

**MineHep** : progetto innovativo per l'estrazione di informazione di rilievo da dati pubblici in Fisica delle Alte Energie.



Conferenza GARR 2019  
*Connecting the future*

**Laura Redapi**  
Andrea Ceccarelli,  
Andrea Cioni,  
Maria Vittoria Garzelli,  
Piergiulio Lenzi,  
Vitaliano Ciulli.

Conferenza GARR 2019  
**CONNECTING  
THE FUTURE**  
Torino 4-6 giugno 2019

## Idea:

Creare uno strumento per rispondere alla domanda:

*I dati raccolti fino ad ora escludono il modello XXX?*

## Di cosa abbiamo bisogno?:

1) Un modo efficace per navigare i dati

- Organizzare i **dati già disponibili** ed estrarre **automaticamente** quante **più informazioni possibili** tramite un **interfaccia** di supporto

*Quali sono le misure rilevanti per il modello che voglio testare?*

*Vorrei confrontare dati di analisi diverse su un unico grafico...*

- 2) Un tool ad hoc per ottenere **previsioni** delle quantità misurate in base al modello XXX
- 3) Uno **strumento statistico** per calcolare l'esclusione del modello XXX

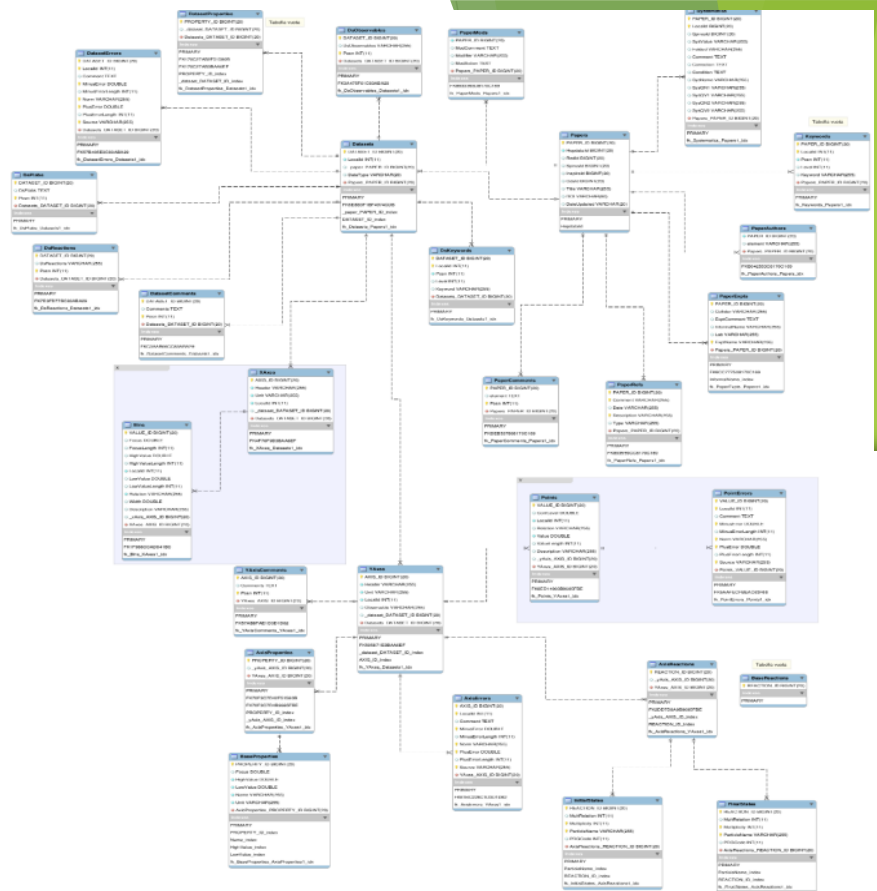
# Start-point: database hepdata (<https://www.hepdata.net>)

Database relazionale con struttura complessa  
(nuova versione: PostgreSQL)

Elemento base della ricerca è l'articolo  
scientifico

*Ricercio tutti quei papers che...*

Possibilità di navigare i dati all'interno di una  
singola analisi selezionata.



Deselect variables or hide different error bars by clicking on them.

**Variables**

Number of signal events ( $M_{LQ} = 400$  GeV)

Number of signal events ( $M_{LQ} = 900$  GeV)

523.0	1036.3	3.7011	0.0519420	4.11105	1.4099	0.1032	0.01
675.0	145.35	10.469	0.060983	2.0782	0.64316	0.26641	0.21
875.0	43.116	57.55	2.299	1.9385	0.6648	0.0002698	0.21

# State of the art: MineHep

❑ Nuovo database organizzato in Schema a Stella

❑ Utilizzo di Metabase, strumento opensource, per interrogare ed estrarre conoscenza dai dati in modo semplice tramite un'interfaccia SQL

**YAxisFact** → dato numerico della singola misura

Le dimensioni identificate sono :

- Properties
- YError
- XAxes

Observable



## Risultati ottenuti fino ad ora:

- ❑ Elemento principale della ricerca è il **dato** della singola analisi
- ❑ Possibilità di fare **query articolate**

*Cerco tutti i dati degli esperimenti che hanno certe keywords e l'osservabile in un certo range di valori scelti*

- ❑ Possibilità di **navigare su più analisi** contemporaneamente:

*Osservo tutte le tabelle di interesse contemporaneamente e decido quali dati, anche di esperimenti diversi, riportare in uno o più grafici*

## Step futuri e altri campi di applicazione:

- ▶ Eliminazione ambiguità creando un **vocabolario** (collaborazione con Università di Durham/hepdata)
- ▶ Creazione di un **interfaccia ad hoc** (linguaggio php e python)
- ▶ Applicazione degli stessi strumenti di data warehousing in **ambienti diversi** (possibili applicazioni a database di fisica nucleare, fisica medica...)