Grid Tools for Mokka and Co.

I. Marchesini (DESY), 2008-01-15,
ILD Workshop in Zeuthen

Introduction

- We propose to do the following centrally:

  - Simulation
  - Digitization
  - Low-level reconstruction

on the grid for all SM processes and to provide the results for everybody.

- In order to do that, we have developed a set of Grid tools, databases and web interfaces, which is at the moment complete for the simulation step.

- In the following, we'll show how that system works.

Simulation: introduction

- Large Mokka simulation:
  - all SM sample from SLAC;
  - both LDC01_05Sc and LDCPrime_01Sc;
  - priority to calibration samples and to processes needed by ongoing analysis.
- With 50 fb^{-1} of signal statistic a possible production would be:

Process	fb^{-1}	#events ($*10^3$)
ee->2f	20	~2500
ee->4f	50	~6000
ee->6f	50	~100
ee->hX	50	~20
ee->γ (n*γ)	10	~600
w (n*γ)	20	~1700
ee->ee	0.1	~200
eγ->eγ	0.1	~650
γ ->X	0.1	~1000
rest	1	~600

Process	#events
light quark 2f at 91.2 GeV	20000
tt (6f) at 350 GeV	20000

- This corresponds to ~14 million events for each detector model: centralized and automated production chain needed!

Simulation: production chain

- **Automated**: scripts chain with few inputs needed.

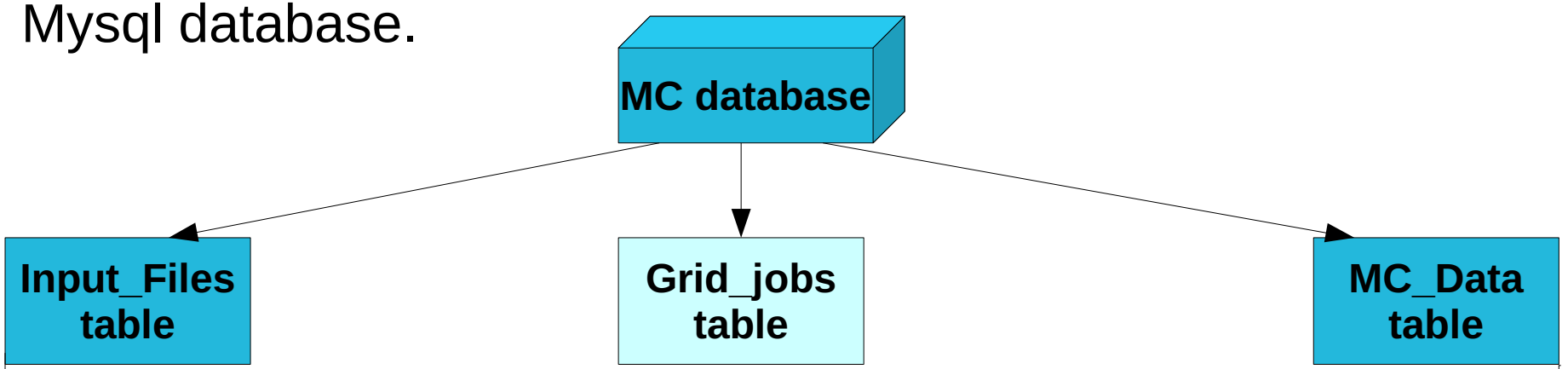
- **Database interacting**: keep track of the generator files and the simulated .slcio files. Temporary information about the running simulations also stored.

- Mokka is run on the **Grid**:
  - Scripts download input files, Mysql, Mokka, G4 Data from SEs;
  - Dump version of the french Mokka-DB used to avoid problems of outbound connection of the CEs;
  - Simulation output stored in SEs.

- Possibility to assign production **priorities**.

Simulation: Database structure

Mysql database.



All about generator files:
- location
- process
- #events
- ...

