EEE-data access interface

Fabrizio Coccetti
Centro Ricerche Enrico Fermi

Outlook

- Acronyms and names
- How collected data are organised
- What is ELOG
- How to request data
- How to manage data
- Conclusions

Acronyms and names

- **ROOT**: *a modular scientific software framework. It provides all the functionalities needed to deal with big data processing, statistical analysis, visualisation and storage.* https://root.cern.ch/
- **DST**: **Data Summary Tape** is a file containing data that are in an intermediate processing state.
- **CSV**: **Comma Separated Value**. A textual format to represent data in a matrix fashion.
- **MC**: **Monte Carlo** is a class of techniques to simulate physical events basing on mathematical models.

The EEE tracks data base

- Hundreds thousands of files are stored at INFN-CNAF
- Tens of terabytes (1 TB = 1024 GB)
- 50+ billions tracks collected so far
- Reconstructed tracks are stored in **DST** files in **ROOT** format

The ELOG service

It is a place where researchers and shifters share information about the Experiment.

http://eee.centrofermi.it/elog

You can consider it as a forum with advanced features.

You will use it to report information gathered everyday during the EEE run (shift).

To access it you will be provided with per-school credentials (user name + password).

ELOG Query

The **ELOG Query** is a software that aims to provide an interface for both the researchers and the students to request EEE data, and recently MC files as well.

Through it, data can be extracted from the EEE data base depending on a set of parameters specified by the user.

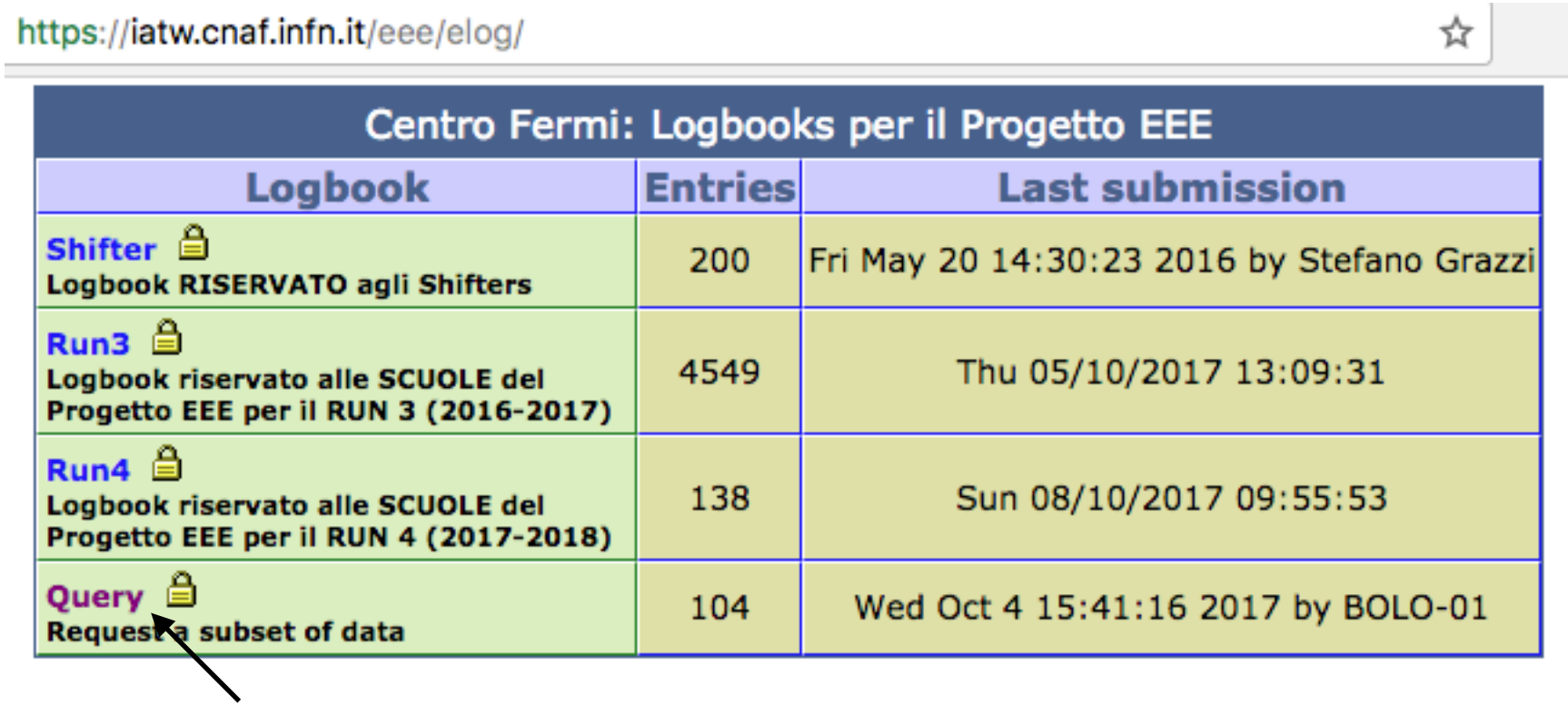
The parameters allow to narrow the search for interesting data out of the enormous set of EEE tracks.

