Data Model Working Group

Laurent Michel, Jesus Salgado -- Bologna - May 2023

Data Modeling

Data modeling

Building an **accurate description** of scientific quantities and of the **way to combine** them

Data Modeling

Data modeling

Building an **accurate description** of scientific quantities and on the **way to combine** them

The IVOA coating

These accurate data descriptions must be **shareable** among different **stakeholders** (data providers, clients, standard editors)

WG Commitment

Core of the group activities

Coordination of the data modeling **activities** in the VO

- **Development** of new models
- **Adaptation** of existing models to new use cases

WG Commitment

Core of the group activities

Coordination of the data modeling **activities** in the VO

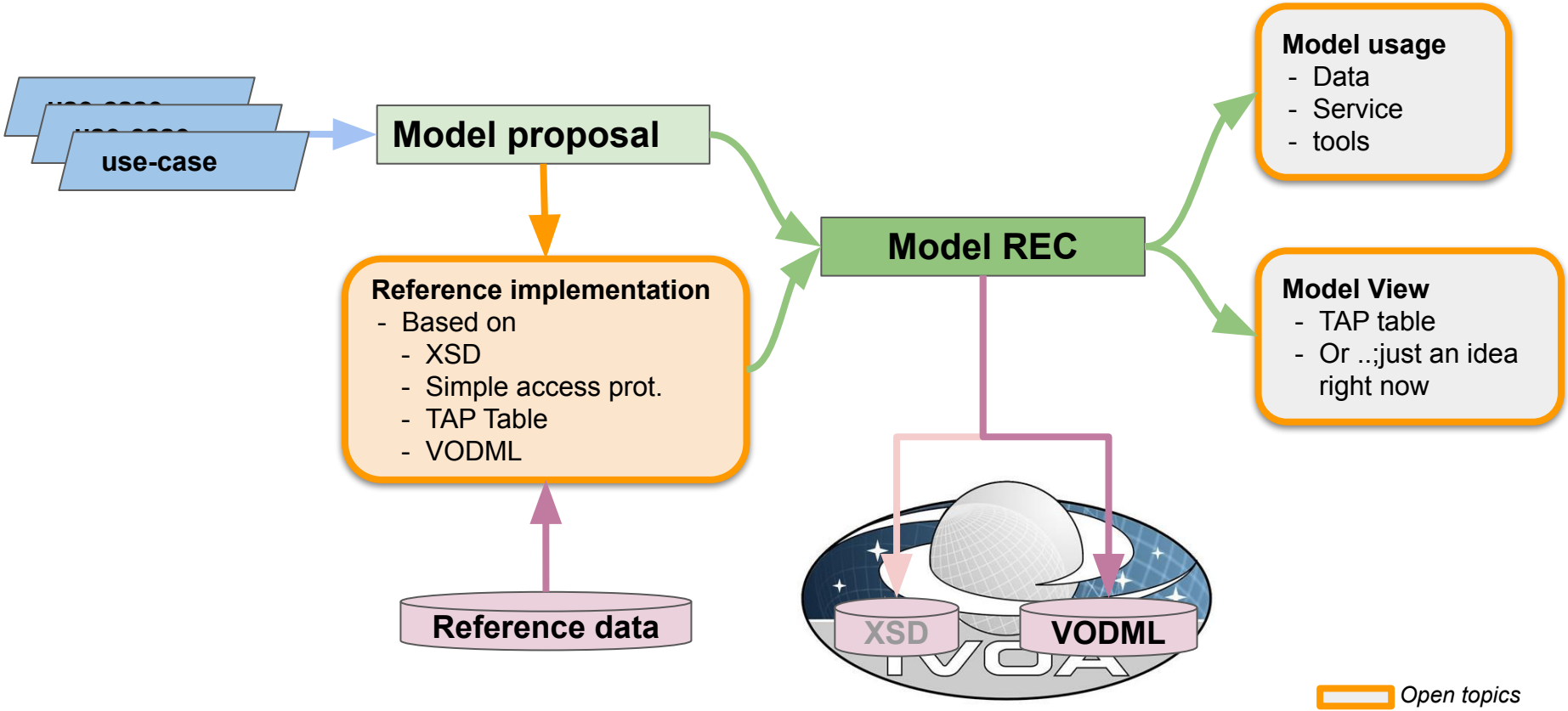
- **Development** of new models
- **Adaptation** of existing models to new use cases

