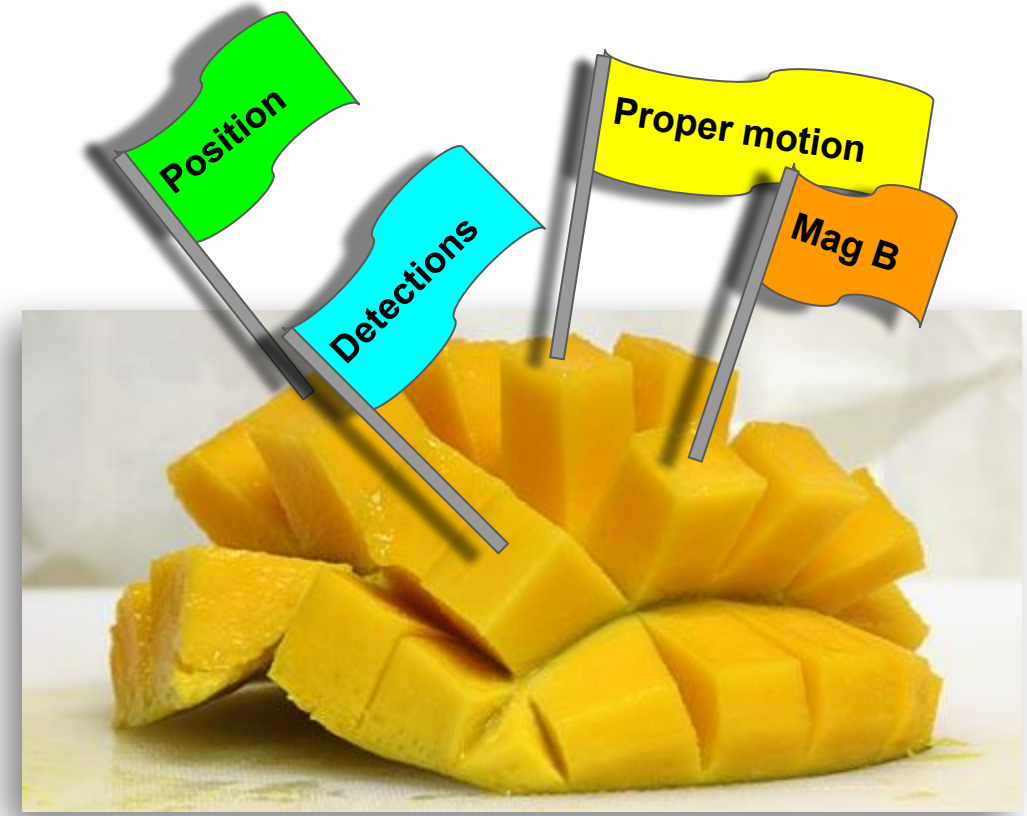
Giving
to archival catalogs
a capability
of interoperability

