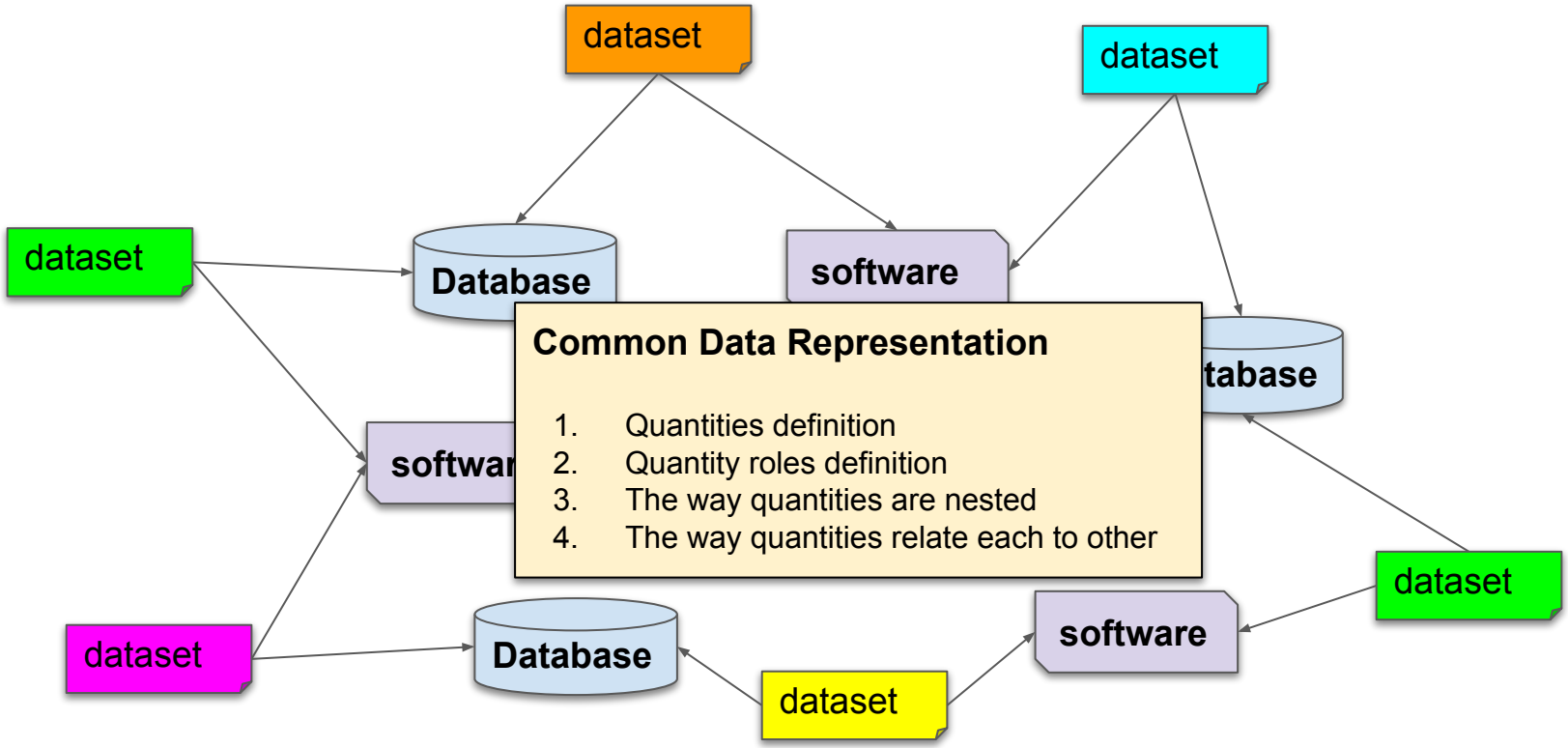
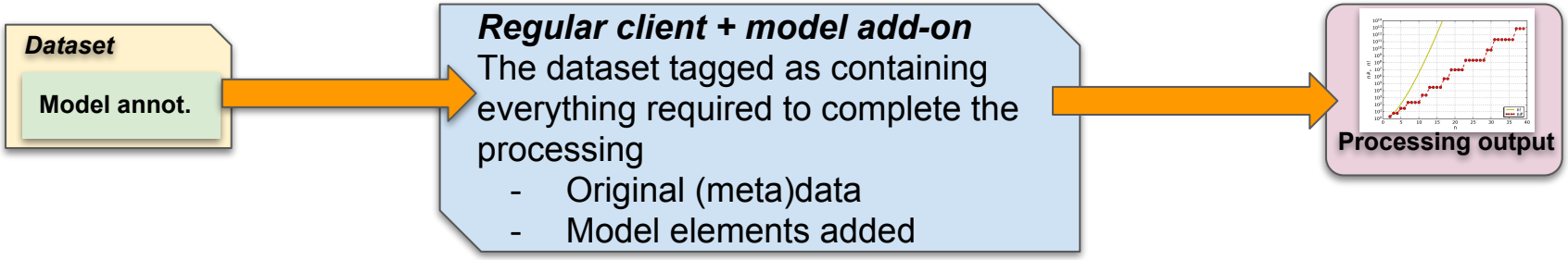
Using Data Models to Get a Better **Interoperability** in the VO