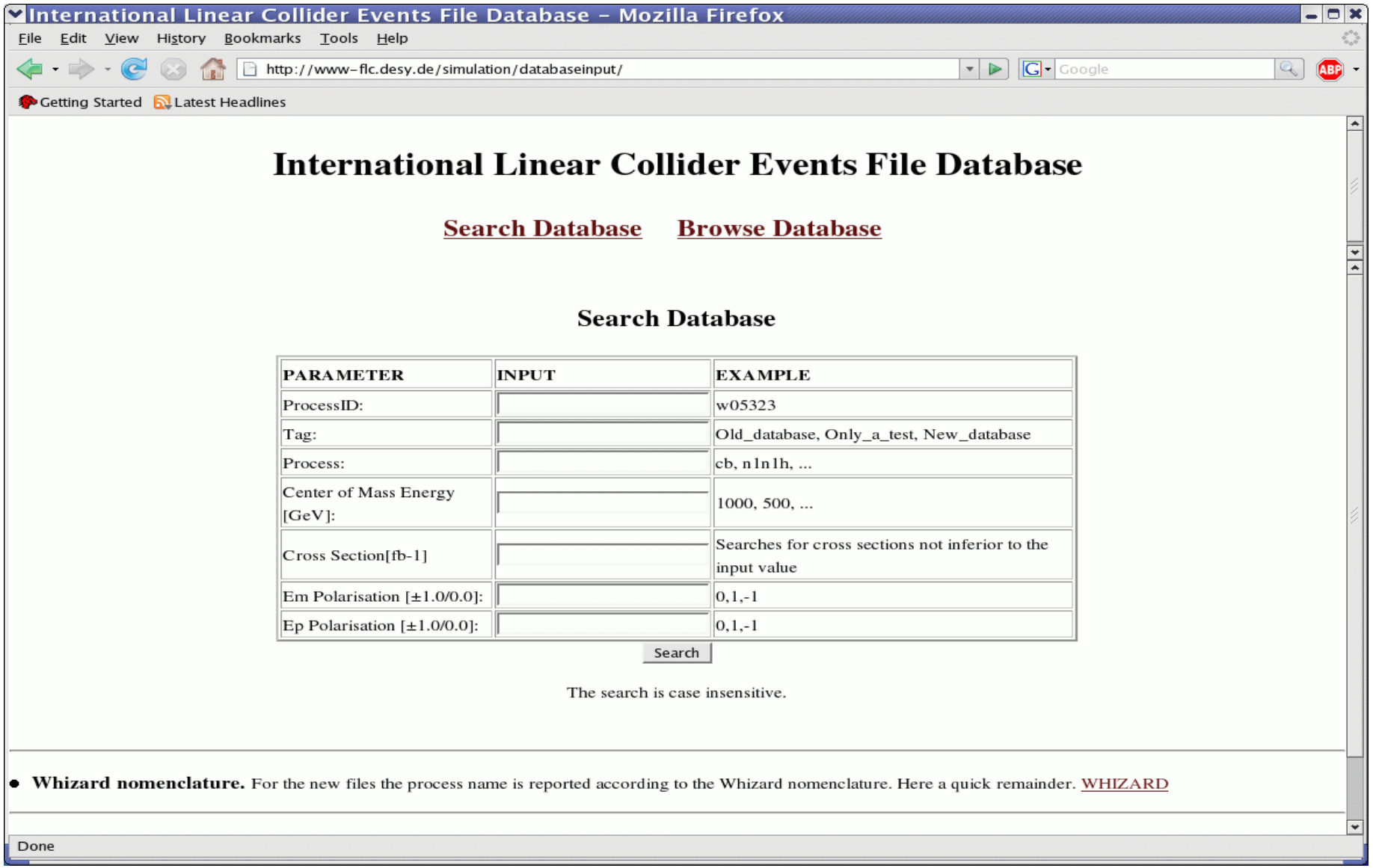
All about the jobs:
- grid id
- status
- input file
- process
- ...

As soon as an ended job is checked successfully, deleted from here!

