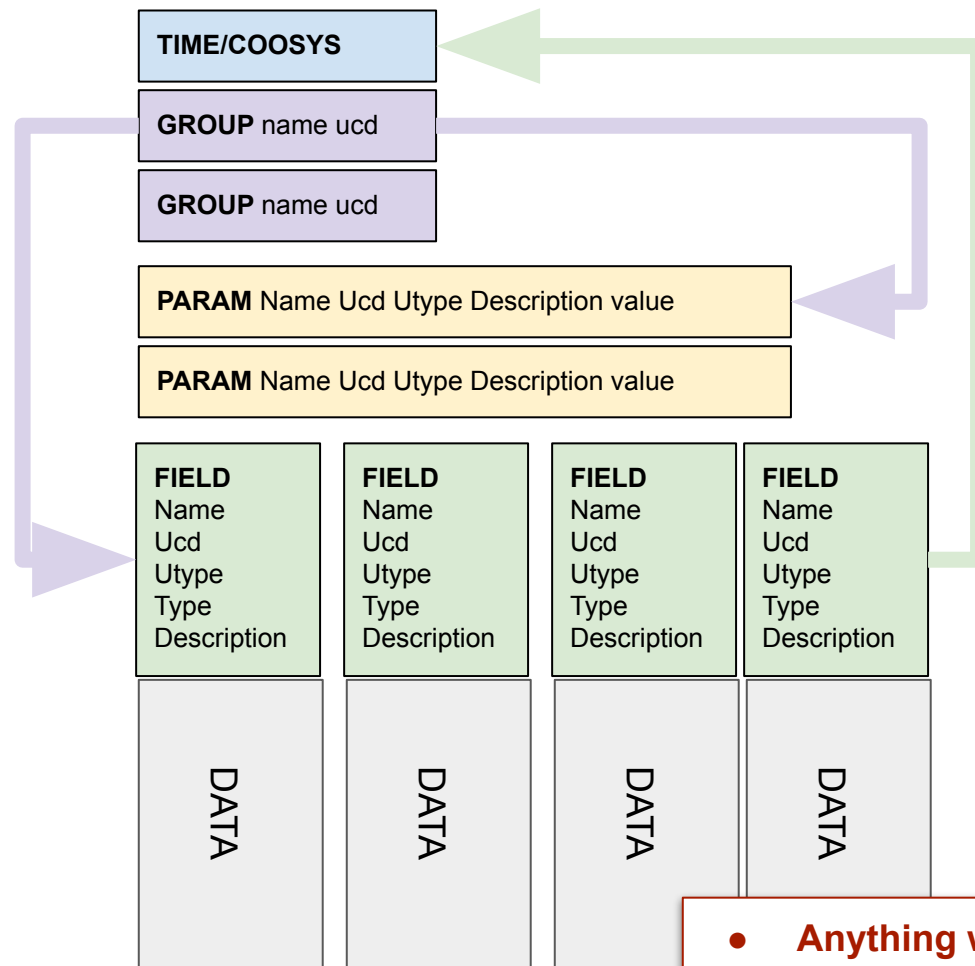# How to use Data Models to Enhance **Interoperability** in the VO

Data Interoperability

- The dataset contains all information required by clients to complete the processing
  - Display
  - Plot
  - Filtering
  - X match
  - Data sharing
  - ...

- Client do not need to run provider-dependant code

VOTable at a Glance

TIME/COOSYS

GROUP name ucd

GROUP name ucd

PARAM Name Ucd Utype Description value

PARAM Name Ucd Utype Description value

FIELD Name Ucd Utype Type Description	FIELD Name Ucd Utype Type Description	FIELD Name Ucd Utype Type Description	FIELD Name Ucd Utype Type Description
DATA	DATA	DATA	DATA

- **Catalog data are well supported by VOTables**
  - VOtable has been designed for this purpose!