Interoperability Landscape



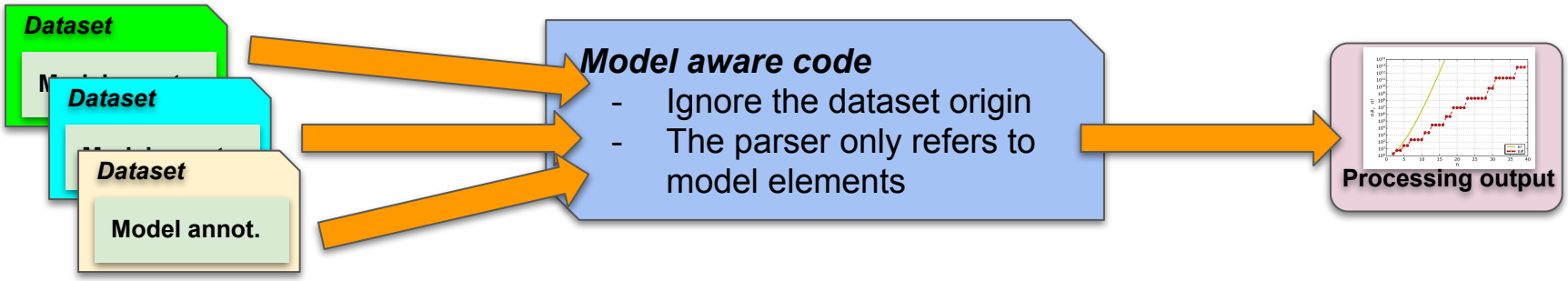
Modeling data <> Making Data models
Building data descriptions that can be shared by different stakeholders and that are independent from any particular data provider

Interoperability Pattern



level 1
Close to what is actually done by current apps

Interoperability Pattern



Level2: (abstract data access layer).

What is in this dataset?