Workflow

- The user logs into the EEE ELOG (the same used for the shifts) with usual user-name and password
- Creates a new entry in the dedicated logbook, specifying the parameters of the query and submits it
- Within a minute from query arrival, an automatic procedure starts and produces an output file accordingly
- As soon as the procedure returns, the output is attached to an entry as a reply to that created by the user.
  All parameters are copied back in the reply

Access to ELOG

http://eee.centrofermi.it/monitor and then click on the “Data Request” link
or
http://eee.centrofermi.it/elog/Query

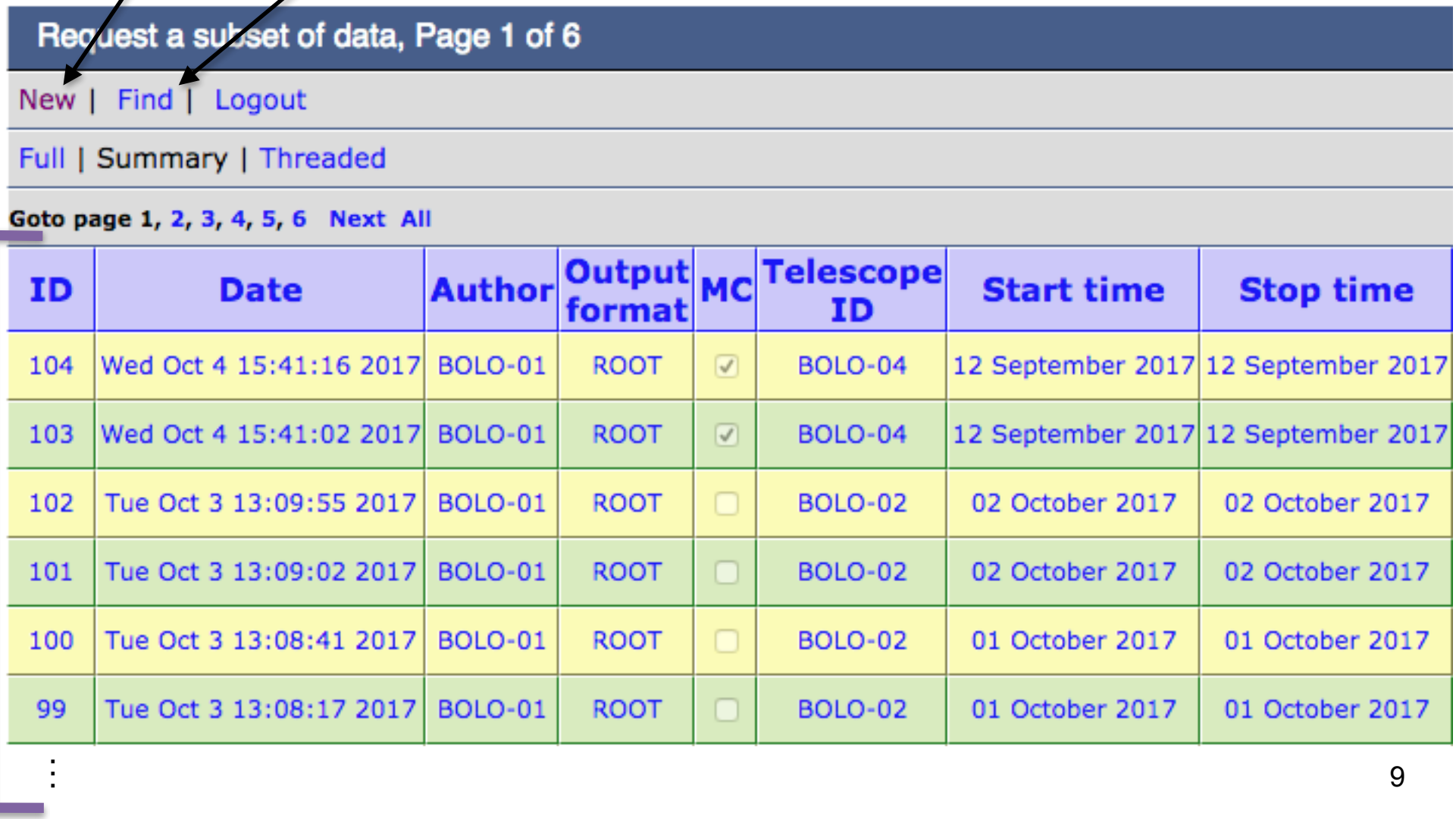


Click here and log-in!

Navigate the Query logbook

New request

Search previous requests/answers



Request a subset of data, Page 1 of 6

New | Find | Logout

Full | Summary | Threaded

Goto page 1, 2, 3, 4, 5, 6 Next All

ID	Date	Author	Output format	MC	Telescope ID	Start time	Stop time
104	Wed Oct 4 15:41:16 2017	BOLO-01	ROOT	☑	BOLO-04	12 September 2017	12 September 2017
103	Wed Oct 4 15:41:02 2017	BOLO-01	ROOT	☑	BOLO-04	12 September 2017	12 September 2017
102	Tue Oct 3 13:09:55 2017	BOLO-01	ROOT	☐	BOLO-02	02 October 2017	02 October 2017
101	Tue Oct 3 13:09:02 2017	BOLO-01	ROOT	☐	BOLO-02	02 October 2017	02 October 2017
100	Tue Oct 3 13:08:41 2017	BOLO-01	ROOT	☐	BOLO-02	01 October 2017	01 October 2017
99	Tue Oct 3 13:08:17 2017	BOLO-01	ROOT	☐	BOLO-02	01 October 2017	01 October 2017