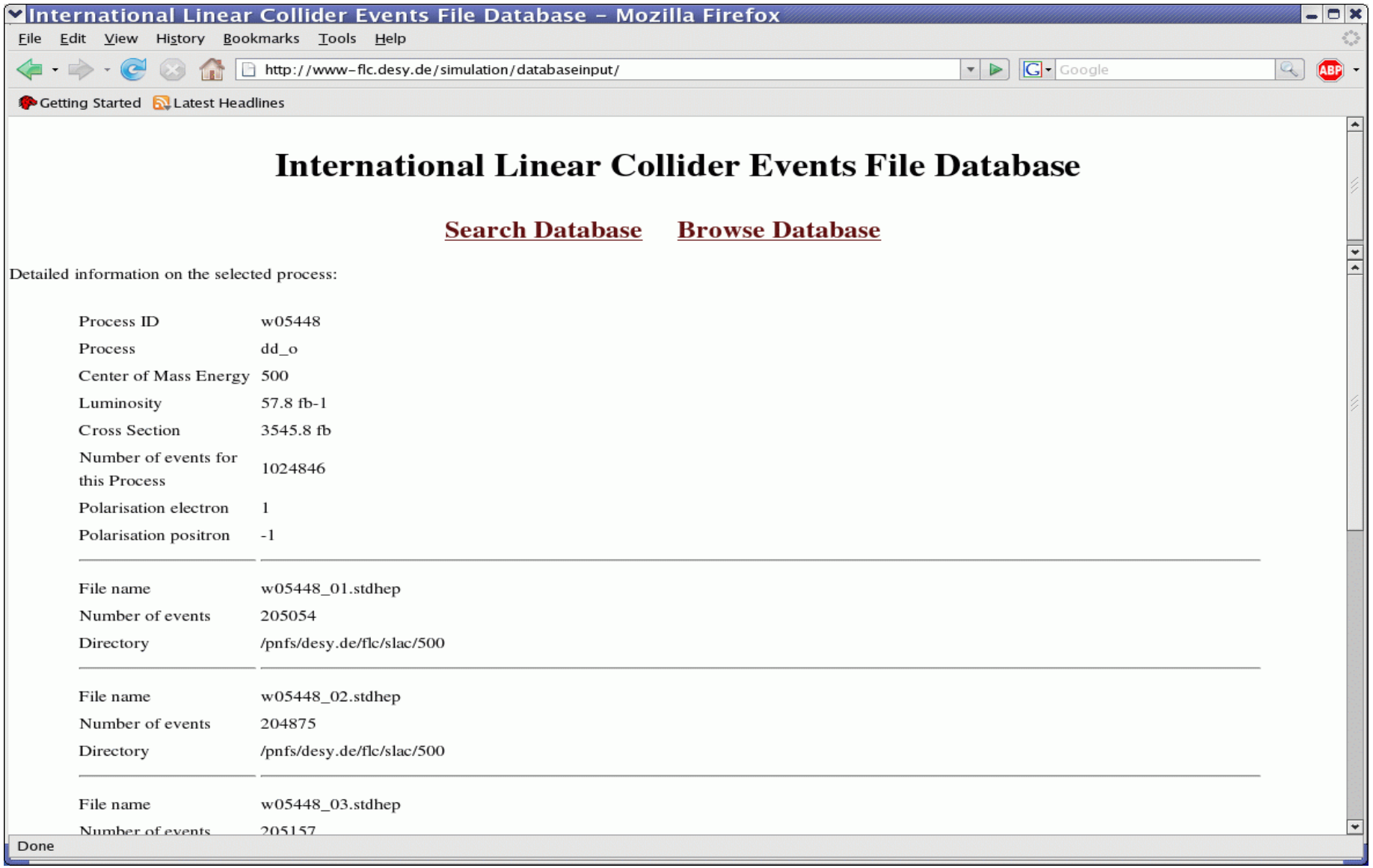
Successful jobs information:
- location (grid and SE)
- process
- #events
- ...

Read only, Php based, web front ends.