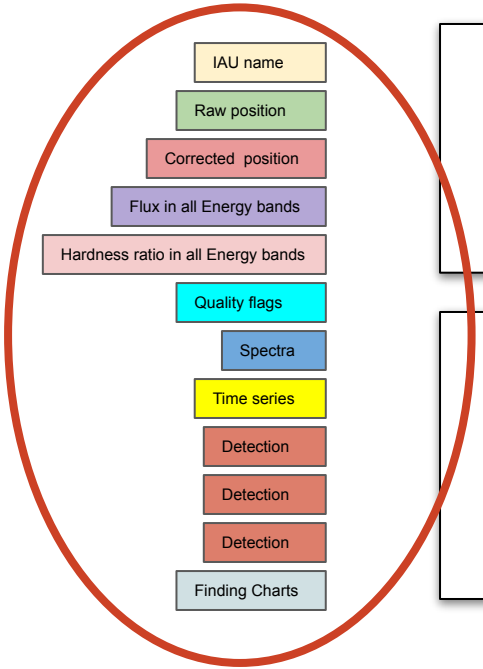


MANGO

A Model for Source Data Model & VOTable Mapping

L. Michel F. Bonnarel G. Landais M. Louys M. Molinaro J. Salgado

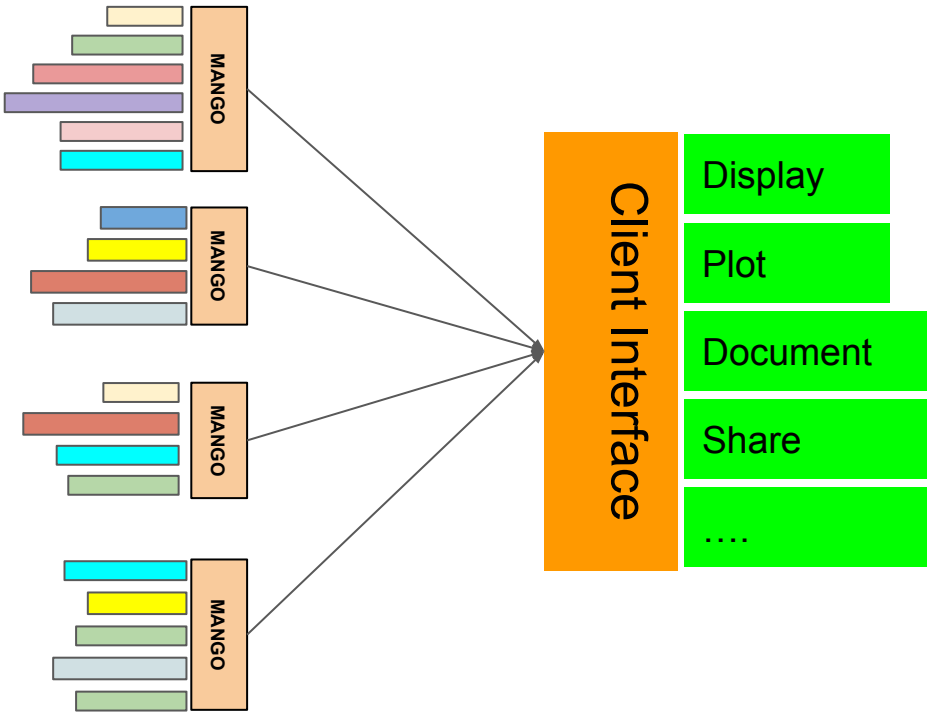
Let’s have a look at some mission data e.g.XMM



- Many different data attached to that source
  - Some are simple quantities
  - Some are complex datasets
  - Some are usual
  - Some are mission-specific

- We would like to make these data understandable for a client that has not been developed for that specific mission.
  - Understanding what does the 0 value means for the detection flag
  - Getting the energy range of the band #1
  - Getting what systematic error applied on corrected position
  - ...

Model for **AN**notating **G**eneric **O**bjects



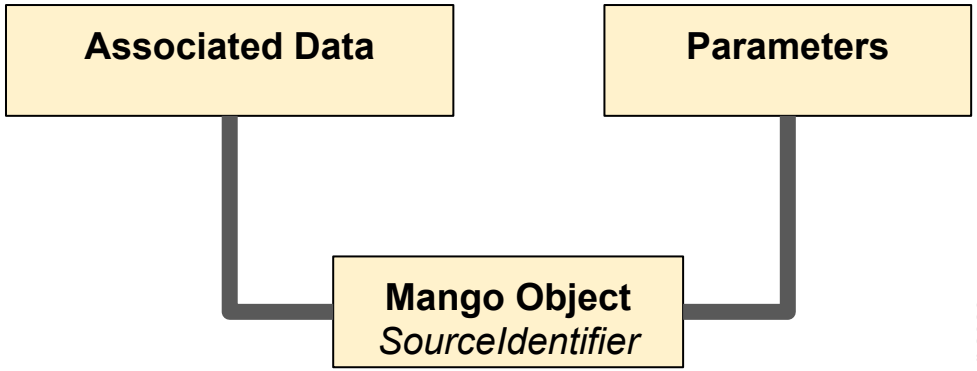
- MANGO is **not** for describing what **sources are**
- MANGO is for **describing source data** that cannot be changed by the curator (archive, TAP) and that can be very **heterogeneous** (Vizier)

- MANGO purposes
  - **Template** to interpret data attached to a source
  - **Guideline** for building **annotations**
- Expectation
  - Clients can get an **accurate description of parameters** whatever the data origin is.

