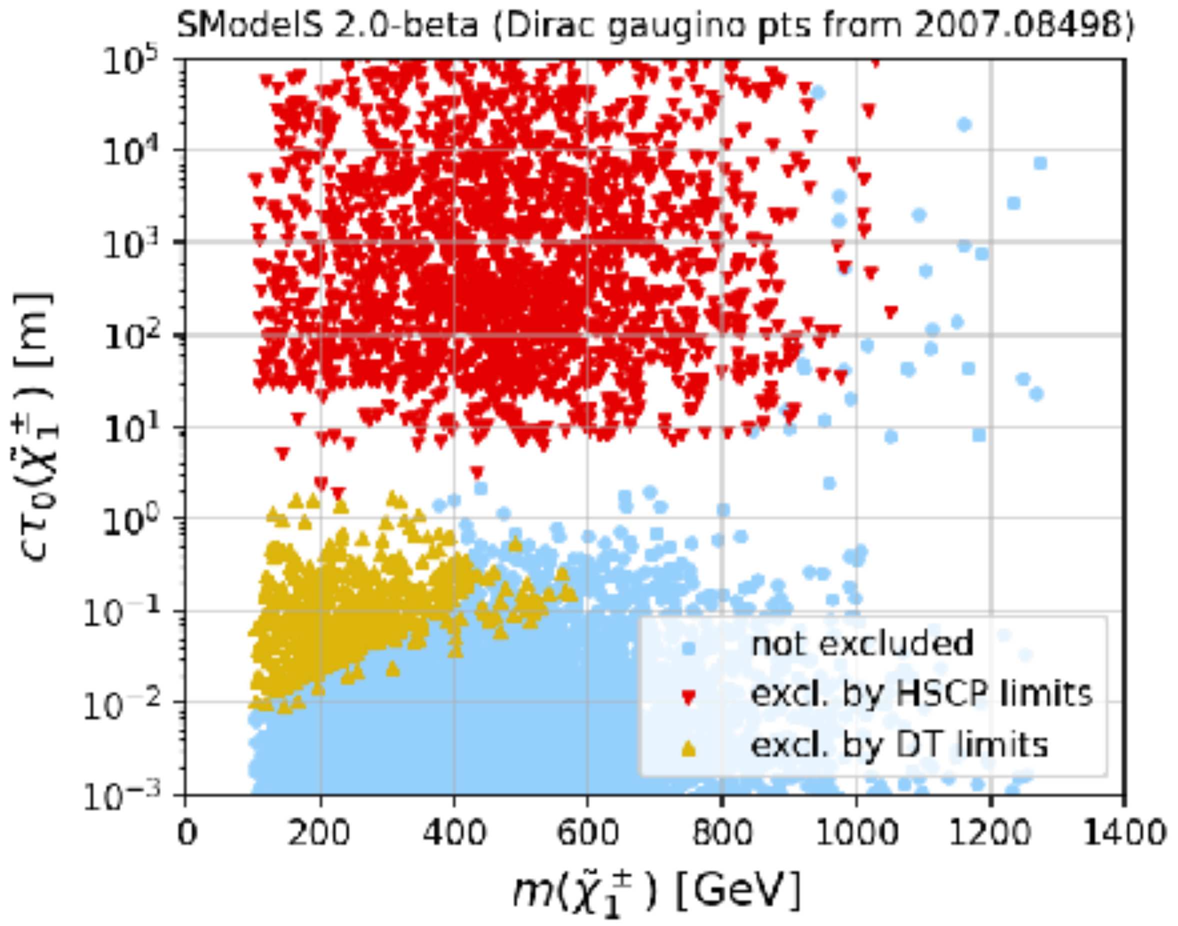
smodels.github.io

Théorie @ IN2P3

Public tool for the automatised interpretation of simplified-model results from the LHC

- Large database of experimental results (currently 100 ATLAS & CMS searches).
- Very fast b/c no need for MC simulation; suitable for large scans and model surveys.
- Can reinterpret ML-based analyses and can treat prompt and long-lived searches on the same footing.
- So far limited to models featuring a Z_2-like symmetry (SUSY-like signals).





(Near) future: machine-learn the database!
Necessary because LLP results have an additional dimension (the decay length), which makes the database very "heavy".

Longer term: generalisation to arbitrary signal topologies, beyond SUSY-like signatures with 2 branches. Needs new internal "language"; complete refactoring of the code and database !

Complicated by the fact that simplified-model assumptions, which are currently exploited won't hold any more.