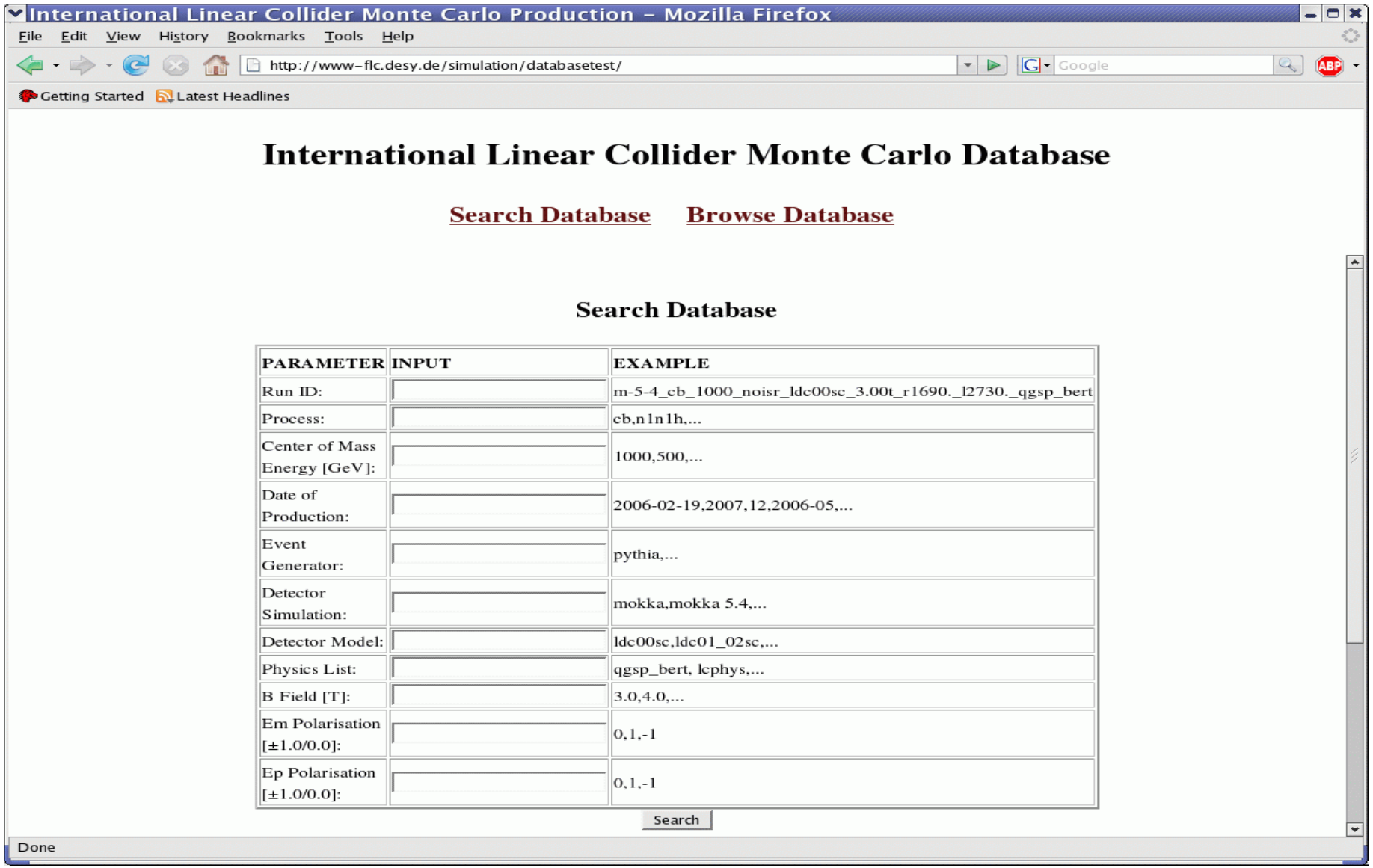
Input_Files table web front



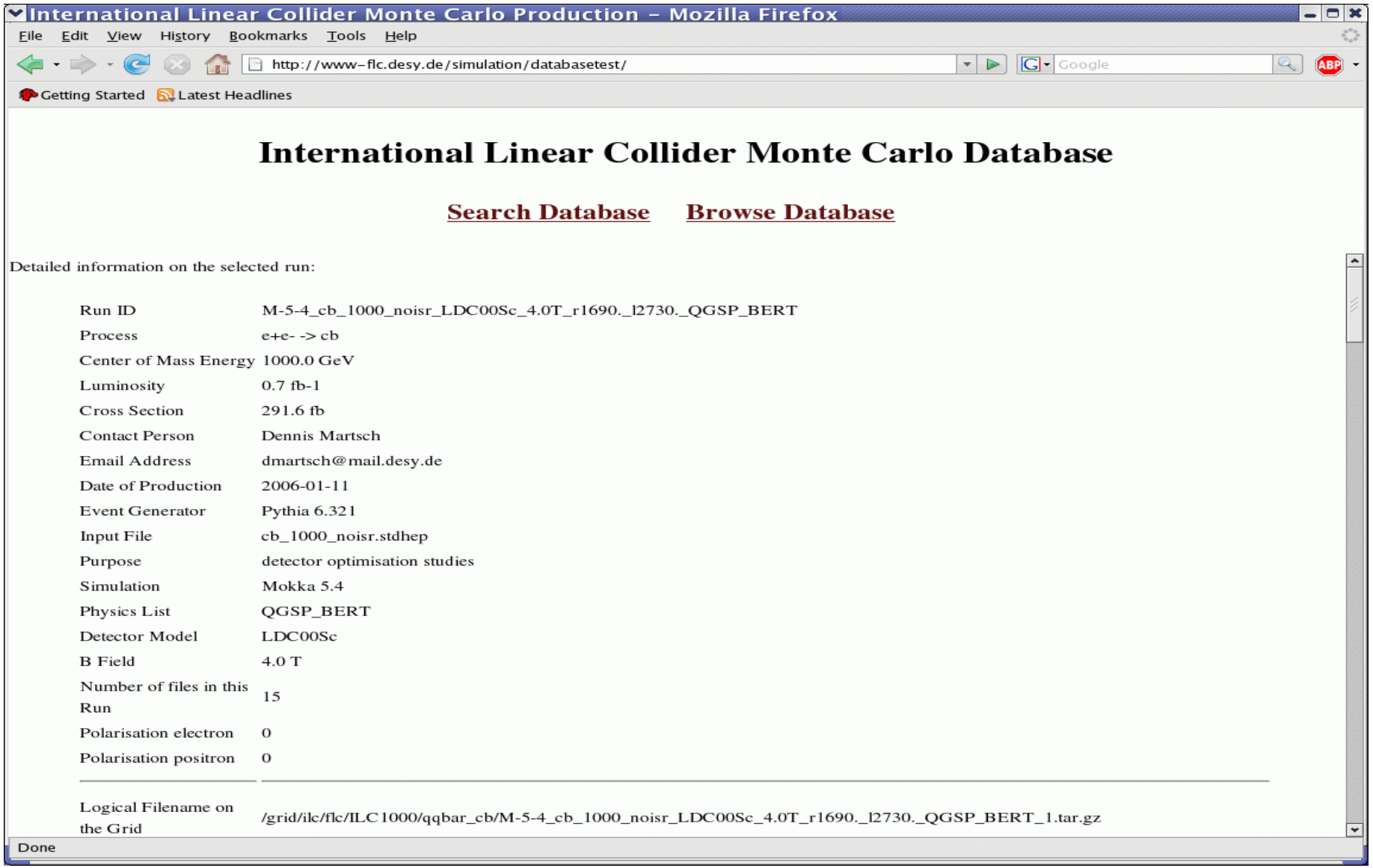
Input_Files table web front