Data Exposed by MANGO

A MANGO source is made with

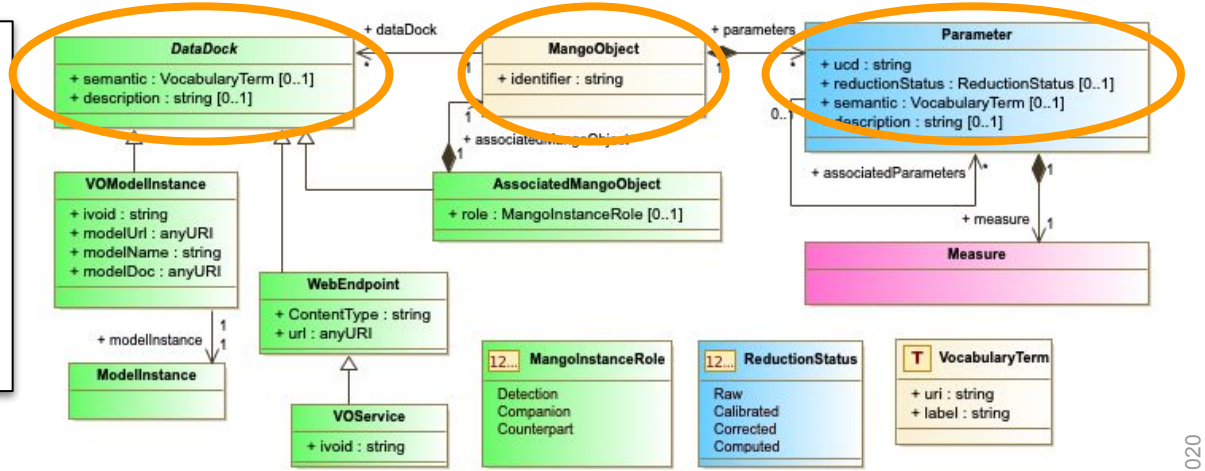
- One source **identifier**
- A set of **Parameters**
  - Set of simple values (string or numerical)
  - Measurements
  - Computed parameters
  - Flags
- A set of **associated data**
  - complex data such as light curves,
  - any sort of counterparts,
  - Pointers VO services



Mango Skeleton

3 **components**

- One source **identifier**
- 2 **Docks**
  - The content of the docks are not defined by the model
  - The model lists possible objects that can be attached to a dock

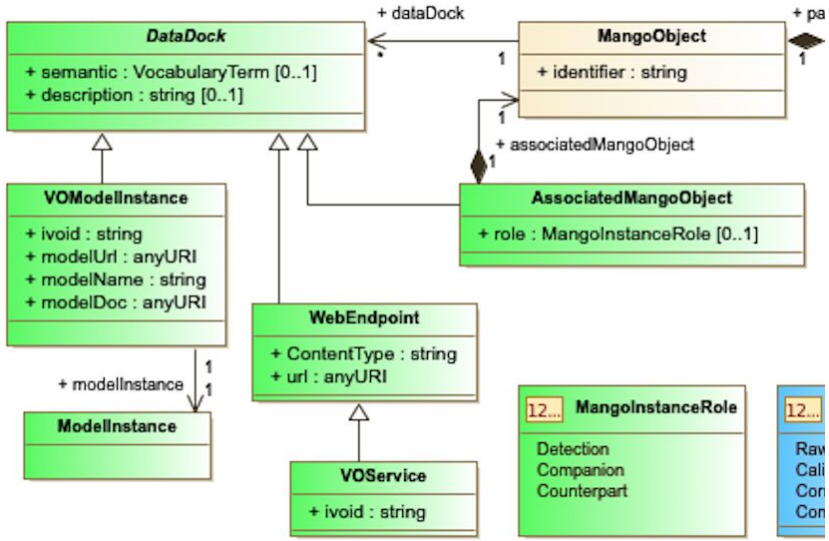


Docks are **open ended** data **containers**

- The model describes which quantities that can be dropped off on a dock
- It does no say which ones have to be there or not
- The content of the docks varies from a dataset to another
- We can have several instances of the same quantity on a dock
  - Multiple positions
  - Multiple counterparts

