



MC data catalogue proposal

Roma Tre Team + Giovanni

TABLE OF CONTENTS

01

Introduction

02

Proposal

03

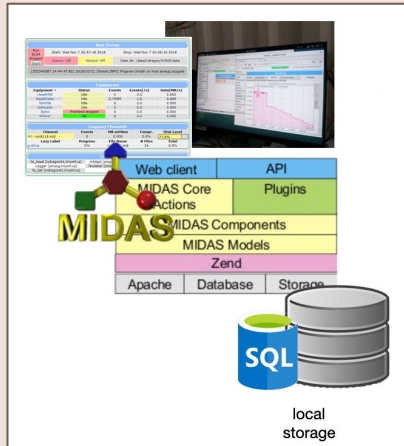
What we would like
to achieve

INTRODUCTION

- Today the simulation data is not yet cataloged;
- Neither the relationship between the MC data and its pedestals or Digitized data;
- On one of the Reconstruction Algorithm latest releases a change was made where the code automatically search for the right pedestal run to the analysis.
- In order to stay coherent with the real data analysis a table of content for the simulation data is necessary (and useful).

Run Catalogue for the REAL DATA

- When the DAQ acquires an entire run it sends all the information to a DB on the Cloud. This DB can be easily access [here](#) through a php web page.



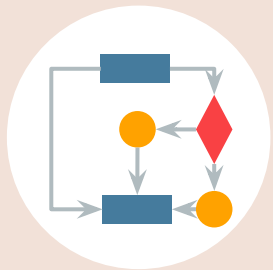
DAQ running locally



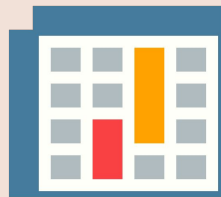
run_number	run_description	start_time	exposure_sec	GEH3_V	GEH2_V	GEH1_V	T2_V	T1_V	DRIFT_V	VOFFSET_V	PH11_V	PH12_V	PH13_V	PH14_V	STATE	stop_time	number_of_events
21	test	2022-03-21 15:17:13	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-21 15:17:26	8
34	test	2022-03-22 17:01:30	1	0	0	0	20	20	240	10	770	935	825	935	0		
35	test	2022-03-22 17:02:12	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:02:19	0
36	test	2022-03-22 17:03:26	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:03:43	10
37	test	2022-03-22 17:08:02	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:08:21	12
38	test	2022-03-22 17:18:26	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:21:40	99
39	test	2022-03-22 17:22:24	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:22:53	18
40	test	2022-03-22 17:23:12	1	0	0	0	20	20	240	10	770	935	825	935	0	2022-03-22 17:28:17	0

Data table of Simulated Events
on the Cloud

PROPOSAL: MC data catalog



Simulation
Algorithm



Data table of
Simulated Events
on the Cloud

Content of the table:

- MC data name
- All the simulation parameters
- Time of creation
- Description of the data
- Who has produced it (?)
- Any ideas?

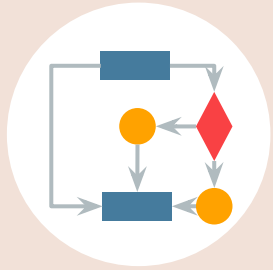
Python automatic* routine that gets all the important information and updates the data table.

*with the option to not register a simulation in case of test.

For the run name we suggest an "prefix+automatic increase number"

For example: MC_00001.ext

PROPOSAL: Same for the Digitized data catalog



Digitization
Algorithm



Data table of
Digitized Runs on
the Cloud

Content of the table:

- Digitized data name
- Source **MC data name**
- All the simulation parameters
- All the **digitization parameters**
- Time of creation
- Description of the data
- Who has produced it (?)
- Any ideas?

Python automatic* routine that gets all the important information and updates the data table.

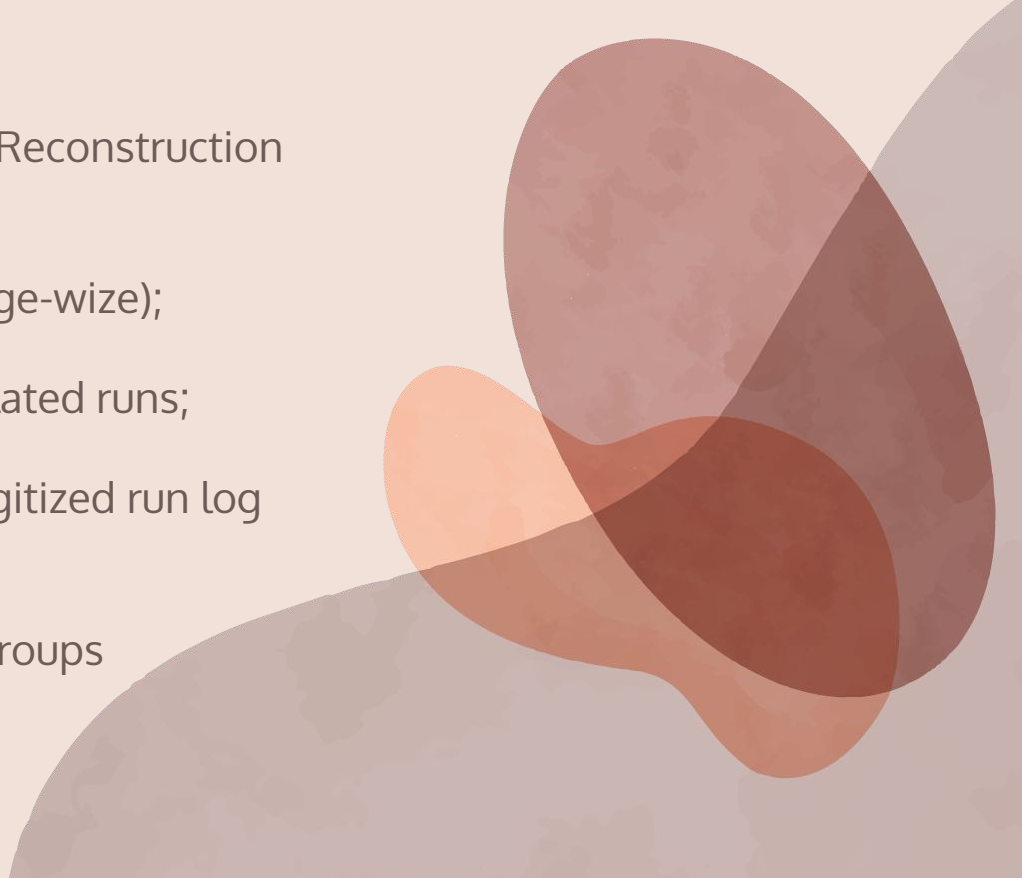
For the run name a suggestion could be:
"prefix+automatic increase number"

For example:

MC_Digitized_00001.ext

What we would like to achieve with this proposal?

- Better coherence/synchrony with the Reconstruction algorithm;
- Easier maintenance of the data (storage-wise);
- An always up-to-date catalog of simulated runs;
- PHP page to access the Simulated/Digitized run log (as the Real Data);
- Simplified synchronization of all the groups working with simulated data.





Thank you!