INAF Science Archives & the Big Data Challenge

17-19 June 2019
INAF
UTC timezone

Search...

The purpose of the workshop is to gather for discussion all main Italian actors involved in the use and management of astrophysical data, also within the interdisciplinary perspective of multimessenger. From an overview of the existing archives and their development, to the discussion of the Archive 2.0 concept for the Big Data, the different functionalities of archives will be presented. The use of modern era archives is no longer circumscribed to the search for scientific information, but it extends to providing the framework for the search, manipulation and analysis of data from telescopes, either terrestrial or satellite, of the new 2020 era.

SOC:

Summary by topic

1. Status session:
    a. Data coming from several providers and released by different sites/institutions for quite all the UTG instruments
    b. What is present: one INAF infrastructure (IA2) + one ASI data center currently dedicated to space missions + several project oriented infrastructures
    c. Workshop contributions and discussion: a lot of hints!
    d. Strong tools for data handling / visualization / analysis
    e. Different perspectives from two international infrastructures: both uses STANDARDS!!
    f. Pipelines not always available and shared
    g. various level of high-level science products
    h. Zerbi: What we need is mentality, effort and infrastructure. New figures in INAF are needed, who are in the middle between science and technology.
    i. Open Data + Open Access = Open Science:
        i. interoperability;
        ii. curation&preservation (data, software, documentation and publications)
    j. No common rules among various experiments/archives

- not important to store everything locally
- lack of computer scientists (difficult to find them), difficult to guaratee permanent positions, salaries are not attractive
- Standards: FAIR, IVOA, RDA, common data models
-
- User representatives in the staff (domain specialist as well as Astronomy specialist)
- machine learning and data mining!! strategic for data exploitation.
- challenging use cases!!! All of them are store+ computation oriented

Criticalities:

reconstruction and reusability ancillary data (telemetry, …..)

private pipelines

no commonly adopted standards

limited interaction between working groups (projects, team, …)

ROADMAP

- conveen around a table with the endorsement of INAF directorate: find consensus upon
  - STRONG NATIONAL interaction on overlapping activities
  - definition of common strategies, strong cooperation and coordination
  - highlight individual / local contribution within the national landscape
- use internationally agreed upon standards!
  - proposing and improving new standards when needed
- data publication:
  - this will increase the scientific return and ensure growth of italian community
- change of user approach: convince the community
  - change of archive perception: data+computing