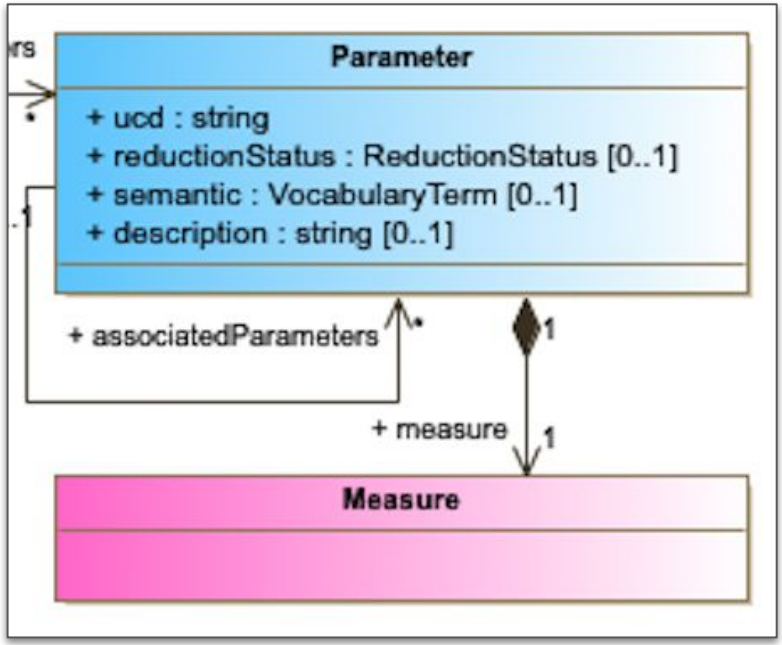
Associated Data

- URL (VO or not)
  - VO endpoints
    - Datalinks, SIA, SSA
    - Serving data related to that source
  - General purpose URLs
    - Documentation
    - Non VO tools
- Other Mango Instances
  - Source detections
  - Counterparts in other datasets
- VODML serialized objects
  - Light curves
  - Provenance
- Purpose
  - Pack data bundles on a VOTable and to retrieve them
  - Can be replaced with Datalinks
- Less tested for now



Mango Parameter

- A MANGO parameter is an object that can be dropped on the parameter dock
- Parameters have 2 components
  - A semantic block giving the role of the measure
    - UCD
    - Reduction status (Model enumeration)
    - Vocabulary entry (label + URI)
    - Text description
  - instance of class deriving from the abstract Measure class
    - See after
- Extended usage of the UCDs
  - UCD scope extended to complex values (values +error)
    - Example: pos;meta.main gives the role of a sky position measure, along with its values, errors and frame.



Measure

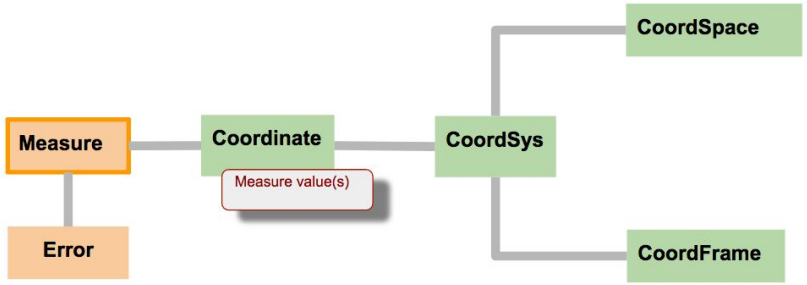
- Broad measure taxonomy
  - Still in expansion
  - Can be measurements or computed values or annotation