Protomodel builder

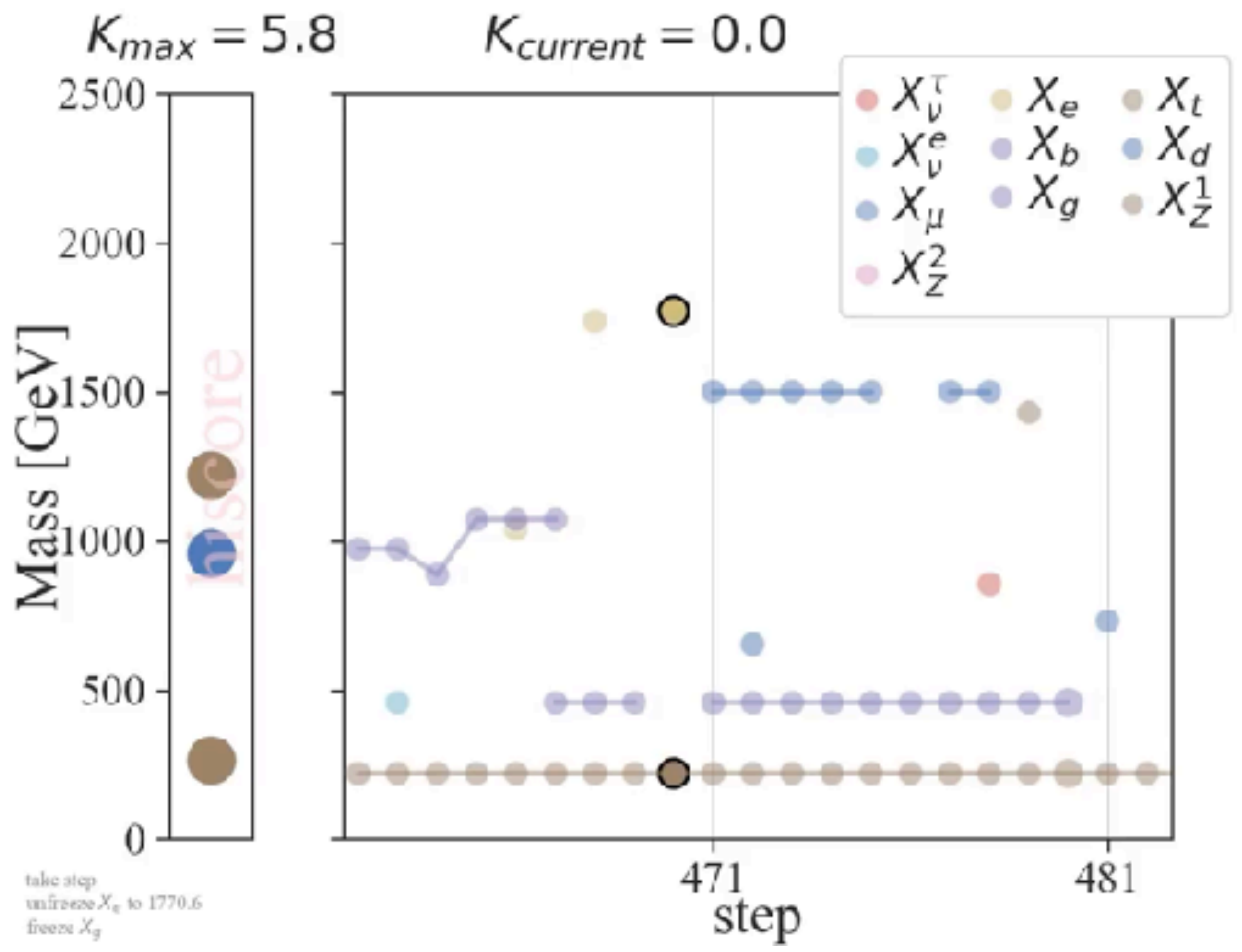
arXiv:2012.12246

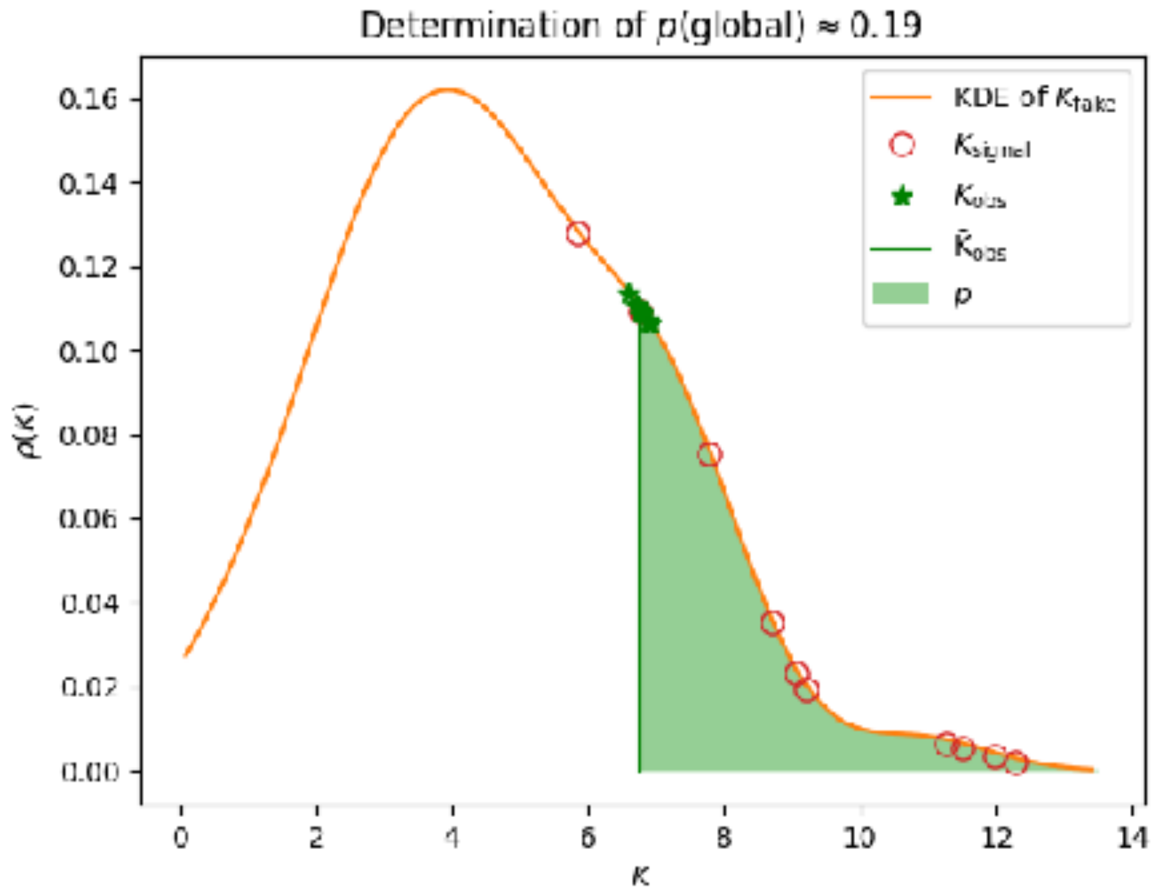
Théorie @ IN2P3

- The LHC currently has no clear sign of new physics; nonetheless there may be dispersed signals* hiding in the slew of data.

  * effects of new particles which are spread out over several search regions or final states

- Developed a novel statistical learning algorithm for identifying potential dispersed signals in the published LHC results while remaining in agreement with constraints from null results.
- Exploits SModelS framework and database.
- Want to develop this into a true global analysis tool (lots of work ahead)
  - PhD grant from IDEX Univ. Grenoble-Alpes
  - ANR PRCI proposal under review
  - BSMGA master project 2020-2022
- Ultimate goal is a complete bottom-up approach to inferring the BSM Lagrangian from the data.





Problem of "relating incomplete data and incomplete theory", see e.g. Binetruy et al, hep-ph/0312248

Public likelihoods

- The statistical model of an experimental analysis provides the complete mathematical description of that analysis

  $p(o|\alpha)$ relating the observed quantities o to the parameters α

- Given the likelihood, all the standard statistical approaches are available for extracting information from it

- Essential information for any detailed interpretation of experimental results

  = determining the compatibility of the observations with theoretical predictions

Les Houches Recommandations (2012)

3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.