- **Data column well defined**
  - UCD, format unit, utype
  - Global parameters (PARAM)

- **Possible column associations**
  - GROUPs

- **Coordinate Systems**
  - Time/space

- **Human readable information**
  - INFO
  - DESCRIPTION

- **Plenty of others features to discover within the 32 pages of the standard**

- **Anything we need to get a good interoperability level!**
- **Do we need more?**

3

Mapping Models in VOTables

#	Interoperability Feature	VOTable
1	Quantities well defined	**Good**
2	Role of the quantities well defined	**Poor** due to UType flaw
3	The way quantities are nested	**Poor** group are little used
4	The way quantities relate each to other	**Very poor** no role - no multiple ref

DMs can help

- The lack of clean role definitions

```
<PARAM datatype="double" name="dec"
  ucd="pos.eq.dec" value="1.18583048057467" ref="system"/>
```

Ref to what?

- Some UCDs miss context (related to)

?????

```
–<FIELD name="TYC2" ucd="meta.id.part;meta.main" datatype= short" width="5">
    <!-- ucd="ID_MAIN:2" -->
    <DESCRIPTION>[1,12121] TYC2 from TYC or GSC (1)</DESCRIPTION>
  </FIELD>
```

Using a Specific Mapping Block

- **Adding something to VOTable**
  - **Improve interoperability**
  - **Not breaking existing stuff**

Model
- Quantities definition
- Description of the roles played by all quantities
- Description of the way quantities are nested
- Description of the way quantities relate each to other

Mapping block
- Bridge binding model(s) and data
- Add missing data description elements

TIME/COOSYS

GROUP name ucd

GROUP name ucd

PARAM Name Ucd Utype Description value

PARAM Name Ucd Utype Description value

FIELD	FIELD	FIELD	FIELD
Name	Name	Name	Name
Ucd	Ucd	Ucd	Ucd
Utype	Utype	Utype	Utype
Type	Type	Type	Type
Description	Description	Description	Description
DATA	DATA	DATA	DATA

5

Agree, Disagree

- **Consensus points**
  - WE need a better description of science (meta) data
  - We have to work with existing models when available (Measure and Coordinates PhotDM)
  - Working first with VOTables which are the most complete data containers in term of meta-data
  - We have to design a mapping structure that can be added within VOTables without altering the original content
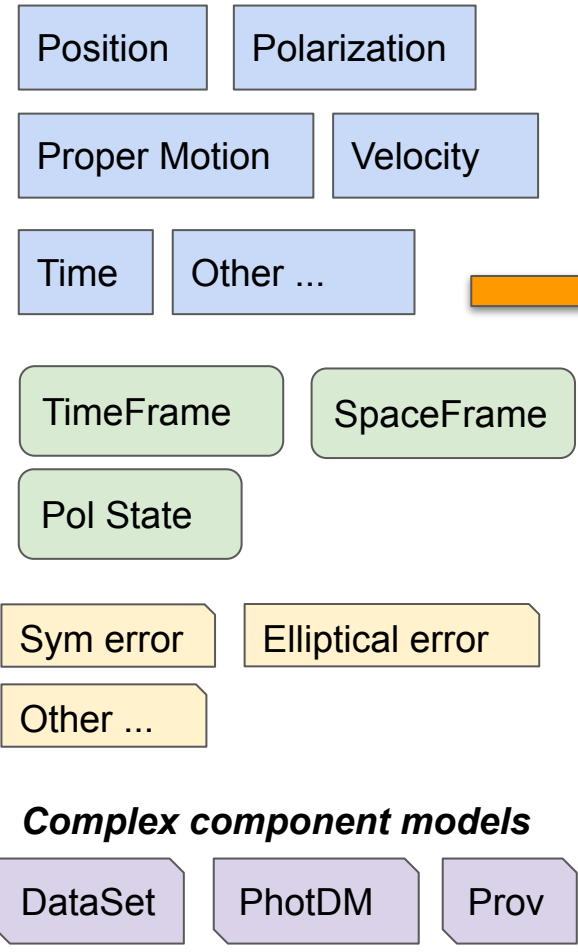
- **Contention points**
  - What is the most appropriate mapping syntax?