After sales service

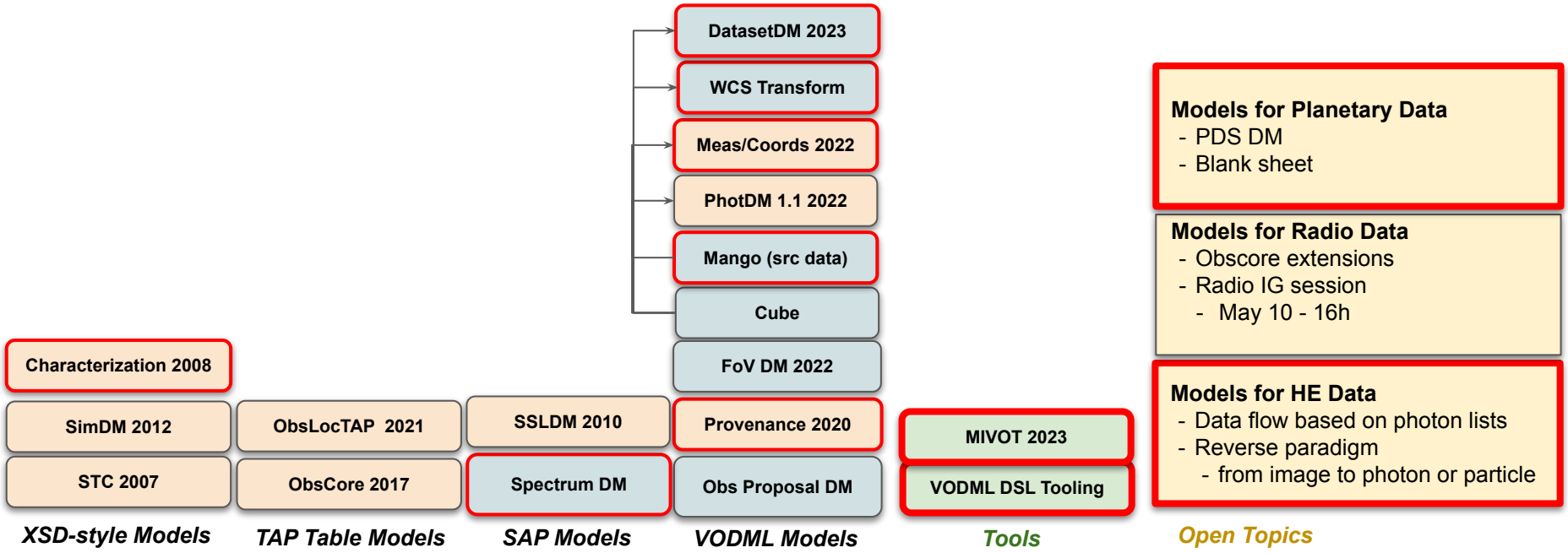
Provide a **framework** for the **DM integration** in the VO activities

- Identify **use-cases** for which the use of models is relevant
- How to bind **legacy data** with models
- How to **process** modeled **data**

DM Workflow


use-case
Model proposal
Reference implementation
- Based on
- XSD
- Simple access prot.
- TAP Table
- VODML
Reference data
Model REC
Model usage
- Data
- Service
- tools
Model View
- TAP table
- Or ..;just an idea right now
XSD
VODML
IVOA
Open topics