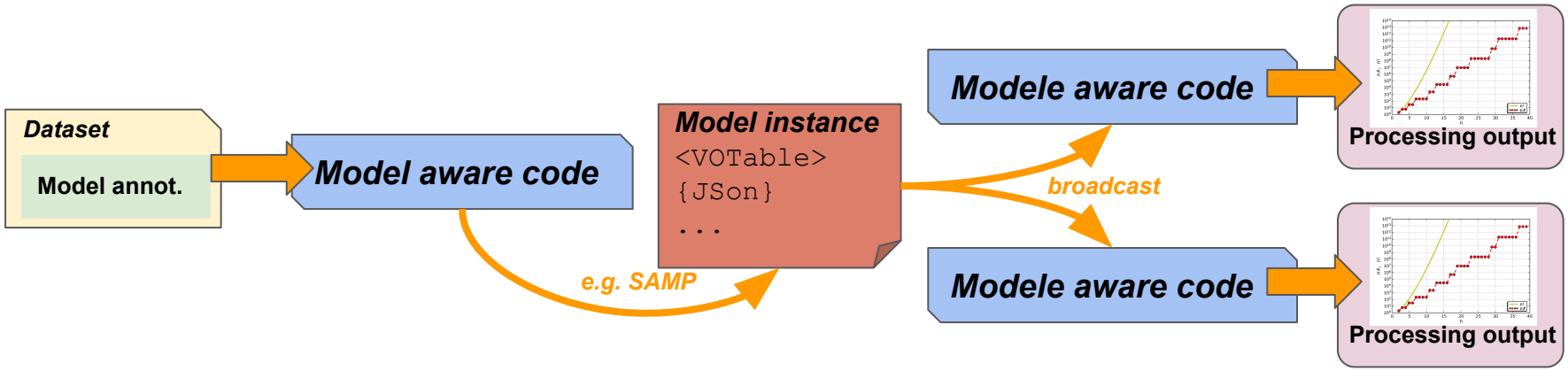
A Time Series as by Cube1.0

OK, Let me process it

Model aware code

- Not a crucial advantage for the simplest cases
- But should be very interesting for complex quantities (orbiting system, combined data ….)