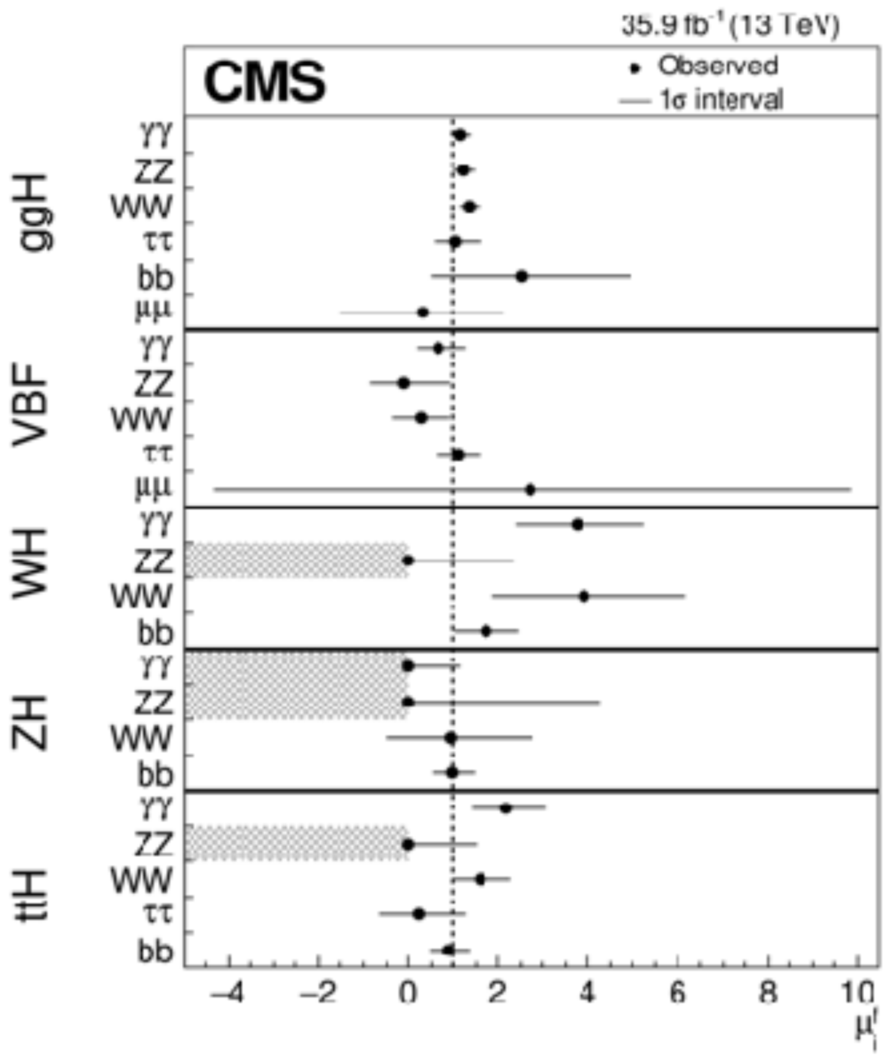
3c: Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.

arXiv:1203.2489

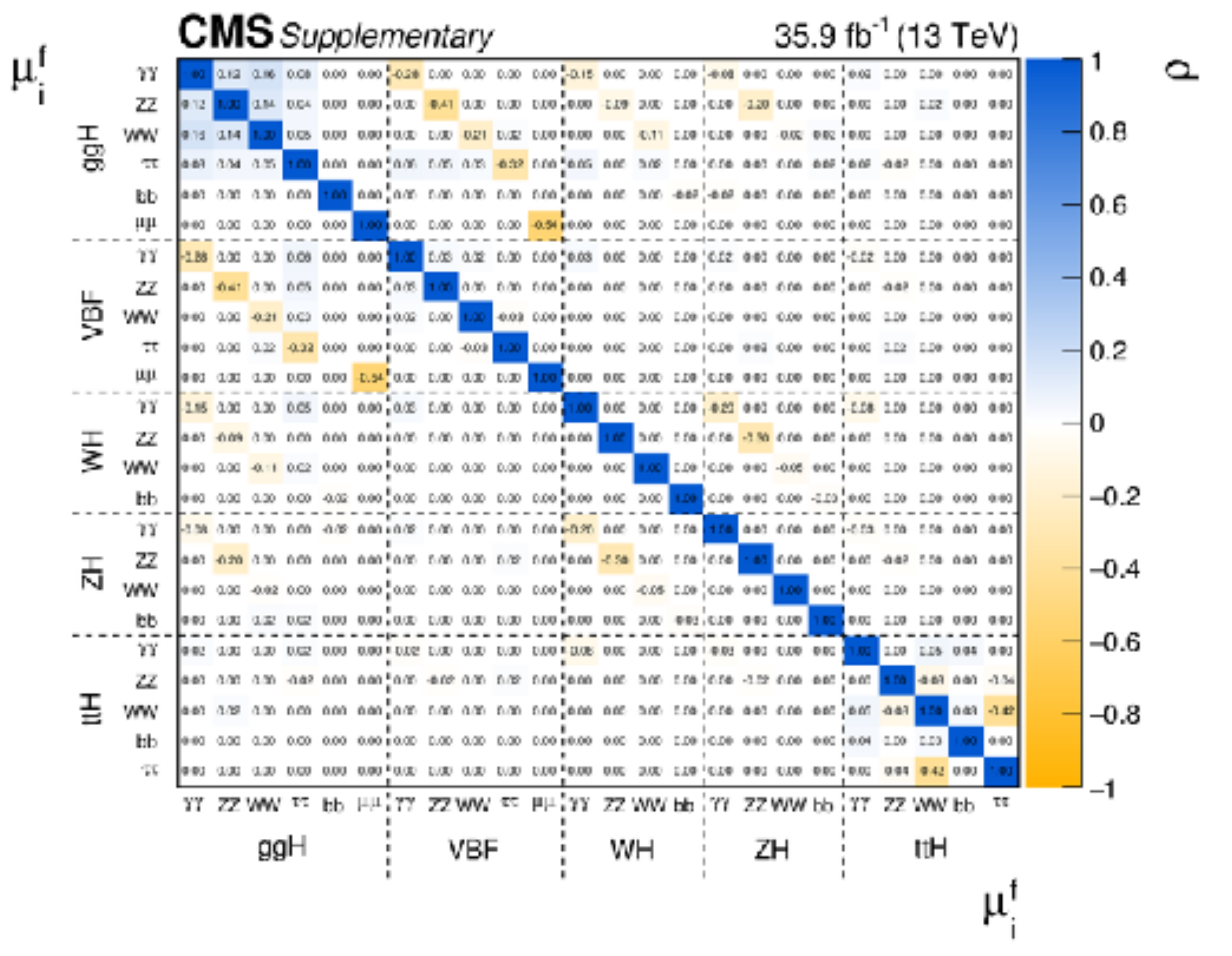
NB the ability to construct a global likelihood is key to the statistical learning procedure of the Protomodel Builder !