WG Landscape in 2023

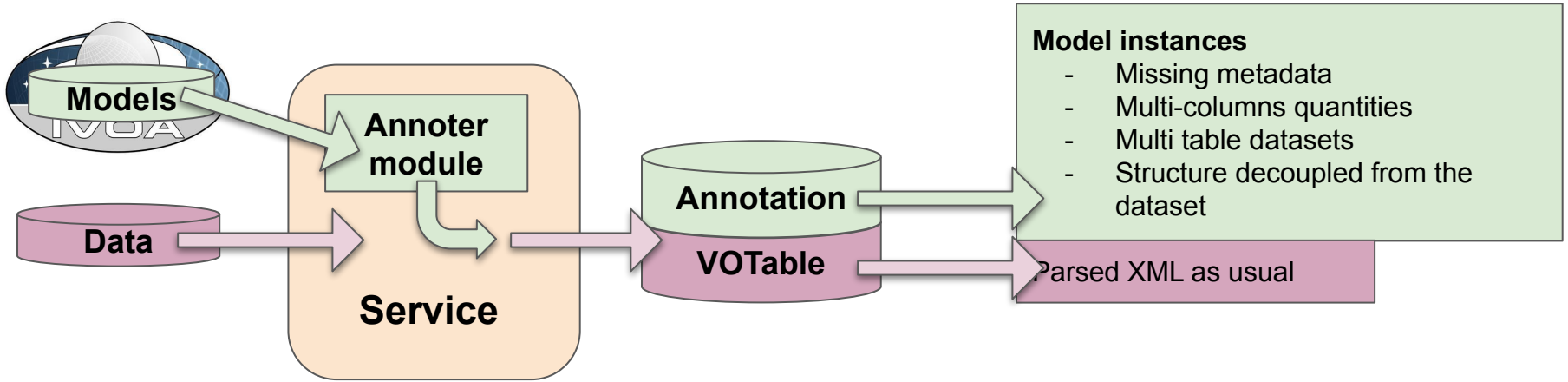


DM I session : The VODML Ecosystem

- **In 2016, the community accepted the principle of VO-DML**
  - A single serialization method for all models
  - VO-DML is an XSD/Schematron schema
- **Models can be handled by generic tools**
  - XSLT e.g. VODML to HTML
  - Code generators
  - Libraries
- **VODML tooling (Paul Harrison)**
  - **A very promising toolbox for handling models**

DM I session : The VODML Ecosystem

- **In 2017, the community agreed on the annotation baselines**
  - Define a syntax for connecting data and models in VOTables
  - This syntax had to be arranged in a way that does not alter the schema of VOTables
  - Must be in XML
  - Must comply with a schema independent of any particular model



DM I session : The situation now

- **In 2023, Model Instance in VOTable (MIVOT) voted by the Exec this Spring**
  - **Strong TCG requirement for guiding implementers**
    - Explain the process
      - How to annotate data
      - How to consume annotated data
    - Provide educational materials
      - Data sample
      - Jupyter notebooks
    - Provide code
      - MIVOT enable PyVO PR
      - MIVOT enable VOLLT extension
    - Complete the model for Archival Data (MANGO)
      - Call for idea/contribution

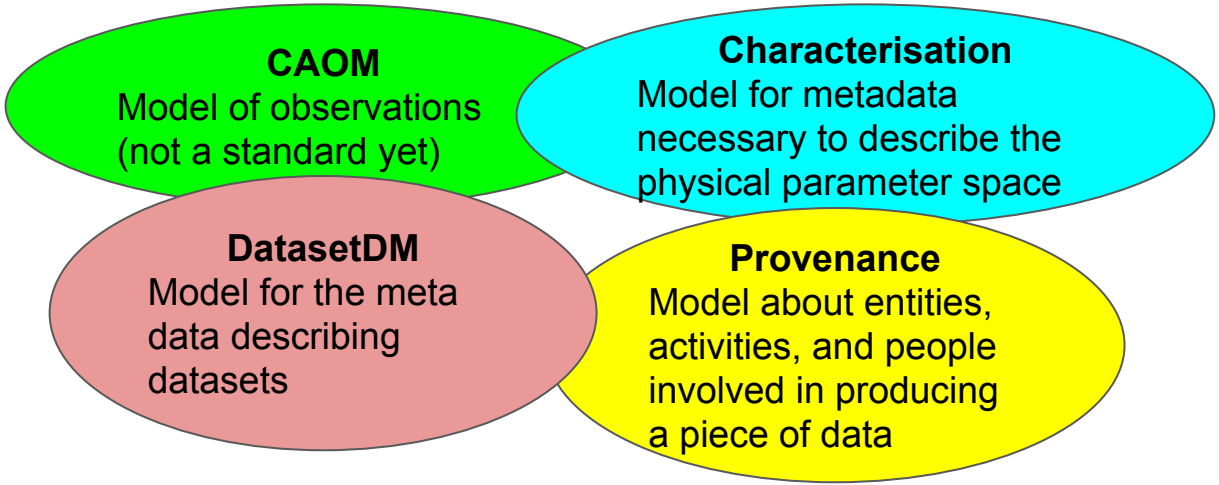
DM I session

DM1 session is focused on things related to VODML and annotation. A large part of the session is open for discussion.