Interoperability Pattern



Level3: Data exchange

This table row is very interesting

Let's broadcast it as a JSON MANGO instance

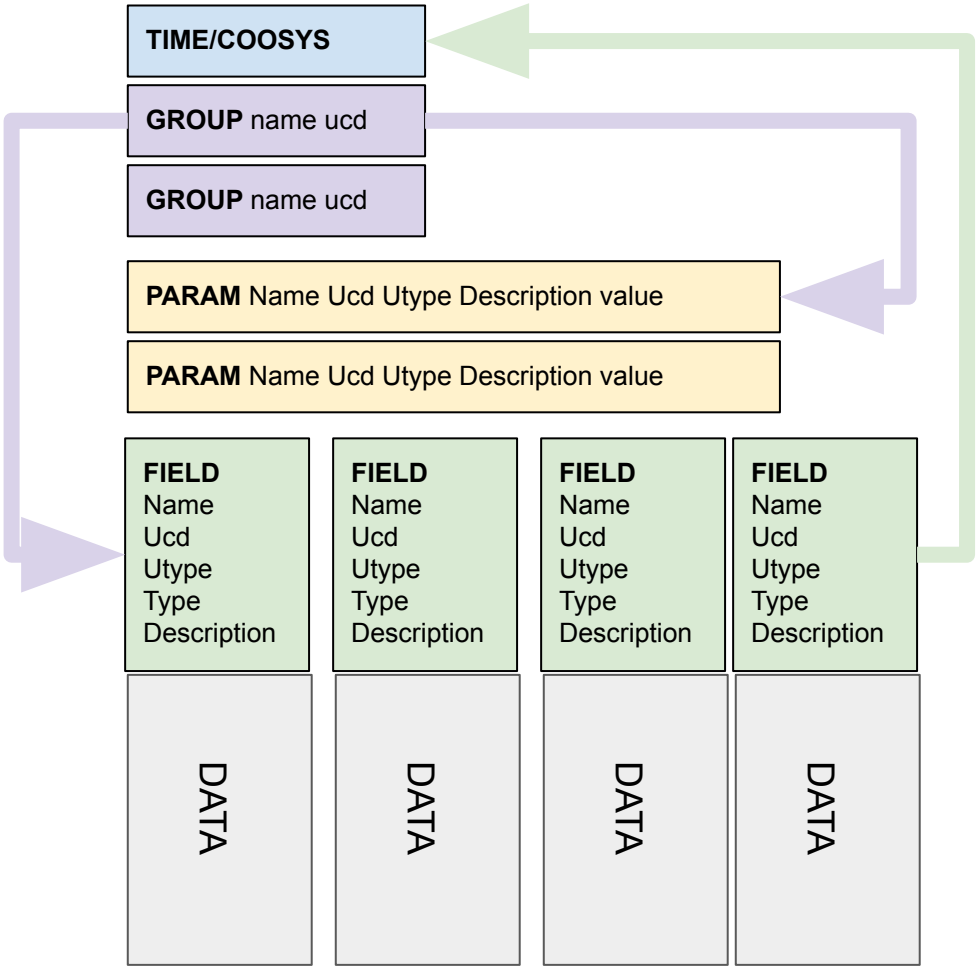
Got it

Me too

Agree, Disagree

- **Consensus points**
  - Getting a better interoperability for science data
  - We have to use the available models for physical properties
    - Measure and Coordinates
    - PhotDM
  - Working first with VOtable which the the most complete data container in term of meta-data
- **Contention points**
  - We should also consider other data containers than VOTables

VOTable at a Glance



- **A VOTable is a container**
  - The VOTable schema can validate the XML structure of the container
  - It cannot validate the content of the container
  - It cannot validate that the data map on a given model

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
4	The way quantities relate each to other	Very poor no role - no multiple ref

- 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>
```

?????

- No clear rule to build UTypes nor context information

```
-<FIELD name="obs_id" datatype="char" arraysize="128*" ucd="meta.id" utype="obscore:DataID.observationID">
   <DESCRIPTION>internal dataset identifier</DESCRIPTION>
 </FIELD>
```

