Understanding data sharing in high energy physics
a data scientist point of view

Mireille Louys , CDS & ICube , Strasbourg University

François Bonnarel, CDS, Observatoire de Strasbourg

Mathieu Servillat, Catherine Boisson, LUTH, Observatoire de Paris

Michèle Sangillon , LUPM, Montpellier

within the IVOA provenance and Data Model working group

Escape - CEVO preparation meeting, 2021 Sept 1

Exploring needs and practice

- about data discovery for high energy data
  - Preliminary meeting June 25th : discussion at LUTH, Paris-Meudon [Minutes available here](#)
  - based on *HESS* data distribution and *gamma_py*
    - what exist in the projects
      - data discovery of event lists, data formats in FITS
      - context for data acquisition is also needed: IRF, good time intervals, etc .
    - what exist to share data in the IVOA : various ways
      - granularity of the data matters

Scenarii for data discovery

- is there some data for this region of sky and at this period of time
    - MOC multi order coverage
        - S spatial MOC
        - T temporal T-MOC
        - both ST -MOC
- give me metadata about the selected event list
    - ObsCore can do
- give me events and observation context
    - how should access be organized?
        - events, then IRF and GTI ? ObsCore+Datalink
        - open GTI and IRF, extract selected events —> trace data selection and processing
            - provenance metadata   IVOA Provenance model /service
            - workflow description   ???

Get detailed scenarios

- to get the best from existing protocols
    - describe scenarios for discovery, data selection
        - spot where the useful metadata is stored : headers , fits keywords etc.
        - define the list of metadata needed for each scenario
- identify what is available already
- identify what can be modified and added for HE and MM astronomy
    - VOEvent extensions ?
    - ObsCore extensions ?
- which steps in processing do we need to describe ?
        - Last step Provenance, Workflow binding
- which description format for a common processing graph ?

Thanks for your attention

welcome for comments and questio