




DArT data transfer, processing and storage

Pablo García Abia (CIEMAT)

DArT general meeting at LSC

April 21, 2022



DArT data status

- DArT running regularly at CIEMAT (Nov. 2019) and LSC (Apr. 2021):
 - Run list: <http://darkmatter.ciemat.es/links/dart-run-list>
 - CIEMAT data: 0.5 TB
 - LSC data: 3.0 TB
- MIDAS frontend writes lz4 files to local disk of the dartaq PC: 4 TB disk:
 - current disk usage is 75%
 - Analysis root files produced at CNAF.
- Regularly transferred to permanent storage (Grid) at CIEMAT and CNAF:
 - data access requires a Grid certificate (VO: vo.darkside.org)
 - semi-manual procedure

Bookkeeping

<http://darkmatter.ciemat.es/links/dart-run-list>

1	Run	status	Date	Time	Duration	Size (GB)	Detector location	Data storage (Pablo)	CNAF local directory	Shield	Rn Free Air	Source	Source position	Trigger Logic	Absolute pressure	PDM-1 [V]	PDM-2 [V]	PDM bias positive [V]	
7																			
8																			
9	212	good	9/3/2022	9:24	19 h 13 m	31.93	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
10	205	good	8/3/2022	11:42	38 h 37 m	36.73	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
11	run00204	bad	26/2/2022	21:00	7 h 21 m	17.39	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
12	run00182	good	22/2/2022	6:31	19 h 41 m	0.42	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
13	run00181	good	21/2/2022	10:22	21 h 56 m	1.79	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
14	run00180	good	20/2/2022	12:22	90 h 12 m	10.21	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
15	run00150	good	8/2/2022	16:51	19 h 34 m	4.38	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
16	run00147	test	4/2/2022	11:58	4 m	0,008	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
17	run00146	test	4/2/2022	11:44	4 m	0,014	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
18	run00145	test	4/2/2022	11:38	8 m	0,025	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
19	run00139	test	3/2/2022	12:38	3 h 1 m	31.95	LSC-Hall A	-	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
20	run00126	good	2/2/2022	15:41	7 h 36 m	1,34	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
21	run00110	good	26/1/2022	10:38	24 h	8,73	LSC-Hall A	CIEMAT	-	yes	no	none		Ext	LAr	30,5	30,5	3,4	
22	run00108	good	25/1/2022	16:42	13 h 20 m	127,34	LSC-Hall A	CIEMAT	dart	yes	no	none		Ext	LAr	30,5	30,5	3,4	
23	run00105	good	25/1/2022	15:02	1 h	0,17	LSC-Hall A	CIEMAT	dart	yes	no	109Cd	outside middle	AND	LAr	30,5	30,5	3,4	
24	run00104	good	25/1/2022	12:18	1 h	7,88	LSC-Hall A	CIEMAT	dart	yes	no	137Cs	outside middle	AND	LAr	30,5	30,5	3,4	
25	run00103	good	24/1/2022	17:55	9 h 32 m	1,66	LSC-Hall A	CIEMAT	dart	yes	no	none		AND	LAr	30,5	30,5	3,4	
26	run00087	good	3/12/2021	12:54	21 m	3,39	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
27	run00086	good	3/12/2021	12:30	24 m	3,32	LSC-Hall A	CIEMAT	dart	yes	yes	none		Ext	LAr	30,5	30,5	3,4	
28	run00084	good	3/12/2021	12:20	10 m	0,03	LSC-Hall A	CIEMAT	dart	yes	yes	109Cd	outside middle	AND	LAr	30,5	30,5	3,4	
29	run00081	good	3/12/2021	11:12	21 m	5,10	LSC-Hall A	CIEMAT	dart	yes	yes	137Cs	outside middle	AND	LAr	30,5	30,5	3,4	
30	run00079	good	2/12/2021	13:13	24 h 10 m	3,70	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
31	run00078	good	1/12/2021	12:09	16 h 7 min	2,70	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
32	run00077	good	30/11/2021	16:19	4 h 35 m	0,69	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
33	run00075	good	16/11/2021	14:35	46 m	1,76	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	
34	run00074	good	16/11/2021	14:07	1 h 30 m	4,78	LSC-Hall A	CIEMAT	dart	yes	yes	none		AND	LAr	30,5	30,5	3,4	

Data storage and transfer



- The local PC is a **4 TB** buffer. **Mind the free space!**
 - Formal procedure needed to backup files and clean up space.
- Only MIDAS files transferred to permanent storage:
 - **What is the fate of analysis root files?**
- CNAF T1:
 - Virtually no space limit: **to be verified**
- CIEMAT T2:
 - The reference site is CIEMAT. Transfer to CNAF is **asynchronous**.
 - New dedicated disk server: **320 TB** (*to be installed soon*)
- Manual operations in this process are an issue:
 - We should move to a fully automatized system: **RUCIO**, <https://rucio.cern.ch/>

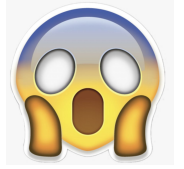


The future: towards an ^{39}Ar measurement

- Plan ahead of time the disk space budget:
 - Event size: 125 kB
 - With 16 channels, 8 μs window and 4000 samples, 80% compression: 8 GB / day / Hz
 - With a 100 Hz rate: 1 TB / day, *we are not doomed!*
 - Bandwidth: for 1 TB / day \rightarrow 12 MB / s (?)
- A formal procedure (centralised) is required for the timely transfer and storage of data files.
- Enforce regular clean up of the local `dartdaq` disk:
 - depending on the rate, consider increasing local disk space (within reason).



Data processing: analysis



- Analysis root files are produced with [pyreco](#) at CNAF:
 - Ludovico copies MIDAS files to CNAF local storage (limited, temporary)
 - Analyzers (Ludovico, Timothée) run [pyreco](#) over those files at CNAF and produce local copies of analysis root files.
 - The root files may or may not be duplicated: **coordination not enforced**.
 - Root files not backed up on permanent storage. Not an issue, **or is it?**
- A formal procedure (centralised) is required for the official production of analysis root files, as well as a storage policy.

Conclusions

- Data transfer, processing and storage is a good old friend:
 - we know the main issues, have most of the tools and some resources.
- The current procedures and resources are doing the job:
 - fine for the testing phase of DArT, fear the future!
- Robust procedures for data transfer, processing and storage must be put in place in the coming weeks.
- Resources have to be secured: not an issue in the one year scale.