```
-<FIELD name="s_resolution" datatype="double" ucd="pos.angResolution" unit="arcsec" utype="obscore:Char.SpatialAxis.Resolution.refval.value">
   <DESCRIPTION>typical spatial resolution</DESCRIPTION>
 </FIELD>
```

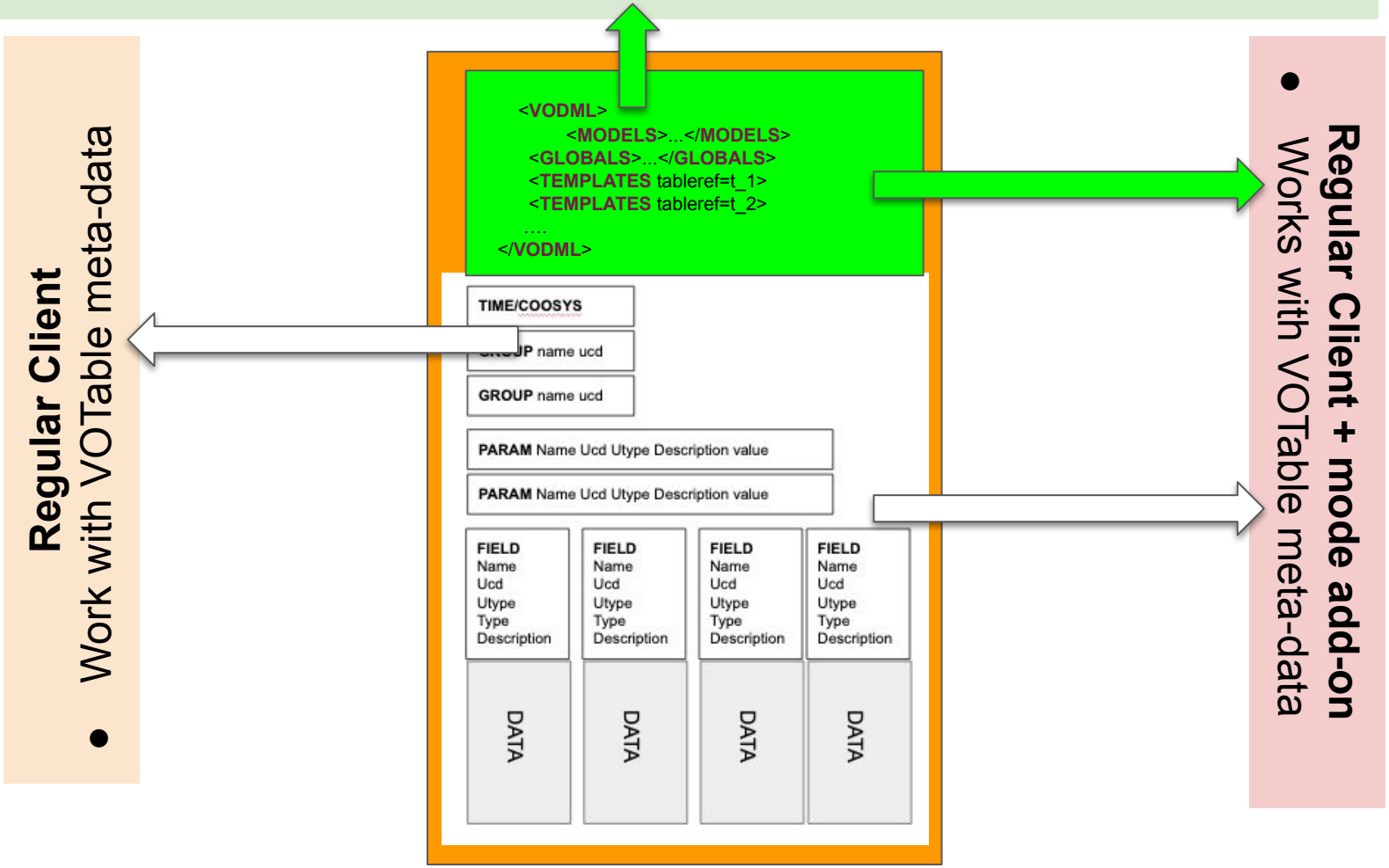
Mapping Models in VOTables: 2 Strategies

- **We need to add something to the VOTable to reach our interoperability**
  - **OPTION #1: Extending the usage of the GROUPs**
    - Adding groups to get a tight coupling between data and models
    - Possible in theory
      - Exercised with complex datasets (François Bonnarel)
    - Invasive annotation process
      - Mix up of original GROUPs with model-related GROUPs
  - **OPTION #2: Using a specific mapping block**
    - Not constrained by the VOTable schema
      - Free syntax
      - Better expressivity
    - Annotation process a bit easier
    - Can map data on multiple models
    - Can be ignored
      - Don’t break a running thing

Using a Specific Mapping Block

Model Aware Client

- Only sees the mapping block
- The model knowledge is sufficient to process any table instance

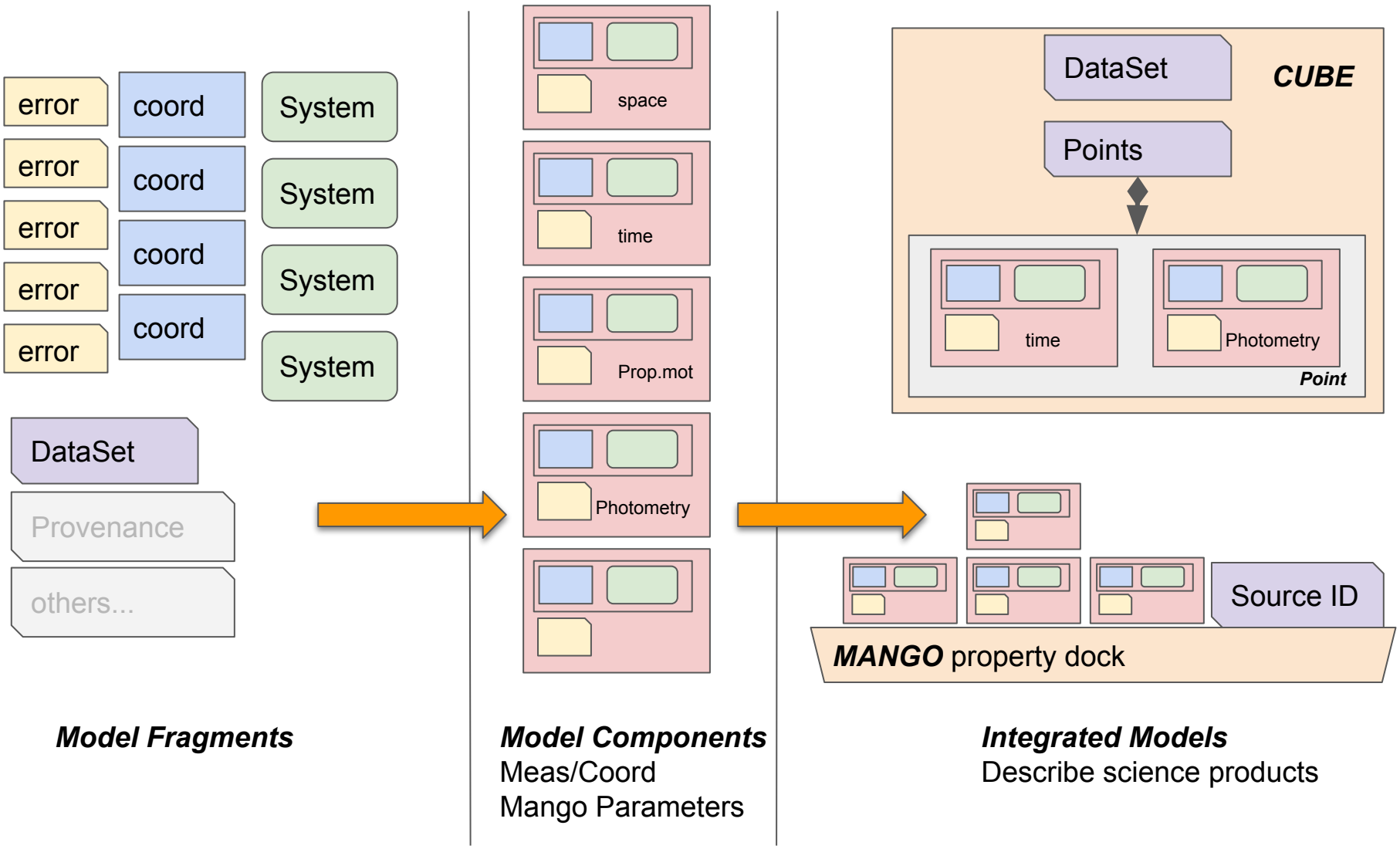