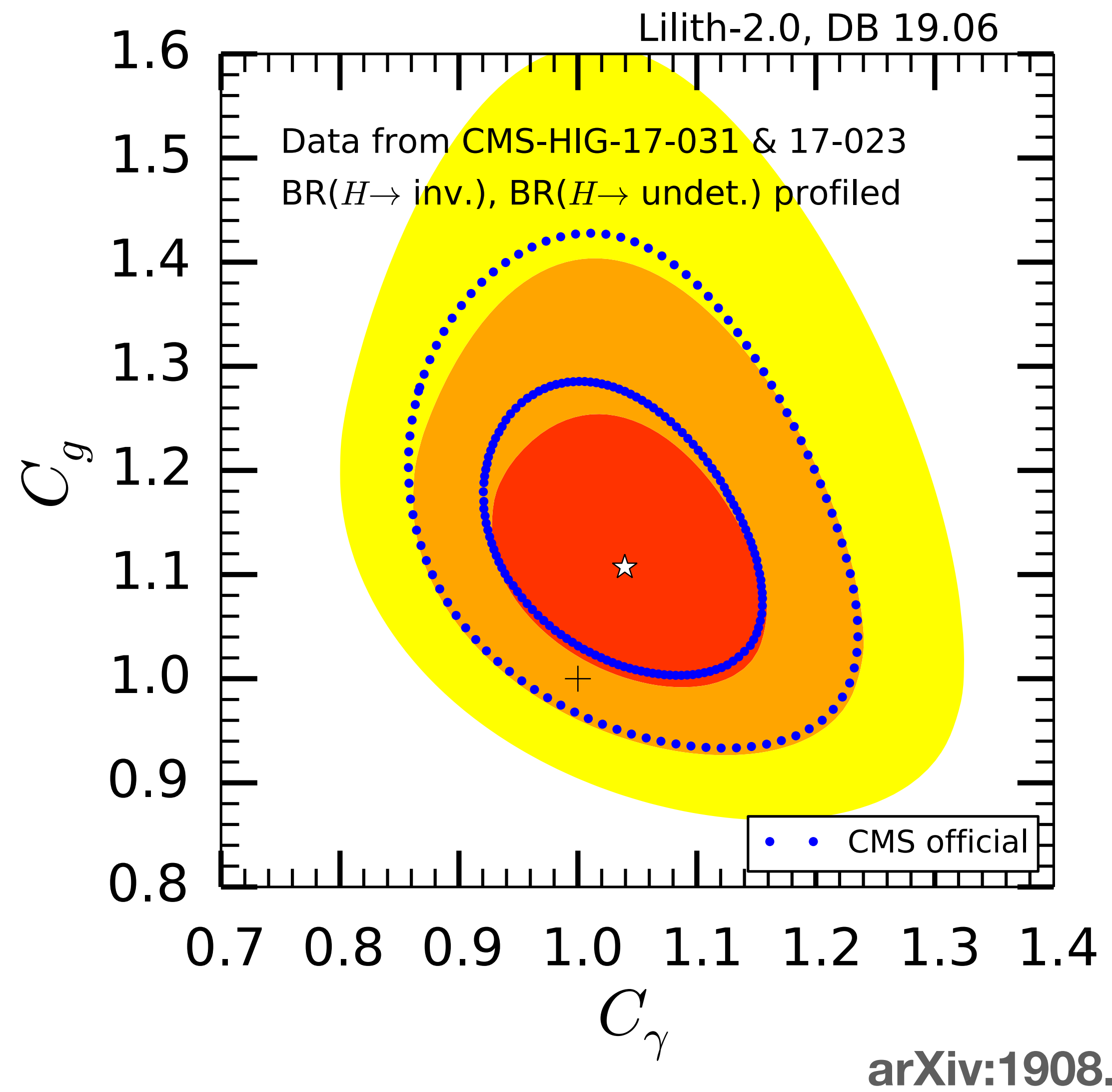
So far: O±δO plus correlations (sometimes)

- Simplified likelihood, Gaussian approximation
  - e.g., Higgs measurements, channel-by-channel correlation matrix



CMS-HIG-17-031





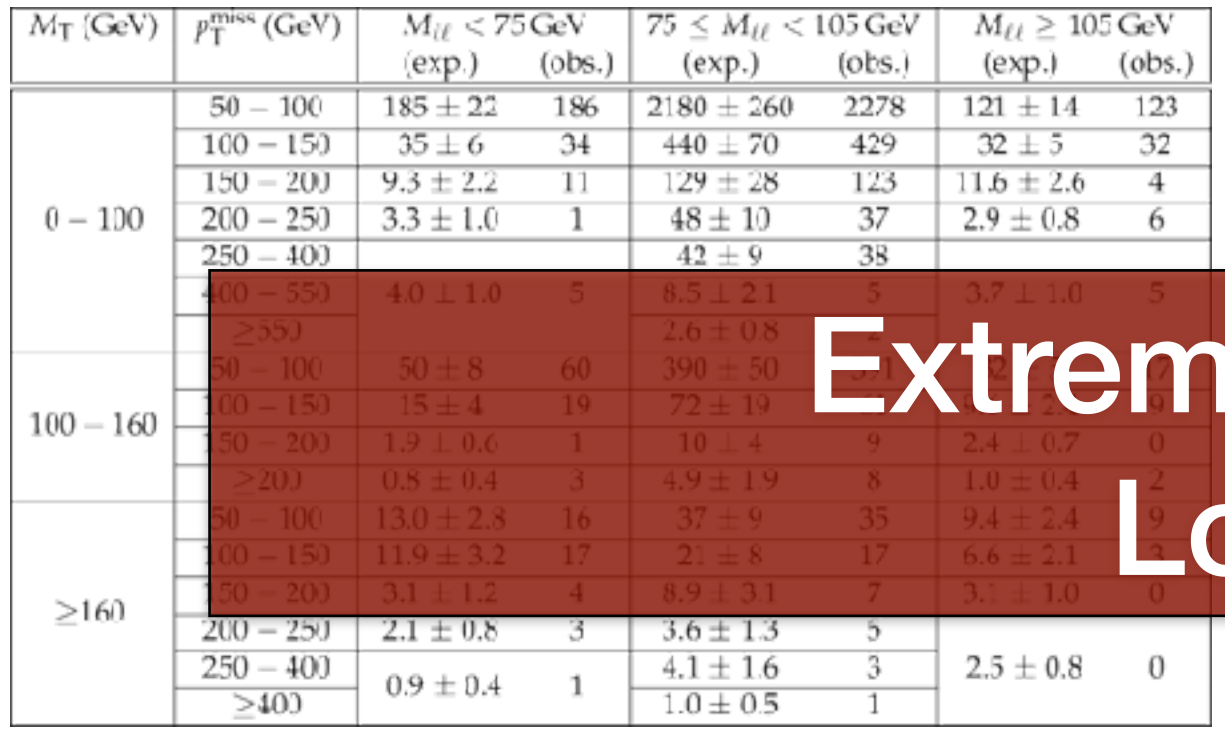
arXiv:1908.03952

«Correlation data […] has proven excellent for stabilising and ensuring better statistical definition in global fits as well as avoiding either overly conservative or over-enthusiastic interpretations.»