Modeling Levels

Model Fragments
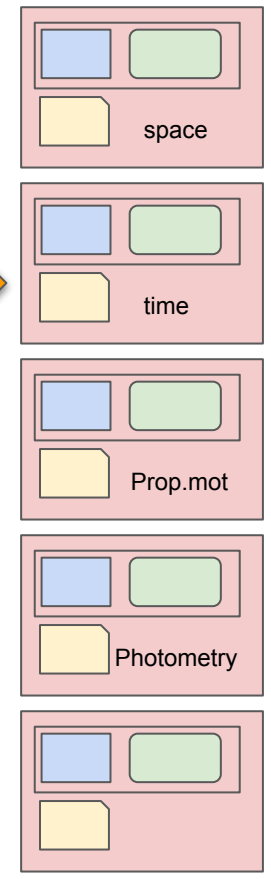Description of elements that constitute the measures
- Coordinate classes
- Measure classes

Position Polarization

Proper Motion Velocity

Time Other ...

TimeFrame SpaceFrame

Pol State

Sym error Elliptical error

Other ...

Complex component models

DataSet PhotDM Prov

Model Components
Complete description of the measures
- Measure model
- MANGO parameter

space

time

Prop.mot

Photometry

Integrated Models
Description of science products
- One product = one model instance

DataSet *Time Series*

Points

time Photometry

Point

Object ID

MANGO property dock

Agree, Disagree
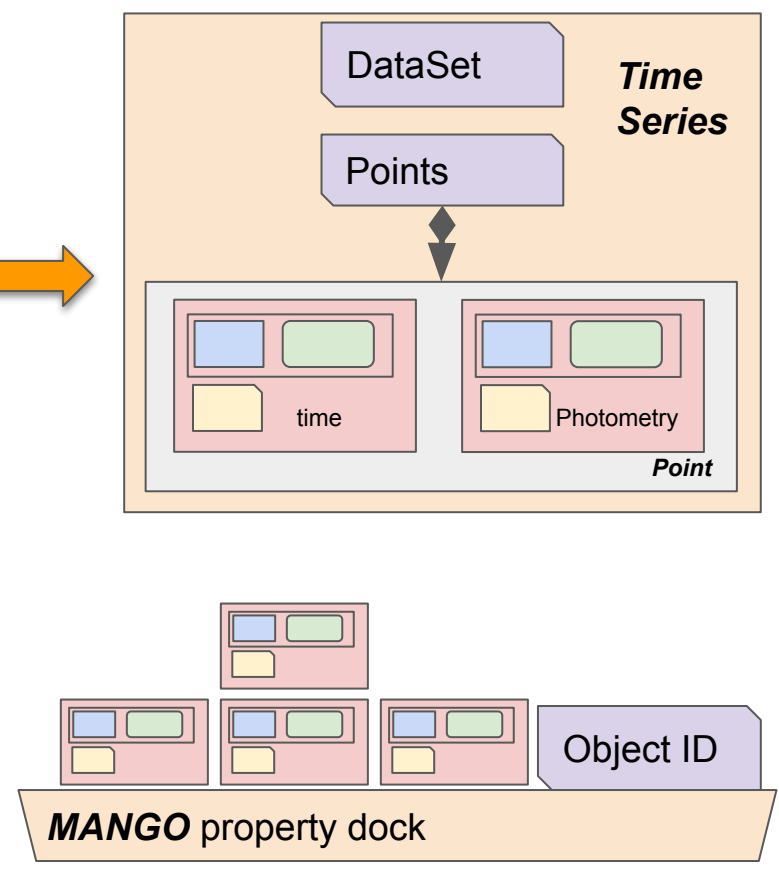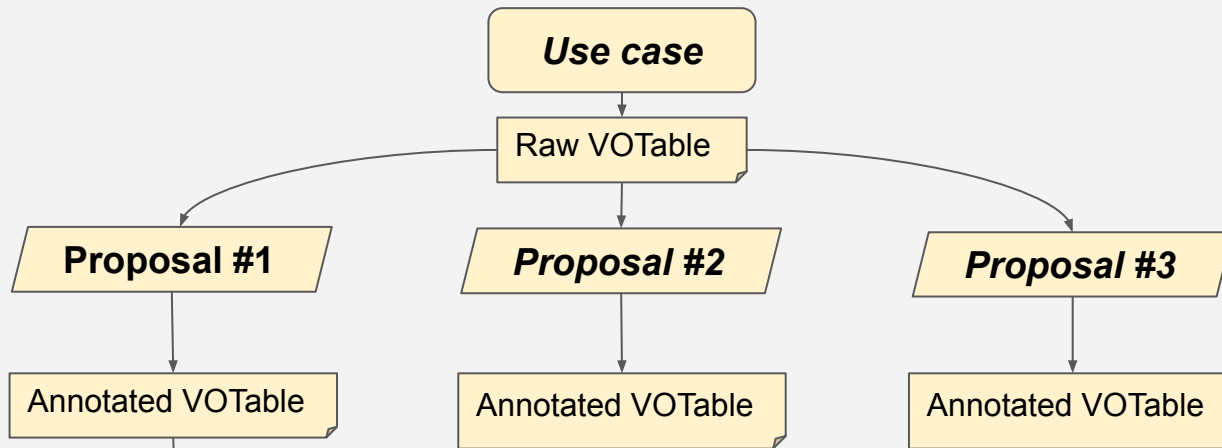
- **Consensus points**
  - WE need a better description of science (meta) data
  - We have to work with models for physical properties
    - Using the existing ones when available (Measure and Coordinates PhotDM)
  - Working first with VOTables which are the most complete data containers in term of meta-data
  - We have to design a mapping structure that can be added within VOTables without altering the original content