Agree, Disagree

- **Consensus points**
  - Getting a better interoperability for science data
  - We have to use the available models for physical properties
    - Measure and Coordinates
    - PhotDM
  - Working first with VOtable which the most complete data container 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**
  - We should also consider other data containers than VOTables

Bottom-up Model Building

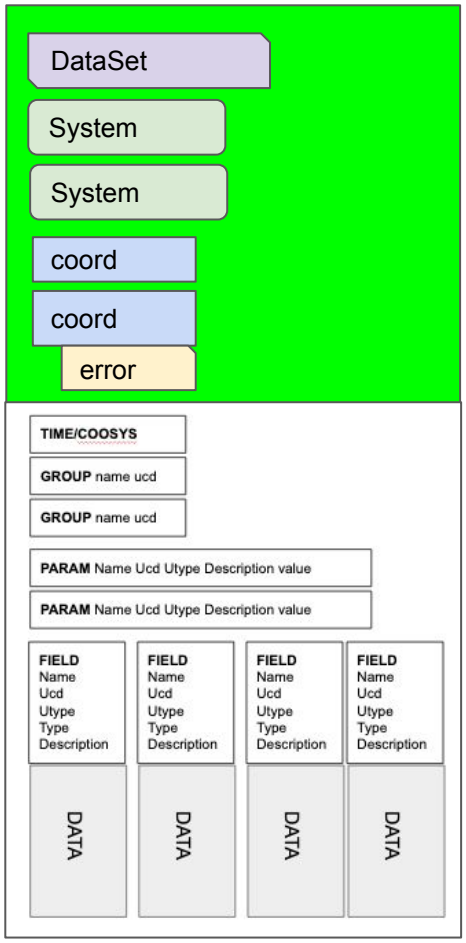


Model Fragments

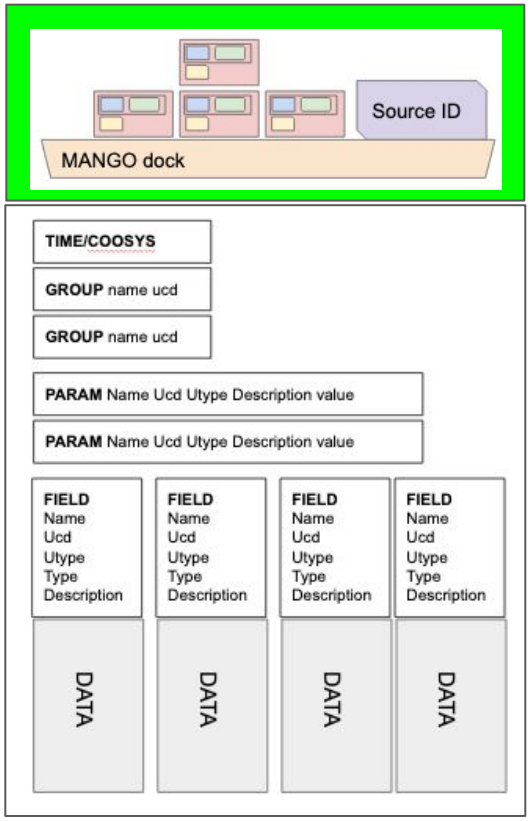
Model Components
Meas/Coord
Mango Parameters

Integrated Models
Describe science products

2 Mapping Approaches



Mapping based on small model components
The client is in charge of building its own model instance for any purpose



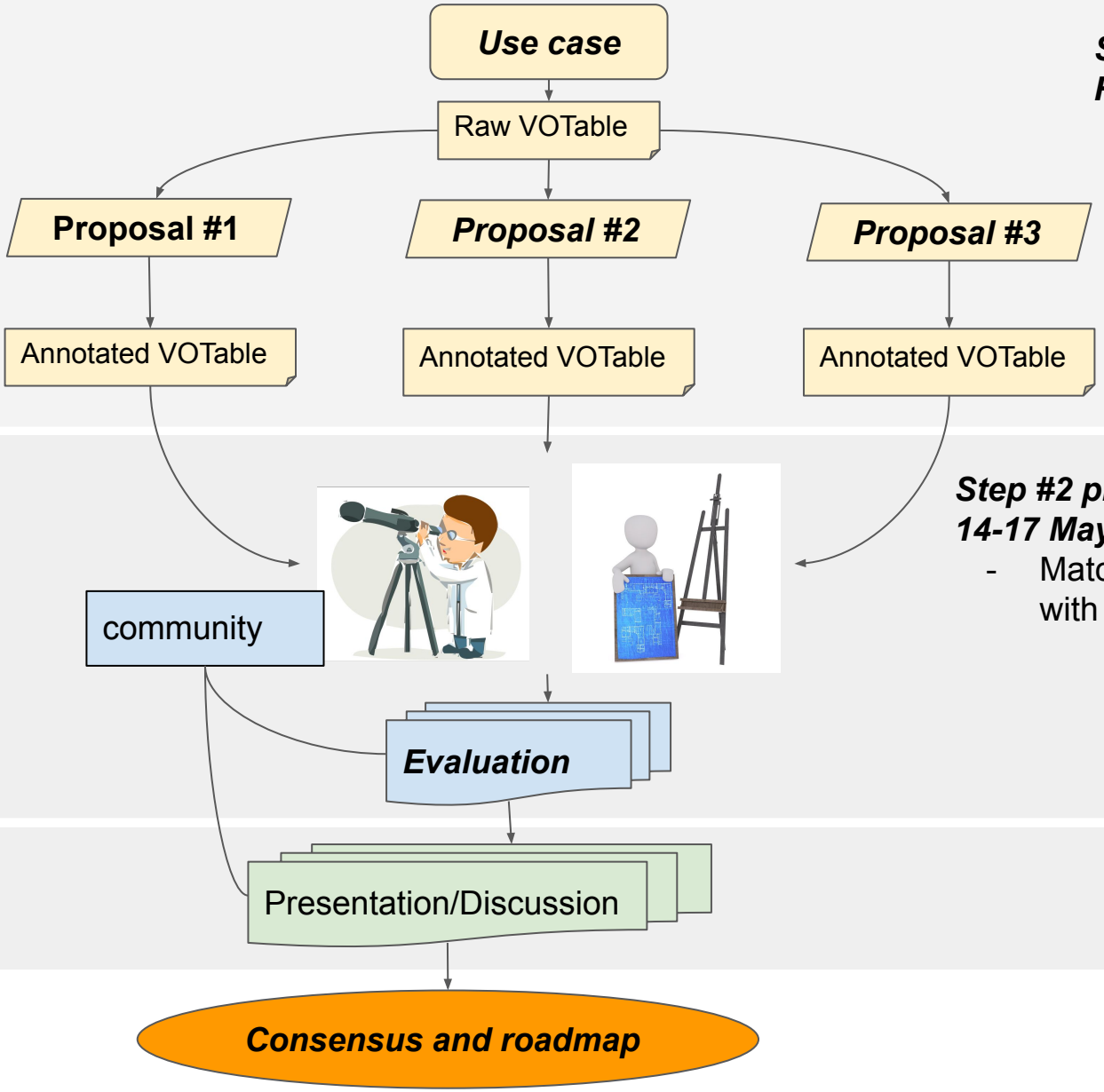
Mapping based on integrated model components
Mapping elements faith to the model structure

Agree, Disagree

- **Consensus points**
  - Getting a better interoperability for science data
  - We have to use the available models for physical properties
    - Measure and Coordinates
    - PhotDM