Reinterpretation Forum Report, 2003.07868

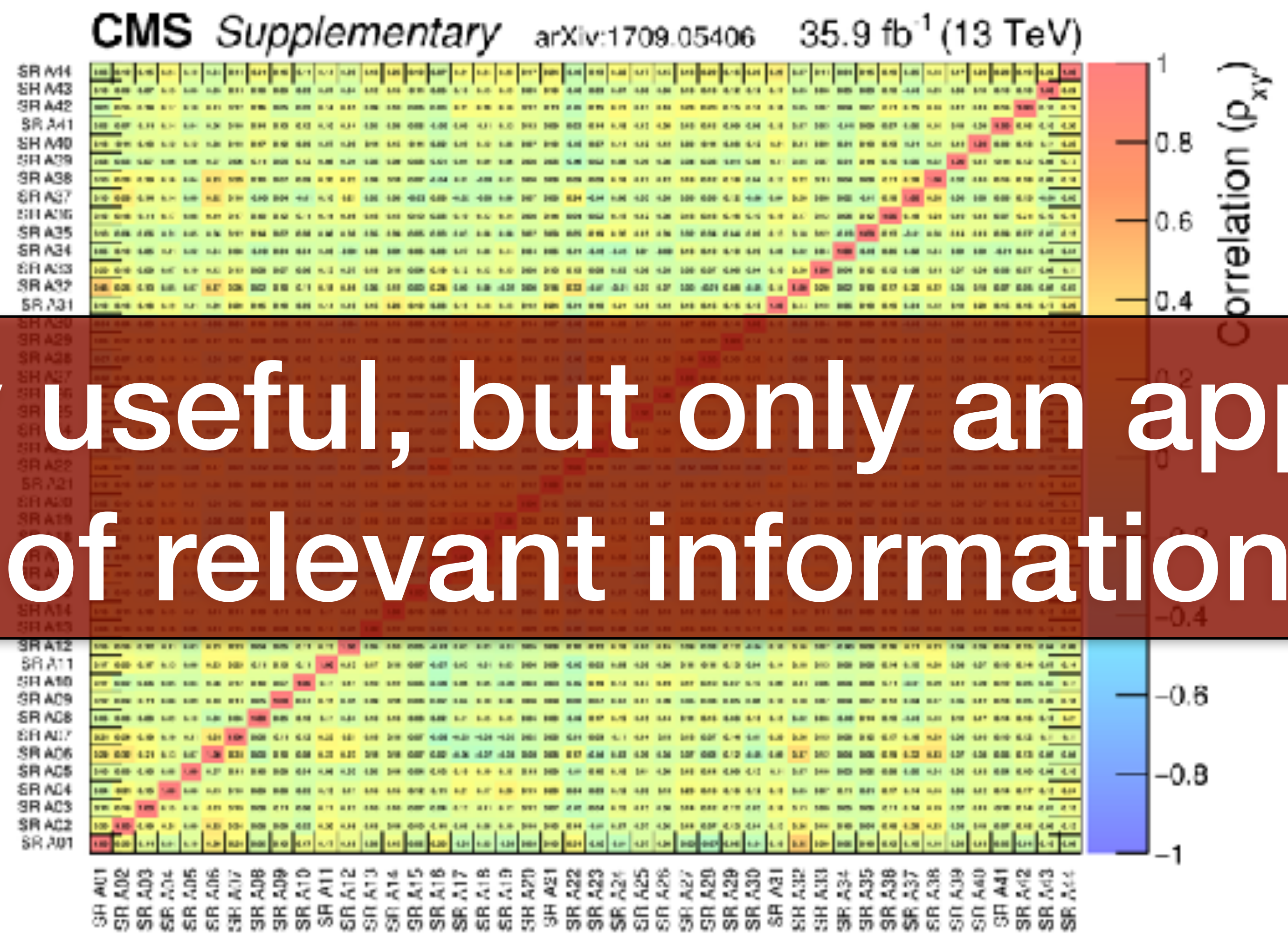
So far: O±δO plus correlations (sometimes)

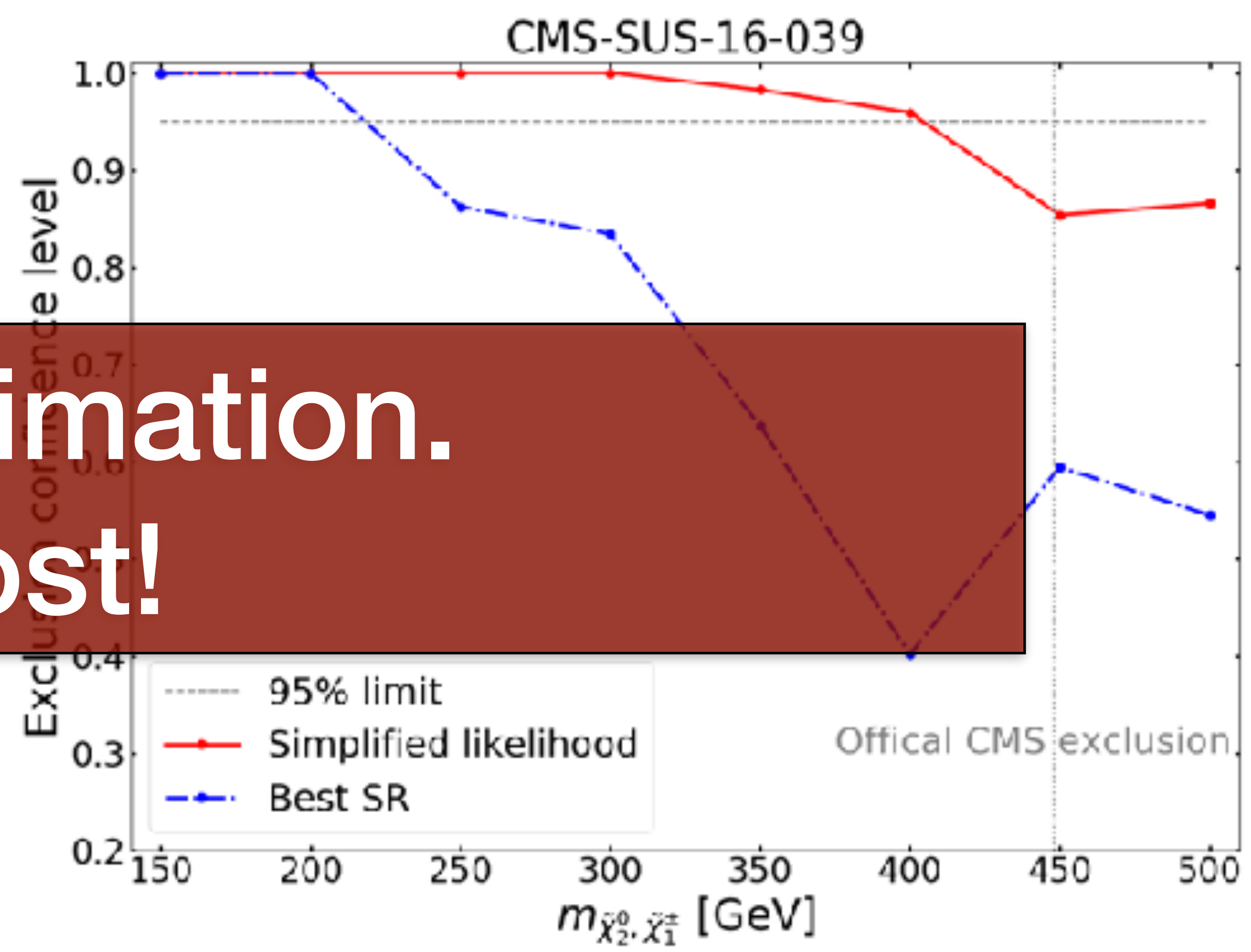
- Simplified likelihood, Gaussian approximation
  - CMS SUSY group: covariance matrices for combination of signal regions