- **Contention points**
  - What is the most appropriate mapping syntax?
  - Do the available models fit our needs?
 - Which modeling level(s) must be used

DM Workshop Process

Use case → Raw VOTable

Proposal #1 → Annotated VOTable

Proposal #2 → Annotated VOTable

Proposal #3 → Annotated VOTable

community

Evaluation

Presentation/Discussion

Consensus and roadmap

Step #1
Feb-April 2021
- Work on use-cases
- Open discussion on Github
- 200 posts

https://github.com/ivoa/dm-usecases

Step #2 pre-interop meeting
14-17 May 2021 ?
- Match use cases implementation with science cases
- **target scientists/implementors who may consider volunteering to relate their experiences at the Interop Workshop.**
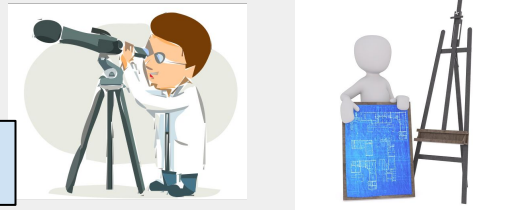
Step #3 Interop workshop
24-28 May 2021 ?

Step #4 Post-interop
- conclusions

9

This Meeting

- **Meeting Objective**
  - Teach scientists/implementers/providers about the ongoing work on the different proposals

- **Contributors will present feedback on their work on use-cases**
  - Work completed so far on a selection of 5 use-cases
  - Vision of where a more mature version of the use case could end up someday.

- **What we expect for/from the audience**
  - We hope to be able to provide a clear view on the issues.
  - To give necessary inputs for people who might consider volunteering to relate their experiences at the Interop Workshop

Use case #1 Standard properties

- **Simple tables with independent properties**
  - Use-case:
    - Providing end-users with a complete view on the data
  - Data source
    - Vizier, Chandra and XMM-Newton
  - Challenge
    - Getting the appropriate coordinate systems
    - Providing a good description of unusual data (High Energy)
  - https://github.com/ivoa/dm-usecases/tree/main/usecases/standard_properties

QUANTITY	COORDINATE SYSTEM
Position	Space frame
Flux	Energy band
Hardness ratio	Energy bands (high and low)
Exposure time	Time unit
Observation time	Time frame
Quality flag	Set of accepted values
Proper Motion	Space Frame