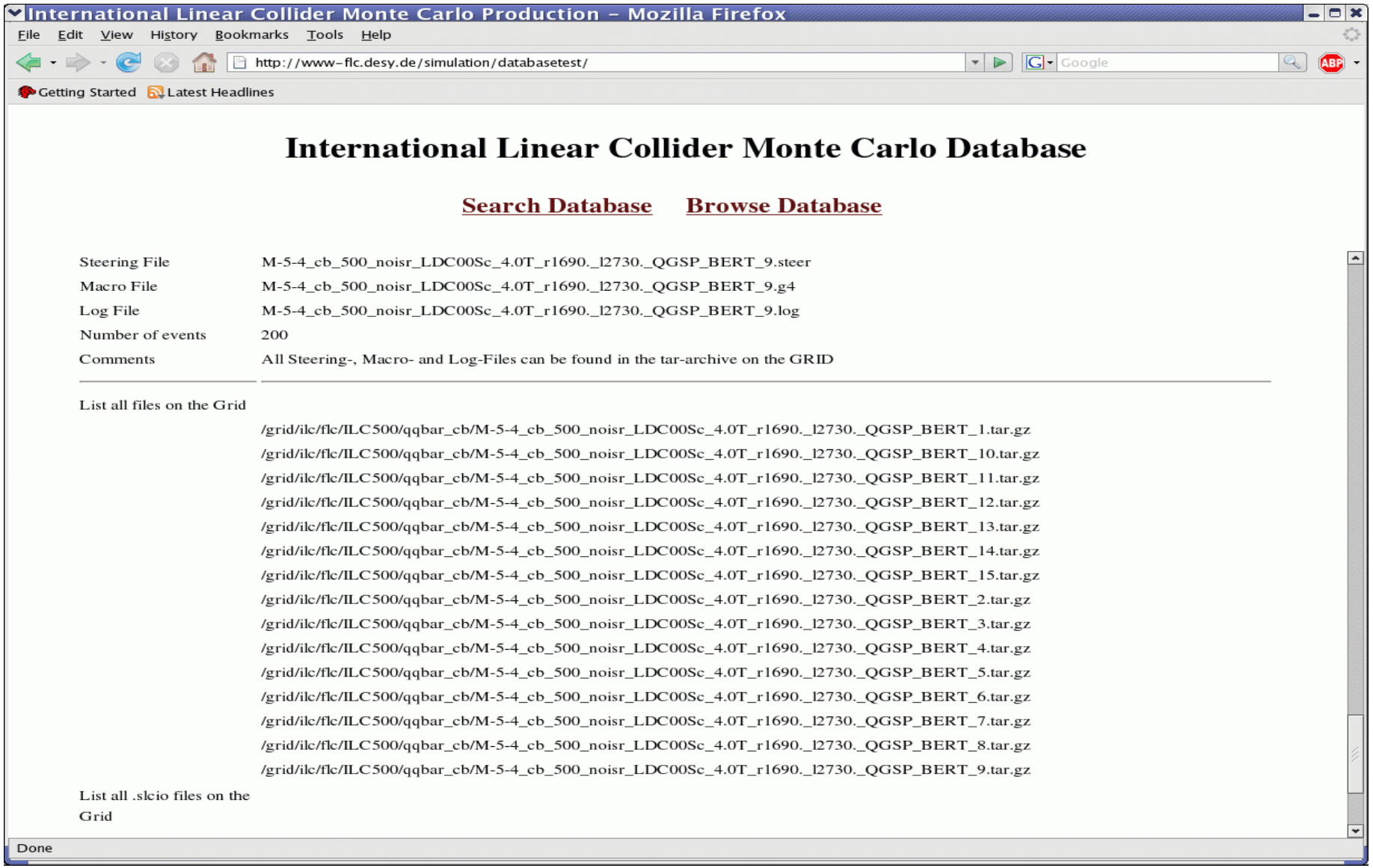
MC_Data table web front



MC_Data table web front



MC_Data table web front



Grid: inputs

- The input for a new simulation is a text file with the **process** (or a specific file), the **number of events** and the **start-event** wanted.

```
ww_500_noisr.stdhep max 1000
w05565 20000 0
```