List of previous requests/answers

Data request

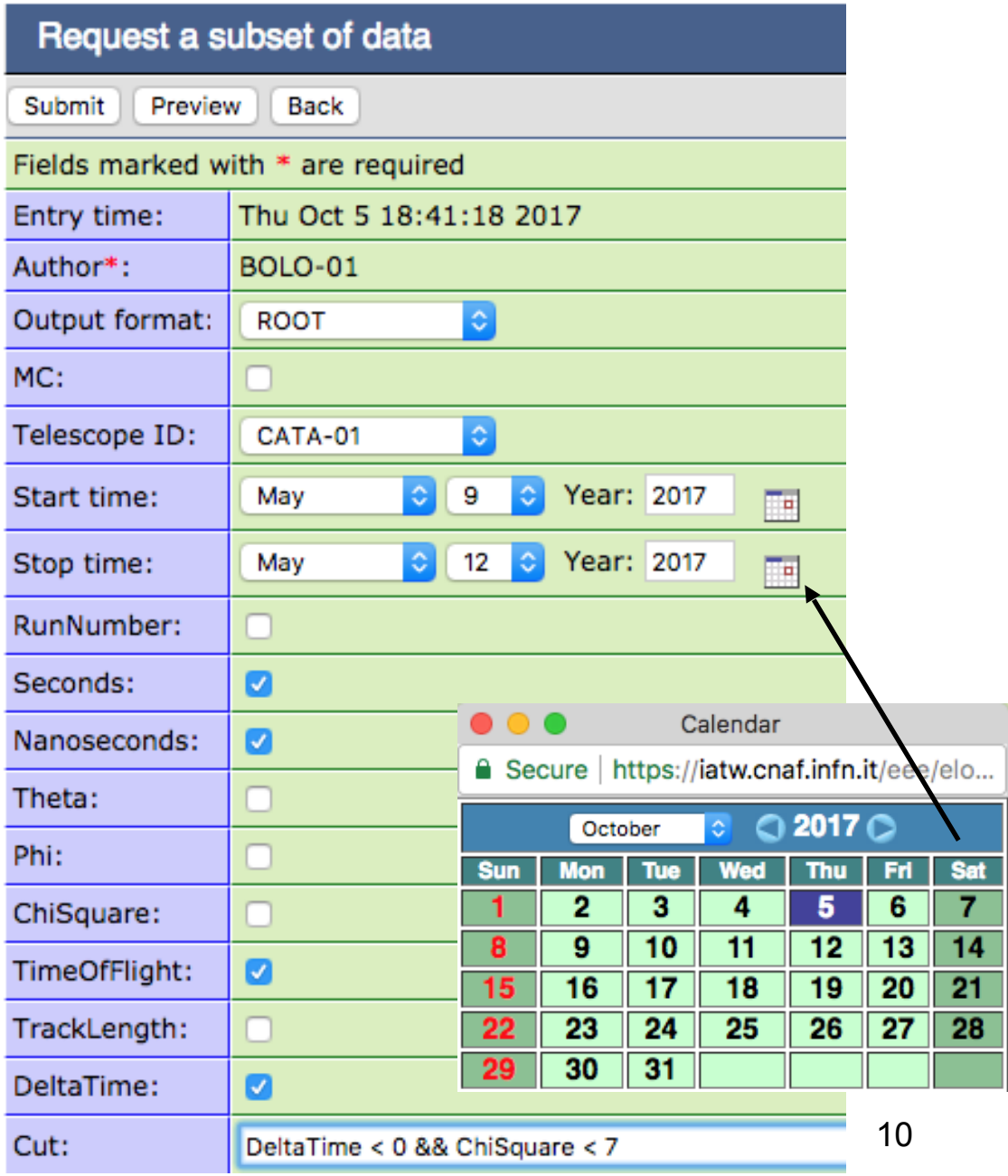
Request parameters

- **Output format**: could be either CSV or ROOT
- **MC**: it is a boolean parameter. If unchecked real data are provided, **M**onte **C**arlo data otherwise
- **Telescope ID**: the ID of the telescope (e.g. LAQU-01)
- **Start time**: initial date (day, e.g. 2017-03-15)
- **Stop time**: final date (day)
- **Cut**: a free-form string where the user can insert custom cuts
  (e.g. “ChiSquare < 9 && Theta < 10”)

Observables to provide in output file (boolean parameters)

- RunNumber
- Seconds
- Nanoseconds
- Theta
- Phi
- ChiSquare
- TimeOfFlight
- TrackLength
- DeltaTime

All observables can be used in the “Cut” field!