  - Working first with VOtable which the most complete data container 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**
  - We should also consider other data containers than VOTables
  - **Which mapping approach should we use?**

DM Workshop Process



Step #1
Feb-April 2021

- Work on use-cases
- Open discussion on Github

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

Step #2 pre-interop meeting
14-17 May 2021 ?

- Match use cases implementation with science cases

Step #3 Interop workshop
24-28 May 2021 ?

Step #4 Post-interop

- conclusions

DM workshop: Usecases

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

Usecase vs Project	XMM	Chandra	Vizier	GLAST	LSST	GAVO	Gaia	Simbad	Other project	Mock data
identity	C	C	C	C	C	C	C	R C		
native_frames			R C							
simple_position	C	C	C	C	C	C	C	C		
standard_properties	R C	C	C	C	C	C	C	C		
precise_astrometry			C		C		C			R
column_grouping			R C							
combined_data	R C	C	C	C	C	C				
complex-shaped-object				R C						
orbital-system				R C						
time-series	C	C	C		C	R C			ZTF: R C	

- **R** Raw data provided
- **C** Concerned by the use-case

DM Workshop: Status

- **Preliminary question**
  - Are the proposed models able to cover the science cases?
    - Components models: Measures Coordinates
    - Integrated Models: Cube Mango

- **Recurrent topics in the discussions**
  - Do we have to limit the exercise to simple VOTables (one <TABLE> well documented)
  - Do we have to anticipate the annotation scheme for complex quantities (radio, combined data)
  - Do we have to consider to map data associations that can be retrieved with WEB services
  - Do we have to consider working with model serializations out of the VOtable context?
  - How is the data annotation scheme robust against major model changes?

- **Background Questions**
  - Which is the most appropriate syntax
    - *VODML mapping* vs *ModelInstanceInVot*

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.

- **Ma**
  -

- **Everyone** is welcome to **contribute**
- We are particularly interested in contributions from people involved in **domains** that are **not well represented** in our cases e.g.
  - Planetary data
  - Radio data
  - High Energy
  - Data centers

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