- The use of different measurements in archival data is very variable
  - Some are very common (**`position`**)
  - Some are less usual (**`shape, flag`**)
  - Some can be duplicated in the docks (**`multipositions, multiple bands`**…)
  - Some are definitely exotic (“**`decimal log of surface gravity`**”)

- A modeling effort weighted by that usage rate
  - The most popular are modeled by specific classes
  - The less popular can be described by generic objects
  - No trouble with adding new quantities that have not been foreseen by the model

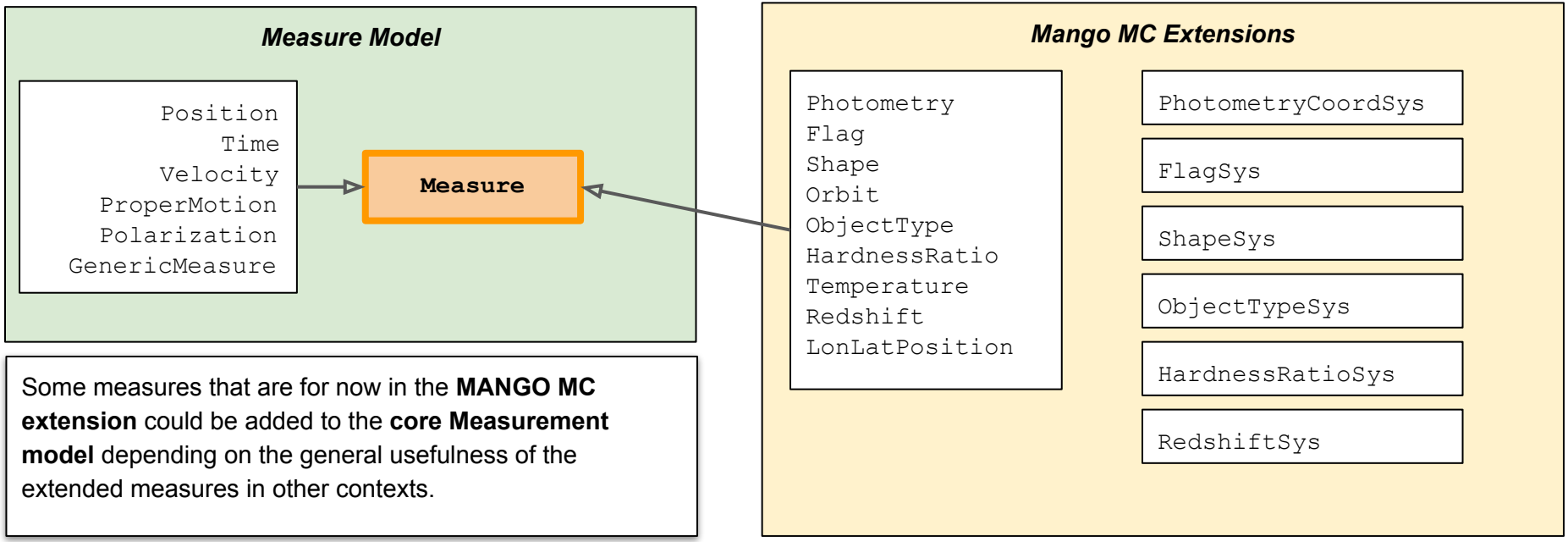
Measure Classes

- All built on the Meas/Coordinates (MC) pattern
  - `Meas` = `coord` + `error`,
  - Using the same pattern for any measure facilitates parsing



- Using MC classes when possible ...
  - `Meas` concrete classes are reused as they are.
  - All measure that have no dedicated classes can be handled by generic `Meas:GenericMeas`
- … or extending MC classes
  - Mango proposes a set of classes that are not in `Meas:Measure`
  - Some are necessary (Luminosity, shape, quality flag)
  - The others still must be discussed
- Using a specific class when available is not mandatory
  - A magnitude can be presented with a generic measure.
  - The semantic block allows us to properly identify the quantity.

