

# WP4 Status

## DADI Data Access, Discovery and Interoperability

Françoise Genova and the WP4 team

# Targets

Three Tasks in support to three complementary targets

- Task 4.1: Support to astronomy ESFRI facilities, their pathfinders and other infrastructures of pan-European interest for implementation of their data in the VO framework (INAF/UHEI)
- Task 4.2: Support to the astronomical community (CNRS-CDS/INTA)
- Task 4.3: Updates of the VO framework from feedback and requirements (CNRS-CDS/UEDIN)

# Activities since the beginning

- DADI is structured around the organisation of Workshops (Deliverables) and participation in external events (IVOA meetings, which are Milestones, plus ADASS, RDA)
- Intense activity during the first 9 months
  - to create DADI community
  - to introduce newcomers to the VO
  - to share information on the partners' expertise, activities and needs
  - to begin to identify topics of common interest and collaboration paths
  - to perform technological work on the initial priorities, in particular multi-dimensional data (IVOA standards) – defining standards and agreeing on them takes time!

# Deliverables

- One deliverable/task during the period
- Deliverables are Workshops, the « text deliverable» is provided several weeks afterwards

#	Title	Lead partner	Due date	Actual date and location
D4.1	First DADI Technology Forum	CNRS/CDS (Task 4.3)	September 2015	Held 17-18 Sept. Strasbourg Del. 3 Nov.
D4.2	FIRST ASTERICs European School	INTA (Task 4.2)	November 2015	Held 15-17 <b>Dec.</b> Madrid Del. being finalized
D4.3	First ESFRI Forum & Training Event	INAF (Task 4.1)	November 2015	Held 3-4 <b>Dec.</b> Trieste Del. Being finalized

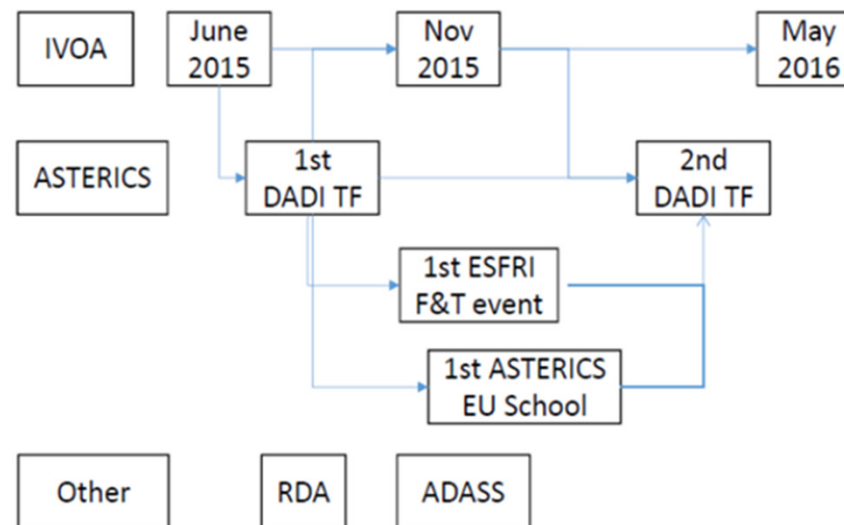
## Milestones and other relevant meetings

- The IVOA Interoperability meetings are milestones

M2	IVOA Sesto (Italy)	14-19 June 2015
M5	IVOA Sydney (Australia)	30 October-1 November 2015

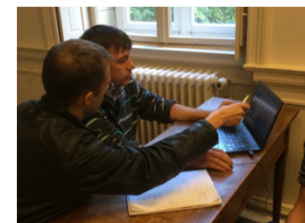
- Other relevant meetings
  - ADASS XXVth Meeting, Sydney, 25-29 November 2015
  - RDA Sixth Plenary, Paris (France), 23-25 October 2015

# DADI events in a nutshell



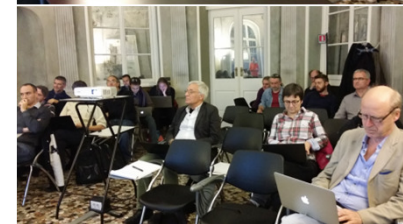
# D4.1 First Technology Forum

- 17-18 September 2015
- 34 participants (except KM3Net-VLVvT), including ESO and ESA
- DADI kick-off
  - Information about ASTERICS and DADI
  - Presentation of the VO Framework by the IVOA Chair (C. Arviset, ESA)
  - Presentation of partners' expertise and on-going work
  - « Hack-a-thon » i.e. discussions by small groups on topics identified on the spot
- Outcomes
  - List of topics of interest for the ESFRI Forum D4.3
  - Preparation of Sydney IVOA meeting
  - Status of multi-D standards
  - HiPS strategy



