

CCF DATA ARCHIVING – THE ARCHIVE SIDE



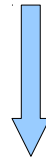
Johan Bregeon, Raquel de los Reyes &
Alessandro Carosi (INAF-ASDC)

on behalf of the Rome archive group:
L.A. Antonelli, S. Gallozzi, F. Lucarelli, V. Testa, S. Lombardi, M. Mastropietro



Main points to be understood:

- Technical: which data have to be archived and who need to access them?
- Technical: approximate total data amount (~few TB/year ?)
- Technical: data transfer rate? (1 tar-file/day?)
- Technical: does CCF take care of the needed drivers/software/interfaces to write data on disk/servers?



- what does CCF need from the archive? (i.e., use cases)

We preliminary identified two possible scenarios for an archive solution that could be managed by our current system in Rome

*[* It is important to remark that these are TECH0 data: during CTA “normal” operations, these will be managed by ACTL through TMCDB *]*

1) Archive is mainly a repository with I/O access

- Search for data file name
- No need (necessarily) for a DB

To browse the archive repository you can structure the physical file names according to the way they should have been accessed.

2) Data are accessed through browsable metadata

- Work on CCF data-model is required
- Dedicated DB has to be populated

it won't likely be compliant with the full CTA archive system where TECH data will be managed by ACTL

- Archive access through client/server resources (solutions currently under investigation/implementation)



+ data backup in the INFN-Frascati dedicated storage resources (already in the GRID)

or alternatively through DIRAC client on the GRID

- iRODS
- ONEDATA

Questionnaire for CCF for site characterization archive ☆

File Edit View Insert Format Tools Table Add-ons Help Last edit was made 7 days ago by Surfinjak

alessandro.carosi@gmail.com

Comments

Share

100% Normal text Lato 10 B I U A

3 2 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Questionnaire to prepare the discussion during the CCF meeting in Barcelona concerning the
Archive and data transfer of the Site Characterization devices

- Hardware setup on-site (please, include the details you already know):
 - ☐ will there be a 24h/24h reliable internet connection?
 - ☐ will the connection have to go through a proxy server ?
 - ☐ will there be a central server, or will each instrument have its own machine ?
 - ☐ will there be an on-site backup of all data ?
 - ☐ which is the expected daily data transfer rate?
- Information about the data model:
 - ☐ will data be pre-formatted on-site ?
 - ☐ will all data be on files or will there be data bases ?
 - if "files": which kind of data formats will produced?
 - ☐ will each instrument team save its own copy of its data ?
- How the archived data will be access:
 - ☐ which is the expected access rate?
 - ☐ do you need access data through browsable keywords or is it enough with a simple