Applying
VO Data Models to
INAF Radio Data Archive

C. Knapic, A. Zanichelli, M. Molinaro, M. Dolensky, S. Righini, A. Orlati, F. Bedosti

follow-up from ...

ADASS XXVII talk (C. Knapic)

Applying VO Data Models to INAF Radio Data Archive

Authors: C. Knapic, A. Zanichelli, M. Molinaro, M. Dolenski, S. Righini, A. Orlati, F. Bedosti
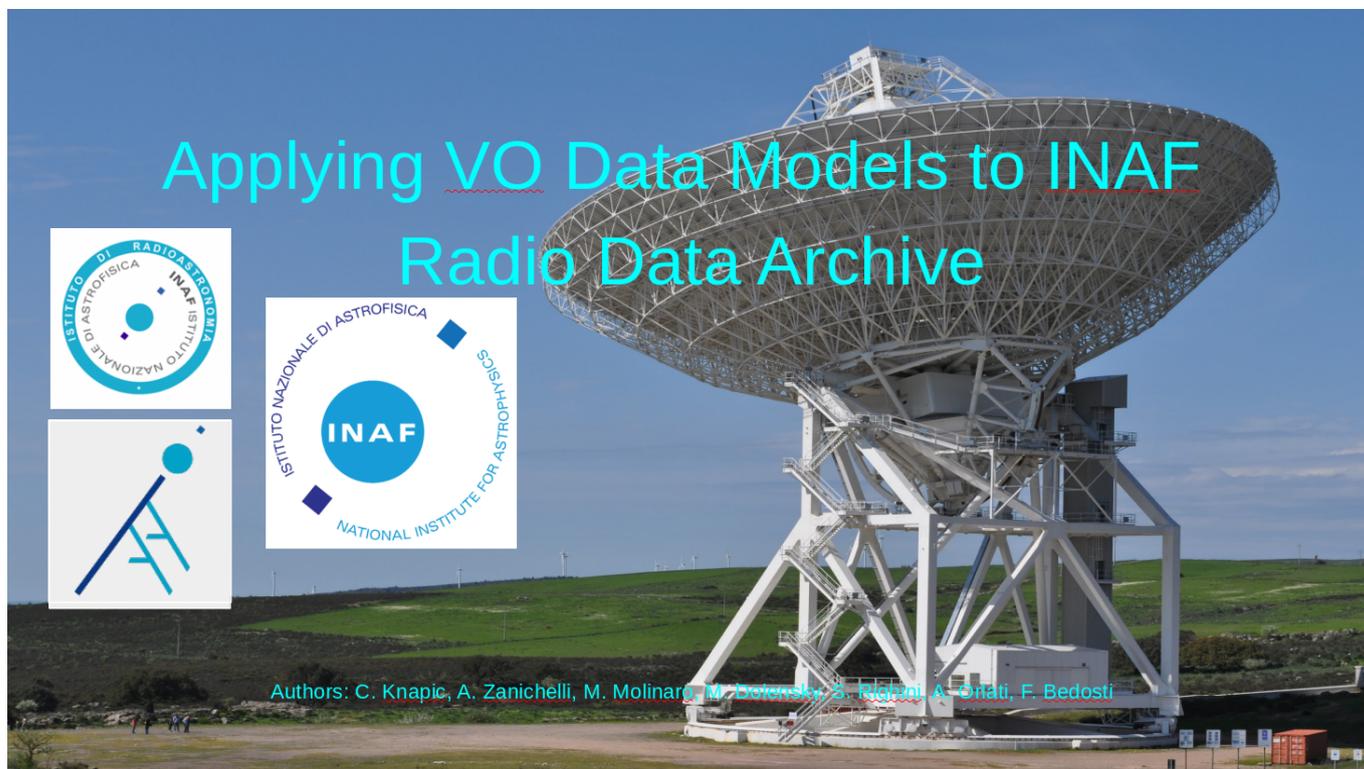
DataSet(s)