M_T (GeV)	p_T^{miss} (GeV)	$M_{\ell\ell} < 75$ GeV (exp.)	(obs.)	$75 \leq M_{\ell\ell} < 105$ GeV (exp.)	(obs.)	$M_{\ell\ell} \geq 105$ GeV (exp.)	(obs.)
0 – 100	50 – 100	185 ± 22	186	2180 ± 260	2278	121 ± 14	123
	100 – 150	35 ± 6	34	440 ± 70	429	32 ± 5	32
	150 – 200	9.3 ± 2.2	11	129 ± 28	123	11.6 ± 2.6	4
	200 – 250	3.3 ± 1.0	1	48 ± 10	37	2.9 ± 0.8	6
	250 – 400			42 ± 9	38		
	400 – 550	4.0 ± 1.0	5	8.5 ± 2.1	5	3.7 ± 1.0	5
	≥550			2.6 ± 0.8	2		
100 – 160	50 – 100	50 ± 8	60	390 ± 50	[illegible]	[illegible]	[illegible]
	100 – 150	15 ± 4	19	72 ± 19	[illegible]	[illegible]	[illegible]
	150 – 200	1.9 ± 0.6	1	10 ± 4	9	2.4 ± 0.7	0
	≥200	0.8 ± 0.4	3	4.9 ± 1.9	8	1.0 ± 0.4	2
≥160	50 – 100	13.0 ± 2.8	16	37 ± 9	35	9.4 ± 2.4	9
	100 – 150	11.9 ± 3.2	17	21 ± 8	17	6.6 ± 2.1	[illegible]
	150 – 200	3.1 ± 1.2	4	8.9 ± 3.1	7	3.1 ± 1.0	0
	200 – 250	2.1 ± 0.8	3	3.6 ± 1.3	5		
	250 – 400	0.9 ± 0.4	1	4.1 ± 1.6	3	2.5 ± 0.8	0
	≥400			1.0 ± 0.5	1		

CMS-SUS-16-039 (EWino search)





Extremely useful, but only an approximation.
Lots of relevant information is lost!

Contribution 15, LH 2019 BSM WG report
arXiv:2002.12220

«Correlation data […] has proven excellent for stabilising and ensuring better statistical definition in global fits as well as avoiding either overly conservative or over-enthusiastic interpretations.»