Measurement Core Model Extension

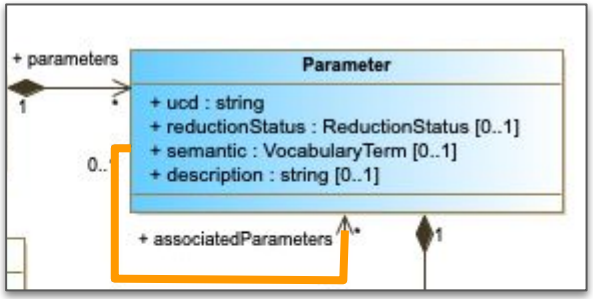


Some measures that are for now in the **MANGO MC extension** could be added to the **core Measurement model** depending on the general usefulness of the extended measures in other contexts.

Measure Sets: A Vizier Request

- Vizier data need a way to **group columns**
  - Grouping columns around a particular quantity
    - `VizieR J/MNRAS/392/19/2slaqqso`
      - `Redshift + quality`
      - `Spectroscopic type + flag`
    - `Gaia`
      - `Position + proper motion` to handle complex errors (Xmatch - work in progress)
  - Such groups have no semantic but **is-related-to**

- Likely one of the most important features to annotate Vizier data
  - Implemented by the ***associatedParameters*** relation



Mapping Guidelines

- We need a convenient way to exercise Mango on Real data

- Client requirements
  - Retrieving data with generic code (no dependency with any particular service)
  - Getting a data presentation that facilitates the comparison with different datasets
  - Being able to restore data hierarchies faithful to the model
  - Being able to gather data spread out within the VOTable

- Provider requirements
  - Facilitate(*) the annotation of heterogeneous and frozen datasets

- In between <GROUP> and a pure ORM (Object Relational Mapping)
  - Compactness
  - Human readability
  - Better than GROUPs to map hierarchical data
  - Do not pretend support a round trip validation (model -> votable -> model)
    - This allows major simplifications

(*) As much as possible

Syntax Fundamentals

- Any complex data hierarchy can be exchanged pair key values, tuples and lists
  - See JSON based Web applications

Key/value pair

```
<ATTRIBUTE
    dmrole="mango:stcextend.PhotFilter.effectiveWavlength"
    dmtype="ivoa:real" value="7740.87" />
```

Tuple

```
<INSTANCE dmrole="mango:Parameter.measure"
    dmtype="mango:Parameter">
    <ATTRIBUTE dmrole="mango:Parameter.semantic"
        dmtype="ivoa:string" value="#position" />
    <ATTRIBUTE dmrole="mango:Parameter.ucd"
        dmtype="ivoa:string" value="pos.eq;meta.main" />
    <ATTRIBUTE dmrole="mango:Parameter.description"
        dmtype="ivoa:string" value="this is the position" />
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="meas:Measure.error"..
</INSTANCE>
```

List

```
<COLLECTION size="-1"
    dmrole="mango:MangoObject.parameters">
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="mango:Parameter.measure"..
    <INSTANCE dmrole="mango:Parameter.measure"..
</COLLECTION>
```

Mapping Block Structure

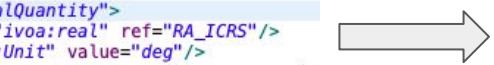
- One block located in the top of the VOTable
- One block maps data for one model

```
<MODEL_INSTANCE>
    <MODEL>
     URI + name of the instanciated model
    </MODEL>
    <GLOBALS>
     Model instances with a global scope
        -   Datatypes
        -   Coord systems
    </GLOBALS>
    <TABLE_MAPPING tableref="Table1">
     Mapping of the data contained in the table labeled Table1
    </TABLE_MAPPING>
    <TABLE_MAPPING tableref="Table2">
     Mapping of the data contained in the table labeled Table2
    </TABLE_MAPPING>
</MODEL_INSTANCE>
```

Other Features

Shortcuts: Model components that are parts of a standard can be folded in compact XML elements

```
<INSTANCE dmrole="coords:Point.axis1" dmtype="ivoa:RealQuantity">
  <ATTRIBUTE dmrole="ivoa:RealQuantity.value" dmtype="ivoa:real" ref="RA_ICRS"/>
  <ATTRIBUTE dmrole="ivoa:Quantity.unit" dmtype="ivoa:Unit" value="deg"/>
</INSTANCE>
```



```
<SC_REALQUANTITY dmrole="coords:Point.axis1"
    ref="RA_ICRS" unit="deg" />
```