Use case #2 Time Series

- **Time Series**
  - Use-case:
    - Plotting time series
  - Data source
    - Gaia (gavo): simple time series
    - Gaia (DPAC): Photometric points in different bands mixed in one table
    - ZTF: Photometric points of different objects mixed in one table
  - Challenge
    - Discovering both independent and dependent axes
    - Processing table containing mixed data
  - https://github.com/ivoa/dm-usecases/tree/main/usecases/time-series

QUANTITY	COORDINATE SYSTEM
Time stamp	Independent axis - Time frame
Magnitude/Flux	Dependent axes - Filter description

Use case #3 Precise Astrometry

- **Simple tables with independent properties**
    - Use-case:
        - Getting the most accurate positions at a given epoch
        - Cross-match
    - Data source
        - Synthetic data table (derived from Gaia)
    - Challenge
        - Restoring matrix errors binding multiple parameters
    - https://github.com/ivoa/dm-usecases/tree/main/usecases/precise_astrometry

QUANTITY	ERROR	ERROR CORRELATIONS
Position	2D error + covariance	ProperMotion/Parallax/RadialVelocity
Proper motion	2D error + covariance	Position/Parallax
Radial velocity	1D error	Position
Parallax	1D error	Position/ProperMotion

13

Use case #4 Column Grouping

- **Column Grouping**
  - Use-case:
    - Showing to end-users additional data attached to one given quantity
  - Data source
    - III/21/gcrv Vizier catalog
  - Challenge
    - Make the column associations machine readable
  - https://github.com/ivoa/dm-usecases/tree/main/usecases/column_grouping

QUANTITY	ROLE
VR	Radial velocity
q_VR	Quality flag of the radial velocity
o_VR	Number of plates used to compute the radial velocity
r_VR	Abbreviation of the observation

Use case #5 Combined data

- **Combined Data**
  - Use-case:
    - Providing users with one VOTable containing sources with associated data
      - Source + detection + counterparts + spectra + …..
    - Complex data transport
      - Registry records
      - Simbad objects
      - Bundle of source data
  - Data sources
    - XMM: sources + spectra
    - Chandra: Sources + detections
  - Challenge
    - Restoring association between data spread over multiple tables
  - https://github.com/ivoa/dm-usecases/tree/main/usecases/combined_data

Agenda

- Workshop Introduction (Laurent MICHEL) - 15'

- VODML Overview (Gerard Lemson) - 10'

- Proposals
    - Mark Cresitello Dittmar - 20'
    - Markus Demleitner - 20'
    - Laurent Michel - 20'
    - François Bonnarel - 10'

- Q/A - 15'

- Wrap-up - 10'

wrap-up

DM session

- **26/05 15h UTC**
  - DM WS session #1: Mapping strategy

- **27/05 15h UTC**
  - DM WS session #2: Model review

- **28/05 15h UTC**
  - DM WS session #3: VOTable mapping syntax

Volunteer to contribute?

laurent.michel@astro.unistra.fr

backup

Interoperability Landscape

Data Interoperability
- The dataset contains all information required by clients to complete the processing
  - Display
  - Plot
  - X match
  - Data sharing

- Client do not need to run provider-dependant code

Modeling data == Building data descriptions
- that can be shared by different stakeholders and
- that are independent from any particular data provider
- That are independent of any particular serialization or container

This is why we are here

Models must provide:
- A clear definition of all modeled quantities
- A clear description of the roles played by all modeled quantities
- A clear description of the way quantities are nested
- A clear description of the way quantities relate each to other

DM Workshop: Timeline

- **February 2021-> April 2021**
  - Git repository open: https://github.com/ivoa/dm-usecases
  - Announcement to the community
  - Contributions / Discussions

- **May 2021: Pre-interop meeting**
  - The proposals documented on the repo are presented to community
  - Expect to involve people who are not part of the current discussions yet but who volunteer to match the proposals with their expectations and to report at interop.

- **May 2021: Interop workshop**
  - 3 focus sessions
    - Summary of the process
    - Scientific vision on this problem
    - Proposal outlooks
    - Discussion

- **June 2021: Post-interop wrap-up**
  - Not setup yet
  - Should end with clear conclusions and a roadmap