Reinterpretation Forum Report, 2003.07868

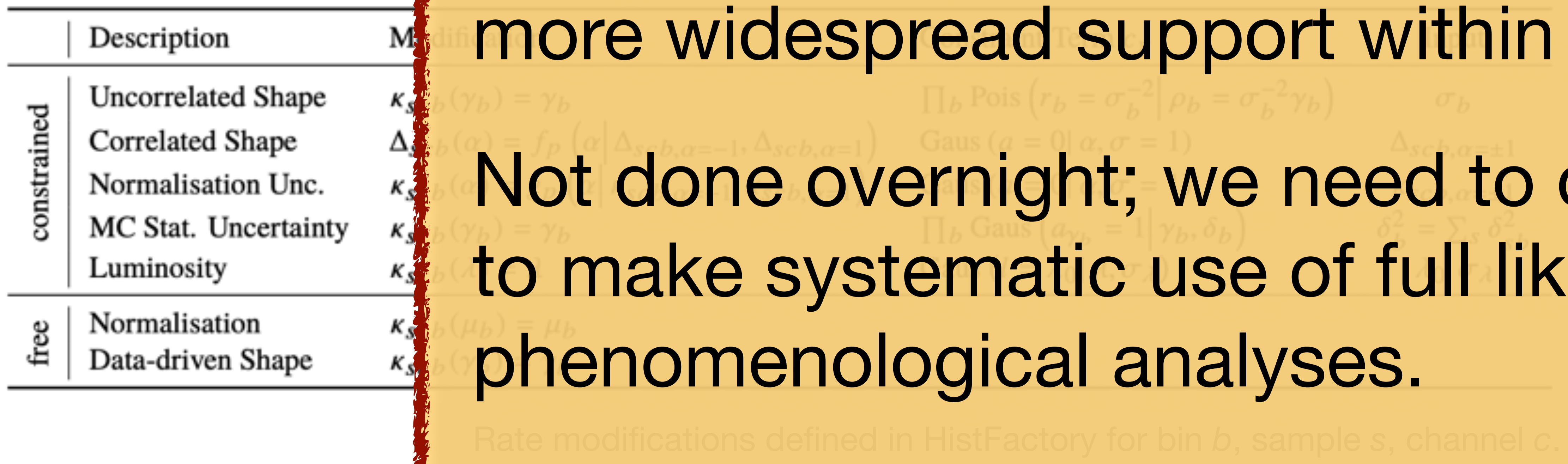
First full likelihoods from ATLAS

Théorie @ IN2P3

- Plain-text serialisation of HistFactory workspaces, JSON format

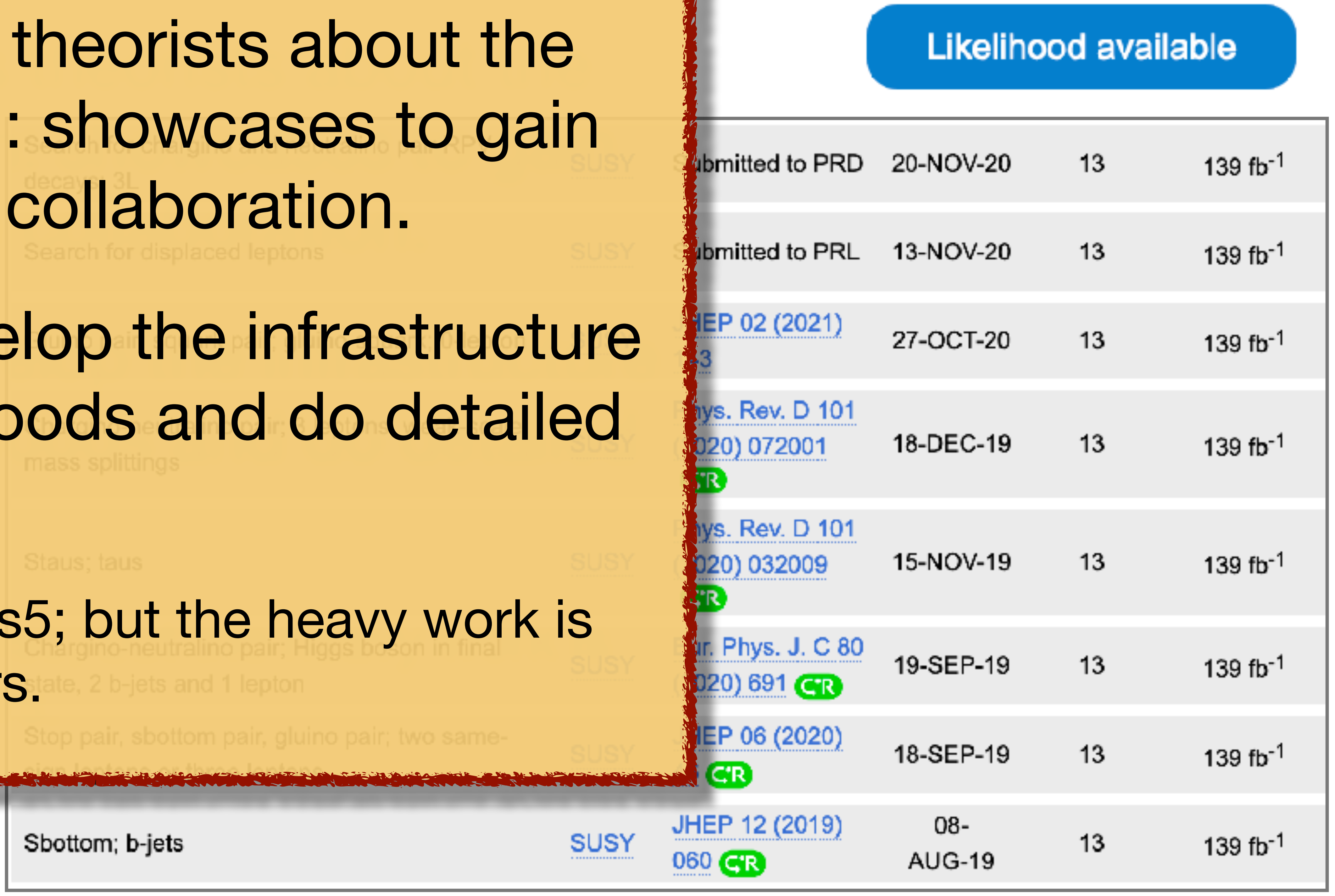
ATL-PHYS-PUB-2019-029

  - Provides background estimates, changes under systematic variations, and observed data counts at the same fidelity as used in the experiment.



  - Usage: RooFit, pyhf

  - Target: long-term data/analysis preservation, reinterpretation purposes



ATLAS now wants feedback from us theorists about the usage and impact of this information: showcases to gain more widespread support within the collaboration.

Not done overnight; we need to develop the infrastructure to make systematic use of full likelihoods and do detailed phenomenological analyses.

Done: interfaces to SModelS and MadAnalysis5; but the heavy work is only starting; timescale is probably a few years.

Théorie @ IN2P3 Les deux infinis

We want public likelihoods to become the standard!

Different experiments/analyses use different modelling specifications, but, whatever the choice, the full statistical model should be made public.

Would enormously benefit the short- and long-term reuse of experimental results

- especially for global interpretations
- reassessment of systematics (theoretical uncertainties, PDFs …)

A white paper on public likelihoods

Théorie @ IN2P3

Chief editors: Kyle Cranmer, SK, Harrison Prosper

Objectives:

- to state as precisely as possible
  - the role of statistical models and likelihoods
  - why it's important to preserve -and publish- them
  - what issues need to be addressed, and
- to urge our field to reach consensus to make them openly available (in a standard form?).