DM I - Tuesday, May 09, 16:00 -- 17:30, Plenary Room		
Speaker	**Title**	**Abstract**
Paul Harrison	Adopting the new VO-DML tooling	A demonstration will be given of how it is possible to use the new tooling with various levels of adoption of new workflow patterns, from still working in the same fashion as previously using UML tooling such as Modelio, to writing models in VODSL. In both cases it is possible to use the source code generation facilities.
MIVOT Session		

Lead by Laurent Michel	DM workflow	Session open for discussion on a roadmap leading to concrete implementation using models.
		
		The talk agenda is not finalized yet but the following topics will be presented
		
		• Reminder of the WG strategy:
		  ◦ Historic reminder
		  ◦ Building component models (MCT/PhotDM)
		  ◦ Design of an annotation syntax (MIVOT) which is model-agnostic, which works with any data arrangement and which is isolated into the VOTable
		  ◦ Design of models aggregating components to annotated archival data (MANGO) or science products (Cube)
		• Plan for MIVOT implementations
		  ◦ Cookbook
		  ◦ Client side: Astropy/PyVO
		  ◦ Server side: VOLLT extension
		• Mango draft:
		  ◦ MANGO is an open model designed to improve the description of archival data by providing mechanisms to connect various columns together for building complex entities and to enhance mapped quantities with some semantic, coordinate systems and coverage information.

Splinter: Global Dataset Description



Some overlaps the need to be consolidated

Splinter: Global Dataset Description

The splinter has no agenda, this is a session open for open discussion (see the Wiki page).

Splinter - Tuesday, May 09, 18:00 -- 19:30, Plenary Room		
Supervisor	**Title**	**Abstract**
Mark Cresitello Dittmar	DatasetDM, Provenance, CAOM and Characterization	Now that the core models of the Cube family (Meas/Coords) are completed, we return our attention to the next group. The information in this model is a consolidation of content from several of the early core models (ObsCore, Spectrum, Characterization). The descriptions of many of the elements are derived from the Resource Metadata standard. The idea of this model is to centralize this information for other datamodels to use (Cube, Mango, Spectrum2, TimeSeries).

DM II session: Ongoing Model and Perspectives

DM2 session contains miscellaneous DM related talks.

DM II - Thursday, May 11, 11:00--12:30, Plenary Room

Speaker	Title	Abstract
Steve Hugues	The PDS4 Information Model - An Implementation-Agnostic Model for Interoperability	The PDS4 Information Model was developed by the Planetary Data System as a science data archive standard to improve interoperability within the planetary science community. It addresses the key requirements for interoperability including standardized data formats, common data models, clear data definitions, and well-defined data governance. In addition, the information model was developed independent of all system implementation choices to insulate it against inevitable changes in implementation technology. It also uses multi-level data governance to localize the impact of changes in the science disciplines. These architectural choices have allowed the PDS4 Information Model to remain relevant within the Planetary Science Community while enabling interoperability across diverse science disciplines, tools, and APIs. It has been adopted world-wide by space agencies involved in Planetary Science. This talk will briefly describe the architectural and design principles used to maintain the independence of the PDS4 Information Model and how the artifacts necessary for the maintenance and operations of the PDS are generated.
Mark Cresitello Dittmar	Model status	This presentation will have 2 components: 1. Spectrum 1.2: Update on the enhancement request for the Spectrum data model and its readiness for RFC 2. With Measurements and Coordinate models in REC, we turn our attention to the remaining models required for representing NDCube data. This portion of the presentation will review the remaining models, with a focus on Dataset and Transform; their current state, open issues and implementation status.
Mathieu Servillat	CTAO Data Model group	
Mathieu Servillat	DM for High Energy astrophysics	
Mathieu Servillat	One step Provenance	We propose a simplified structure to describe the provenance of an entity as a succession of steps, based on the IVOA Provenance Data Model. With such a structure, the "last-step" provenance of an entity may be stored as a flat list of attributes inside the entity (e.g. with keywords in the header of a FITS file), as a separate file or in a relational database. By iterating the request of provenance step by step, one can reconstruct the full provenance of an entity.

If you want to join the WG

- WG Wiki page https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaDataModel
- SLACK IVOA#data-models
- MIVOT RFC https://wiki.ivoa.net/twiki/bin/view/IVOA/DataAnnotation
- Follow us on GitHub https://github.com/ivoa-std/
  https://github.com/ivoa/
- Mailing list dm@ivoa.net