## D4.3 First ESFRI Forum & Training Event

- 3-4 December 2015
- 25 participants, all DADI partners and ESO
- Centered on the ESFRI and pathfinder requirements
- Organised around the topics identified during the Tech Forum
  - Multi-dimensional data access
  - Time domain data access (<> WP5)
  - VO registry of resources
  - Authentication & Autorisation (WP3)
- VO standards and tools already in use
- Discussion on specific support activities initiated with the ESFRIs and pathfinders





# D4.2 First DADI School

- 15-17 December 2015
- Specific talk tomorrow morning



# IVOA

- DADI technological activities (Task 4.3) are performed in the IVOA context
- IVOA priorities well aligned with ASTERICS's
  - Get NOW the first set of multi-D standards as defined by the IVOA Standing Committee on Science Priorities
  - Time domain: no ASTERICS requirement on time domain for the moment (WP5 involved)
- Sesto (June 2015): presentation of DADI and work on multi-d standards
- Sydney (October 2015)
  - Many meeting highlights relevant to DADI
  - Well prepared by the Tech Forum
  - Significant impact of participants linked to DADI
    - At least one chair or co-chair of all the Groups present at the meeting except one
    - Significant participation in the Groups

# ADASS XXVth

- The place to be to discuss with large projects about their data
- ASTERICS talk (F. Pasian)
- IVOA talk (C. Arviset)
- VO ubiquitous in the talks and posters, very positively mentioned in B. Schmidt keynote address
- Two ASTERICS posters



## IVOA Provenance Data Model: Hints from the CTA provenance prototype

Michèle Ségouin<sup>1</sup>, Mathieu Scrufer<sup>2</sup>, Maricela Lopez<sup>3</sup>,  
François Dubillard<sup>4</sup>, Catherine Besson<sup>5</sup>, Johan Bergman<sup>6</sup>

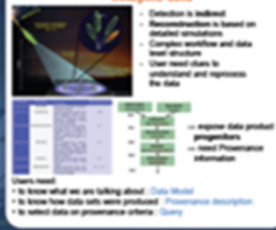
<sup>1</sup>ICST/INRAE, <sup>2</sup>SOAR/INRAE, <sup>3</sup>ADP/INRAE

**Abstract**

We present the last developments on the IVOA Provenance data model, mainly based on the W3C PROV concept. In the context of the Cherenkov astronomy, the data processing stages imply both assumptions and comparisons to established simulations. As a consequence, Provenance information is crucial to the end user in order to interpret the high level data products. The Cherenkov Telescope Array (CTA), currently in preparation, is thus a perfect test case for the development of an IVOA standard on Provenance information. Contrary to previous Cherenkov experiments (CTA will serve as an open observatory to a wide astrophysics community and high level data products), specific, key aspects will be made available through the Virtual Observatory. We describe general use-cases for the computational Provenance in the CTA production pipeline and explore the proposed W3C notations (PROV) format, as well as Provenance access solutions.

### Cherenkov Astronomy Context: Complex data

- Distortion is induced
- Reconstruction is based on detailed simulations
- Complex workflow and data level structure
- User need data to understand and reproduce the data




**Users need:**

- to know what we are talking about: **Data Model**
- to know how data sets were produced: **Provenance description**
- to select data on provenance criteria: **Query**

### Provenance Data Models


**W3C Provenance Data Model**

Simple and general concept to trace the activities performed on Entities by Agents.




**VO Provenance Data Model**

adapted to astronomy data and data processing activities.



### Provenance description

**INTEROPERABILITY: STANDARDIZED DESCRIPTION LANGUAGE PROV-O**



The W3C model does not include the concepts of Activity Collections or Instances. We get around the problem by describing on the one hand a workflow and on the second hand independently the activities that comprise it. The non-formality relation is used to indicate the beginning and end of the workflow.

### Provenance access


The objective is to enable user queries on Provenance information, in a standardized way.

Selection criteria could be:

- Name of attributes of the W3C or IVOA Provenance data model
- Name of the attributes specific to the CTA context (run number, arbitrary conditions, ...)

**Identified** - We need to identify the specific provenance terms for each data level.

The VO user need to query which specific attributes could be a criteria.