List | New | Reply | Find | Duplicate

Wed Sep 27 17:19:49 2017, BOLO-01, CSV, BOLO-04, 12 September 2017, 18 September 2017, 1 , 1 , 1 , TMath::Abs(Phi) < 0.1 && ChiSquare < 20

Wed Sep 27 17:23:14 2017, BOLO-01, CSV, BOLO-04, 12 September 2017, 18 September 2017, 1 , 1 , 1 , TMath::Abs(Phi) < 0.1 && ChiSquare < 20

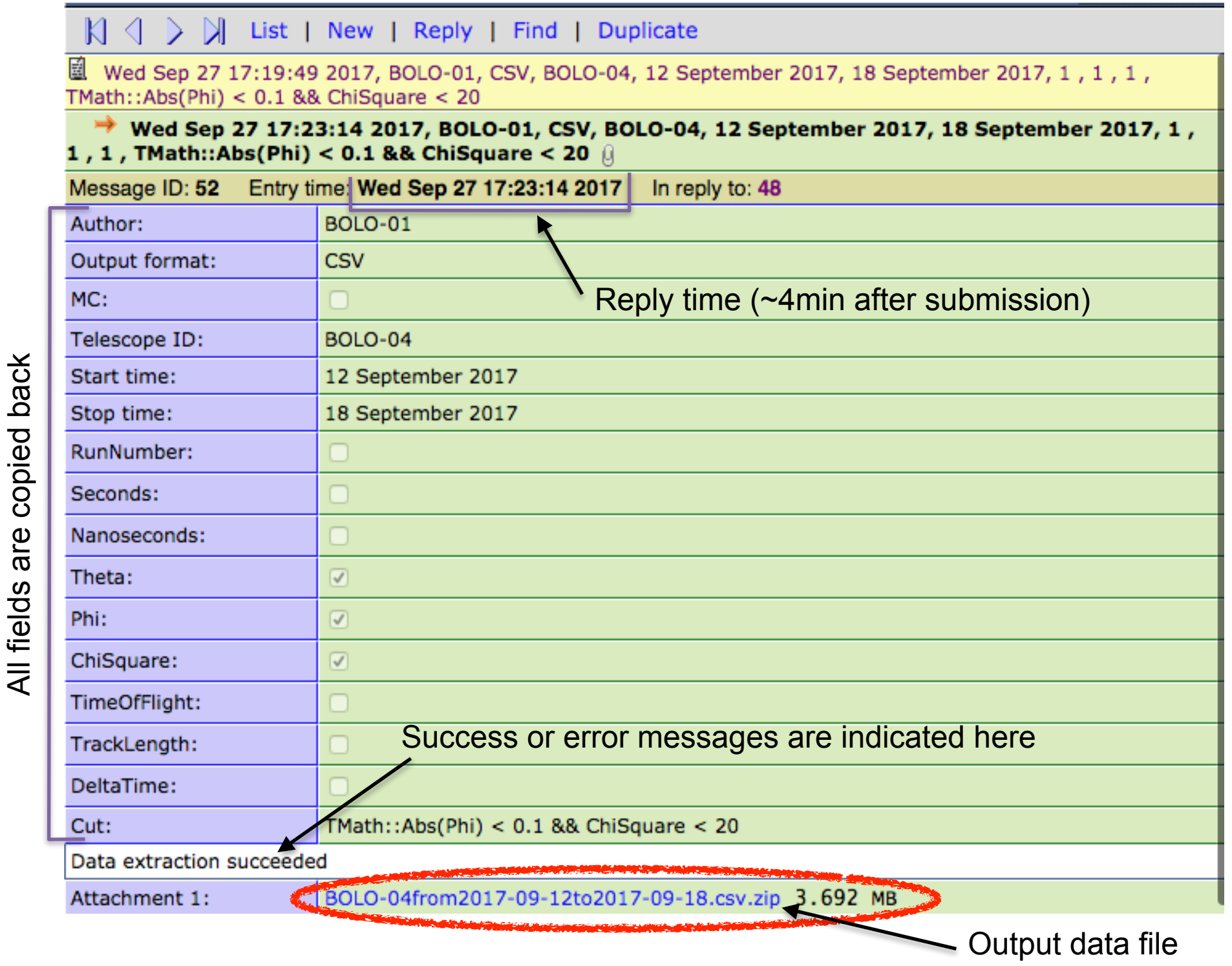
Message ID: **48** Entry time: **Wed Sep 27 17:19:49 2017** Reply to this: **52**