Row filtering: Only processing data with a certain field value

```
<COLLECTION dmrole="dmrole">
        <TABLE_ROW_TEMPLATE>
                <FILTER ref="ref" value="value"/>
                <INSTANCE dmref="dmref" dmrole="dmrole" />
        </TABLE_ROW_TEMPLATE>
</COLLECTION>
```

Foreign keys: Joining data from different tables

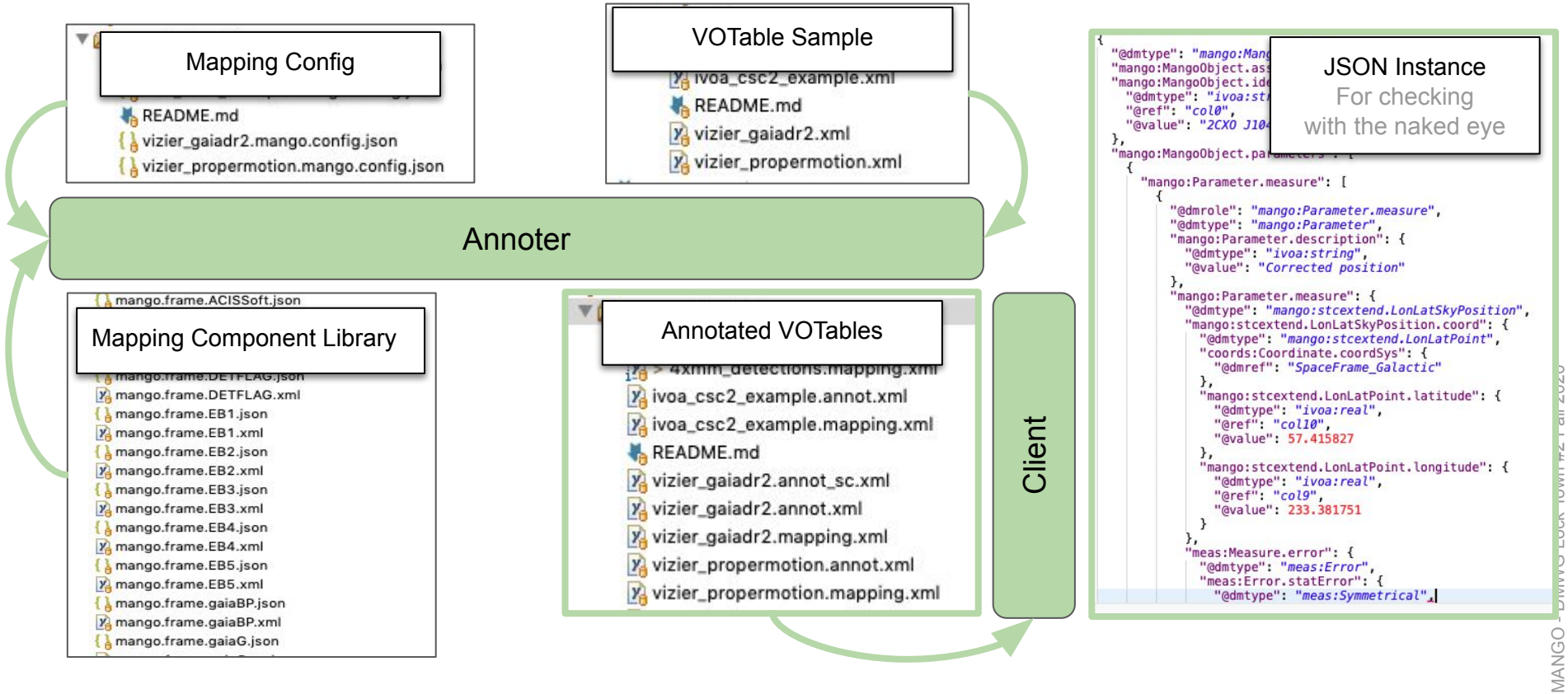
```
<TABLE_MAPPING tableref="fgdgfddf">
        <COLLECTION dmrole="dmrole">
                <JOIN tableref="tableref" primary="primary" foreign="foreign">
                        <INSTANCE dmref="dmref" dmrole="dmrole" />
                </JOIN>
        </COLLECTION>
</TABLE_MAPPING>
```