## IVOA Data access layer project

Bonnafant Francois, Dowler Patrick, Nockle Keith, Tody Douglas [COS/CADC, University Edinburgh, NRAO]  
Contact: [francois.bonnafant@cea.fr](mailto:francois.bonnafant@cea.fr)

### IVOA DAL Working Group Project and Objectives

A group formed at the very beginning of international VO efforts. Data is no longer produced in homogeneous:

- data discovery
- data retrieval
- data access processing

providing other data access, in dynamically composed subsets of service astronomical processes:

- data Model: query based on data Model concepts, responses are detailed instance of data Model
- Provenance: use of provenance for metadata
- Registry: standardized description and discovery of the services
- VOI: application made support

### Historical protocols

- Discovery**
  - simple search: simple SQL constructs for listing of records, VO/VOI support
  - single image browser: constraints and other additional parameters constraints
  - VO/VOI query responses for data discovery giving standardized image description, instance and Model: generation
  - single data browser
- Generation**
  - single image browser: the DAL has implemented a series of a comprehensive special data Model
  - VOI: interoperable resources make query services

### Simple Spectral Access

- Standard query parameters in description of data
- Additional spectral related parameters (WAVELENGTH, SPEED, CHANNEL, etc.)
- Query responses is a VO/VOI
- Organized in VO/VOI based from Spectrum Data Model packages: Model, Instance, Retrieval, Transformation, Constraints, etc.
- Model instance register with registry
- Virtual data generation
- services query color or image generate a "fast search" to query parameters "locate"

### TAP and ADQL

A standard protocol to describe and query relational tables.

ADQL: The active where expressions made is consistent with IVOA Cherenkov Monitor Services (example) Instance and Model registered with the DAL, VO/VOI and VOI/VOI descriptions: standardized metadata (query via ADQL, language provenance extension of SQL).

The DAL: standard data Model protocols appear as a TAP service to support uniform cross observatory data discovery.

### Recent or current developments of new protocols

- collection of service capabilities (parameters) to provide multi-dimensional data discovery and access (parameter priority)
- Calculus: Transformational methods to link resources to instant datasets
- SQL: parameterized query (PQL) of the OData 1.1 data model for simple interoperable data discovery
- Association: driving formal data processing for accepting information from the datasets

### Lessons learned from DAL development

- service use cases: often complex using a collection of services to a database of single
  - in general, one service center always clearly representative
- having clearer functionality
  - single purpose service capabilities that can be combined to support various use cases
- common patterns and features
- additional data access layer interface (DAL) for easy to use and maintaining
- Test a framework of protocols sets
  - standardizable: processes (more complex, more powerful) which align-model: processes (simple, implementation flexibility, stability)
  - open generation of Parameter Query Language/purpose
    - Application homogeneity: VOI/VOI

## Among the other activities

- Technical work on Provenance (VO/CTA)
- Collaboration between CDS and EGO on Aladin customization
- Publication of ANTARES data by GAVO
- Autorisation&Authentication
- Collaboration with EUDAT (i.e. with the « generic » data infrastructure)

## Next steps

- Discussion on specific support actions with the ESFRIs
- Cape Town IVOA meeting (May 2016)
  - Focus sessions on large projects, organised by M. Allen
  - VO standards and tools including multi-D and time domain – Convergence on multi-d mandatory
- RDA Plenaries March (Tokyo) and September (International Data Week, Denver)

# Next Deliverables

#	Title	Lead partner	Due date	Actual date and location
D4.4	Second ASTERICs DADI Technology Forum	UEDIN (Task 4.3)	March 2016	7-8 March Edinburgh <b>Us, here, now!</b>
D4.5	Second ASTERICs European School	CNRS-CDS (Task 4.2)	November 2016	Date TBC Strasbourg
D4.6	First European Data Provider Forum & Training Event	UHEI (Task 4.1)	November 2016	15-16 <b>June</b> Heidelberg + training 17 June

Next Tech Forum : March 2016 (D4.7)

**An additional Tech Forum before Trieste Interop?  
we need to liaise regularly on technological activities**

## Paths for support to ESFRIs/pathfinders

- Work to do for all the tasks
- Task 4.2: Strong interest in science training
- Specific information & training events for scientists in addition to participation of a few individuals into Schools
  - Collaboration needed to define and develop relevant tutorials
  - Can be coupled with project meetings, disciplinary schools, etc



# Paths for support to ESFRIs/pathfinders

- Tasks 4.1/4.3
  - Cubes
  - Time domain: transients & time information in data
  - Provenance
  - A&A (with WP3)
  - HiPS
  - Implementation CharDM, ObsDM
  - At some point: Visibility data
  - Important to provide (if possible) ranked list of topics and concrete examples to define the actions
  - Tech Forums (Hack-A-Thon), ESFRI Forum & Training
  - Specific focus working meetings can also be organised (e.g. Provenance)