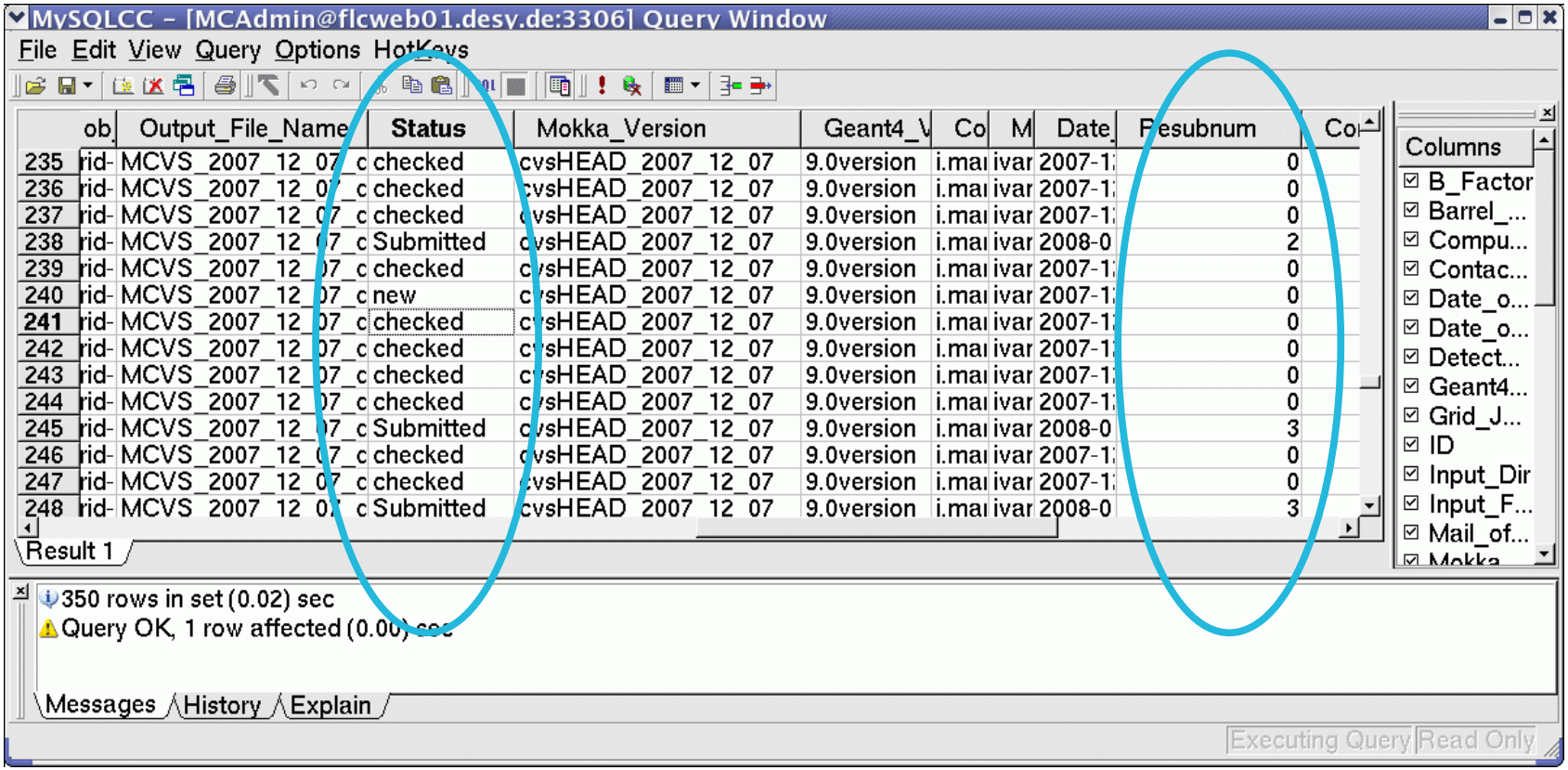
Other requirements like detector model, physics list, ... to be set just once.

- The number-of-events is splitted into jobs automatically, according to default values (that can be changed).

- Jobs are submitted to the Grid.

Grid: job status

- The status of the jobs together with the number of failures in their submission is updated and registered in the temporary Grid_jobs table.



Grid: checks and output

- The jobs are checked automatically. In the log file with the output from Mokka, the last event expected is required to be scanned. The .slcio file size is also checked.

>>> Event 9999, scanning sub-detectors

VXDCollection from the VXD sensitive detector has 347 hits.
SITCollection from the SIT sensitive detector has 162 hits.
FTDCollection from the FTD sensitive detector has 52 hits.

- If a job is checked successfully, information on the simulated file is written in the MC_Data table, otherwise it is resubmitted.

Comments

- The set of tools for the simulation has been tested and works.

- Development of similar tools for digitization and reconstruction ongoing. Same structure:
  - Running jobs on the grid;
  - Database-oriented;
  - Automated.

- Other institutes are invited to share these tools with us, to split the burden of the simulation in a coordinated way.

- The output, grouped into proper datasets, will be provided on the grid. The web interface of the database will provide location and all the information about the files.