Author:	BOLO-01
Output format:	CSV
MC:	☐
Telescope ID:	BOLO-04
Start time:	12 September 2017
Stop time:	18 September 2017
RunNumber:	☐
Seconds:	☐
Nanoseconds:	☐
Theta:	☑
Phi:	☑
ChiSquare:	☑
TimeOfFlight:	☐
TrackLength:	☐
DeltaTime:	☐
Cut:	TMath::Abs(Phi) < 0.1 && ChiSquare < 20

Submission time

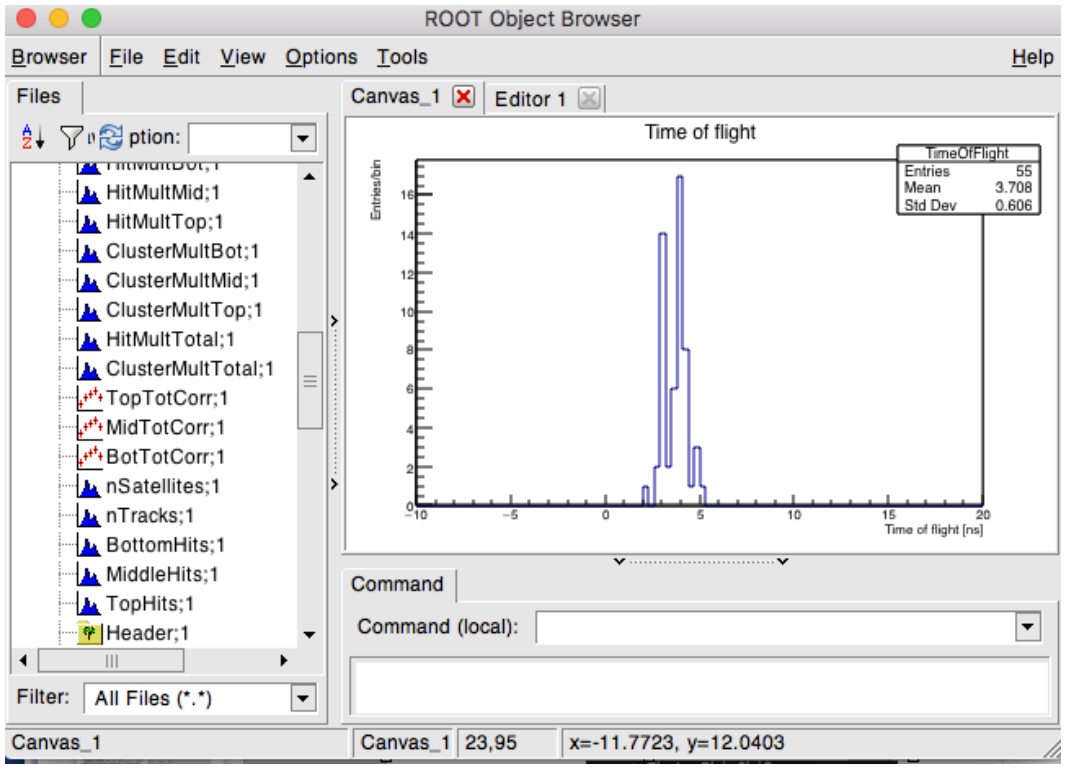
Requested observables

Provide only the tracks satisfying the condition

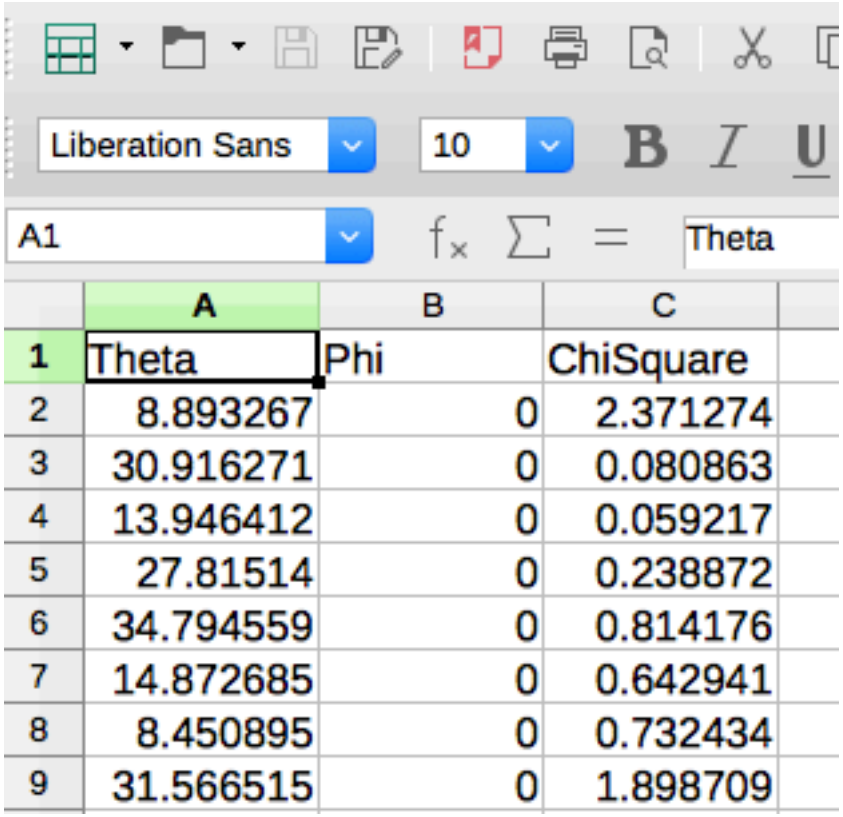

List | New | Reply | Find | Duplicate
Wed Sep 27 17:19:49 2017, BOLO-01, CSV, BOLO-04, 12 September 2017, 18 September 2017, 1 , 1 , 1 , TMath::Abs(Phi) < 0.1 && ChiSquare < 20
Wed Sep 27 17:23:14 2017, BOLO-01, CSV, BOLO-04, 12 September 2017, 18 September 2017, 1 , 1 , 1 , TMath::Abs(Phi) < 0.1 && ChiSquare < 20
Message ID: 52 Entry time: Wed Sep 27 17:23:14 2017 In reply to: 48
Author: BOLO-01
Output format: CSV
MC:
Telescope ID: BOLO-04
Start time: 12 September 2017
Stop time: 18 September 2017
RunNumber:
Seconds:
Nanoseconds:
Theta:
Phi:
ChiSquare:
TimeOfFlight:
TrackLength:
DeltaTime:
Cut: TMath::Abs(Phi) < 0.1 && ChiSquare < 20
Data extraction succeeded
Attachment 1: BOLO-04from2017-09-12to2017-09-18.csv.zip 3.692 MB
Reply time (~4min after submission)
Success or error messages are indicated here
Output data file
All fields are copied back

ROOT or CSV format?

ROOT is for advanced use.
It is programmable and can be used to ease automatic operations.



CSV is better suited for spreadsheet application, e.g. Excel



	A	B	C
1	Theta	Phi	ChiSquare
2	8.893267	0	2.371274
3	30.916271	0	0.080863
4	13.946412	0	0.059217
5	27.81514	0	0.238872
6	34.794559	0	0.814176
7	14.872685	0	0.642941
8	8.450895	0	0.732434
9	31.566515	0	1.898709

Conclusions

- ELOG “Query” logbook (http://eee.centrofermi.it/elog/Query) provides access to EEE data and MC.
- Requests are fulfilled within few minutes.
- Data selection possible via cuts.
- Different output formats: ROOT or CSV.
- Output files limited to 30 days & 50 MB: file exceeding one of these two limits are truncated.
- Heavy production load still not tested (might introduce some slowdown as the requests are processed serially).

Please report any problem to:

- carmelo.pellegrino@cnaf.infn.it
- francesco.noferini@bo.infn.it