Timeline: soon, next few months

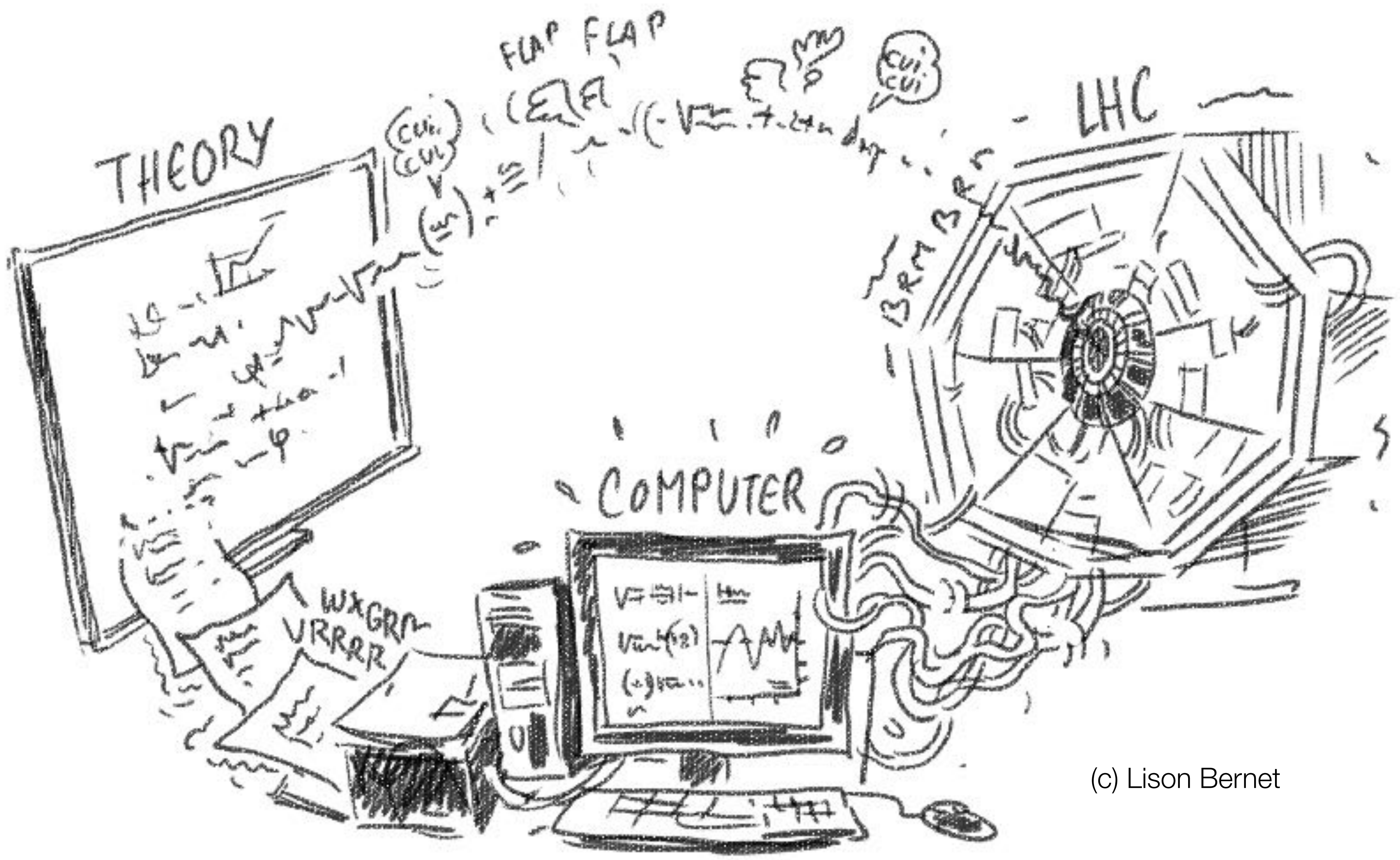
Effort started this April; attracts a lot of interest also beyond the LHC community (Belle-II, dark matter direct detection)

to be followed by a series of dedicated workshops

Conclusions

Théorie @ IN2P3

Analysis preservation and reuse is a very active topic — originated from theory side but now increasing interest also in the experimental collaborations themselves.*



(c) Lison Bernet

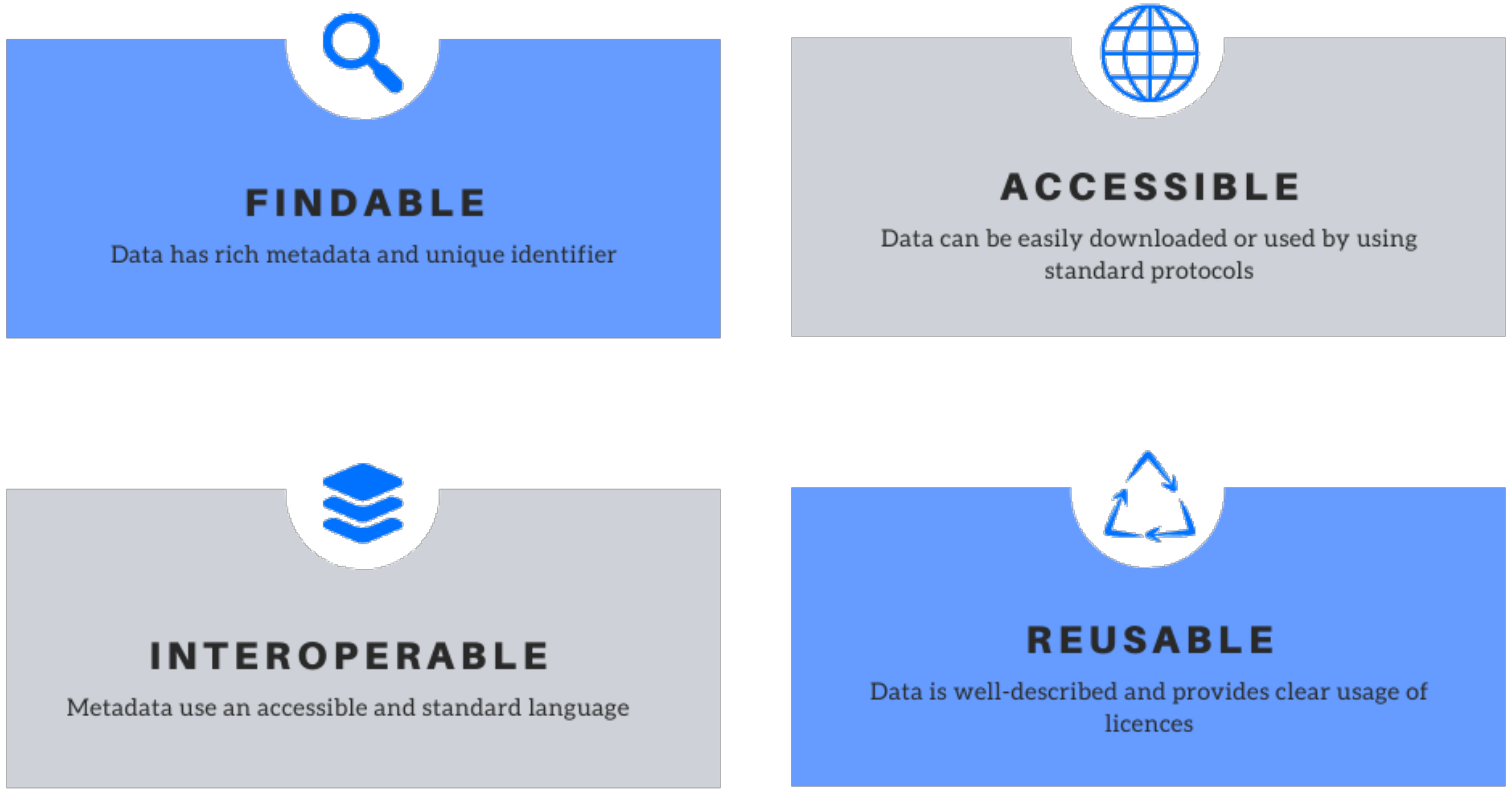
Strong implication of French community since the turn-on of the LHC; to be continued throughout the LHC physics programme.

Heavy lifting!
Needs close th-exp interaction and institutional support.

*) e.g., was asked to coordinate a white paper for Snowmass 2021 on this topic

F.A.I.R. principles … for theorists

Théorie @ IN2P3



Pheno studies are usually findable and accessible (arXiv)

However, they are often hard to reproduce in practice

Would like to organise a FAIR data initiative for theorists