- **Single Dish & VLBI-IT raw data**

  - 2 data formats for Single Dish

    - FITS + summary.FITS

    - M(ulti)B(eam)FITS

  - 1 data format for VLBI-IT

    - Visibility FITS + summary.XML

- **All of them: global.tar file as the DataSet**

Archive & User Interface

- **The resulting archive model is complex**

  - Because it has to be exhaustive

    - It can cope with multi-faceted requirements

      - Science, technical, …

  - Results in a complex RDB solution

- **Scientific users are not interested in this complexity**

  - Discovery (and access) are simplified by

    - Custom UDFs mappings

    - Exposing a use case derived set of filtering parameters

Model in VO: goals

- **Porting the resource in a VO solution to**

  - Allow interoperable access

    - Using existing models and metadata

      - Discovery and access use cases

  - Side-by-side to UI

    - Extended users community

  - Improve knowledge of needed metadata

    - In existing or revisioned models

    - To allow data processing

      - Bring (correlated) raw data to science ready level

INAF approach

- **TAP Service**

  - IA2 implementation

  - custom User Defined Functions to improve query performance

    - get_table = f(pos, freq,…) ;

      - pos= f(ra,dec,rad,..);

  - ObsCore [and CAOM?] deployment

  - UDF at TAP capability level or transparently mapped behind

    - yet to decide

- **Datalink access to datasets**

  - Access to the atomic datasets

  - planned to provide additional metadata description

    - probably requires an addition to Datalink semantics vocabulary

CSIRO/ASKAP solution

- The MWA project is using GAVO's DaCHS for IVOA layer

- ASKAP archive (CASDA) uses DataLink to package and access complex data products (screenshot)

- SKA has derived an observation model from these precursors and verified against CADC's CAOM

- NGAS is used for bulk storage of MWA and ASKAP precursor data (among many other telescope sites)

CASDA Datalink Result

The SIAP services are primarily intended for use with scripts and programmatic data access. Please refer to the CASDA User Guide for information and examples.

ID	access_url	service_def	error_message	description	semantics	content_type	content_length
visibility-6	https://casda.csiro.au/casda_data_access/data/sync?id=gyVwQJeH8404JmSRNJdoXKlmak6rwMCUnr3HjwRDhsbGTwzdJru3jw0GDbugq9sS			Download data product via web	#this	application/tar	3957964800
visibility-6	https://casda.csiro.au/casda_data_access/data/sync/pawsey?id=gyVwQJeH8404JmSRNJdoXO--tUrnV889GwQ2A2knHzEAjrUjrMmaiNwmxvQ8Rvt7			Download data at Pawsey Centre. Note: Only users from within PSC network can access the data through this link. Only use if you currently have access to Pawsey facilities.	#this	application/tar	3957964800
visibility-6		async_service		Scripted file access via Web	#this	application/tar	3957964800
visibility-6		async_service		Scripted file access via Pawsey Centre. Note: Only users from within PSC network can access the data through this link. Only use if you currently have access to Pawsey facilities.	#this	application/tar	3957964800

ObsCore metadata components

- **Dataset metadata content and RDB content are enough to cover ObsCore MUSTs**

- **Discovery requirements are met**

  - Considering the atomic.tar DataSet

- **Interoperability of radio DataSets should not be a problem**

  - Supposing considered datasets are representative

CAOM

- **A possible solution to map the complexity of the existing RDB**

  - CompositeObservation, Artifact(s) and Chunk(s) could help

    - But would require denormalisation to represent FEBE and waveband sets properly

  - Collaborate to adapt CAOM to our use case?

Conclusions

- **ObsCore core model seems to fit atomic dataset discovery for Single Dish & VLBI-IT collections**

  - Thus allowing interoperability through a TAP + Datalink solution

- **Applying CAOM solution may preserve the inherent complexity of the atomic data**

  - Cost: denormalization

- **Open question: what metadata is needed to bridge the gap from data access to data processing**

  - Once we define what's needed to go from correlated to science-ready where do we set that information?

Conclusions

- **ObsCore core model seems to fit atomic dataset discovery for Single Dish & VLBI-IT collections**

  - Thus allowing interoperability through a TAP + Datalink solution

- **Applying CAOM solution may preserve the inherent complexity of the atomic data**

  - Cost: denormalization

- **Open question: what metadata is needed to bridge the gap from data access to data processing**

  - Once we define what's needed to go from correlated to science-ready where do we set that information?

Thank you!