Row grouping: Grouping data of the same source spread over multiple rows

```
<GROUPBY ref="ref">
        <INSTANCE dmref="dmref" dmrole="dmrole">
                <ATTRIBUTE ref="ref" dmrole="dmrole" dmtype="dmtype"
                        value="dmvalue" />
        </INSTANCE>
</GROUPBY>
```

MANGO - DMWG Lock

Test Bench


Mapping Config
README.md
vizier_gaiadr2.mango.config.json
vizier_propermotion.mango.config.json
VOTable Sample
README.md
vizier_gaiadr2.xml
vizier_propermotion.xml
JSON Instance
For checking
with the naked eye
Annoter
Mapping Component Library
mango.frame.ACISSoft.json
mango.frame.DETFLAG.xml
mango.frame.EB1.json
mango.frame.EB1.xml
mango.frame.EB2.json
mango.frame.EB2.xml
mango.frame.EB3.json
mango.frame.EB3.xml
mango.frame.EB4.json
mango.frame.EB4.xml
mango.frame.EB5.json
mango.frame.EB5.xml
mango.frame.gaiaBP.json
mango.frame.gaiaBP.xml
mango.frame.gaiaG.json
Annotated VOTables
ivoa_csc2_example.annot.xml
ivoa_csc2_example.mapping.xml
README.md
vizier_gaiadr2.annot_sc.xml
vizier_gaiadr2.annot.xml
vizier_gaiadr2.mapping.xml
vizier_propermotion.annot.xml
vizier_propermotion.mapping.xml
Client

AstroPy Wrapper

The MANGO validation requires to show up a good level of **compliance with AstroPy**.

The **ModelInstanceInVot** code includes an Astropy wrapper

- Produces Astropy objects from MANGO annotations
- Very few features for now

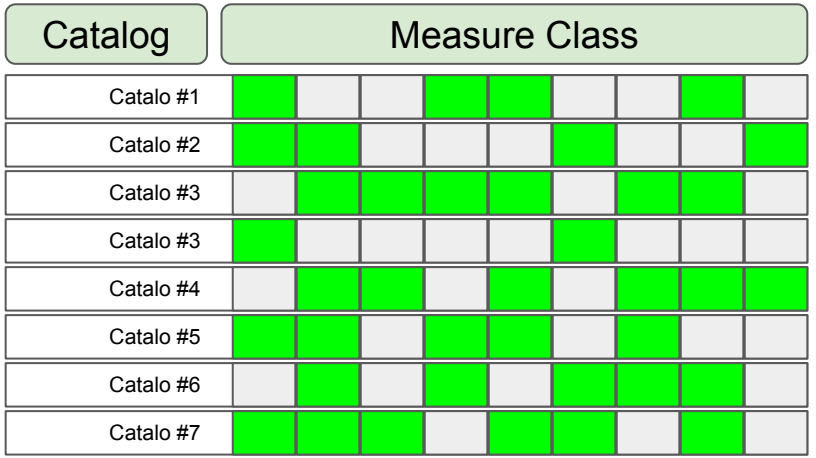
```
wrapper = AstropyWrapper(vodml_instance, mapper_name)

print(f"Astropy space frame: {wrapper.get_space_frame(inst)}")
print(f"Astropy time frame: {wrapper.get_time_frame(inst)}")

'''
# output

Astropy space frame: <ICRS Frame>
Astropy time frame: ('tcb', <EarthLocation (0., 0., 0.) m>, 'mid')
'''
```

Discovering Catalog data in TAP services with MANGO

Issue raise by Christophe Arviset (ESA) at last interop

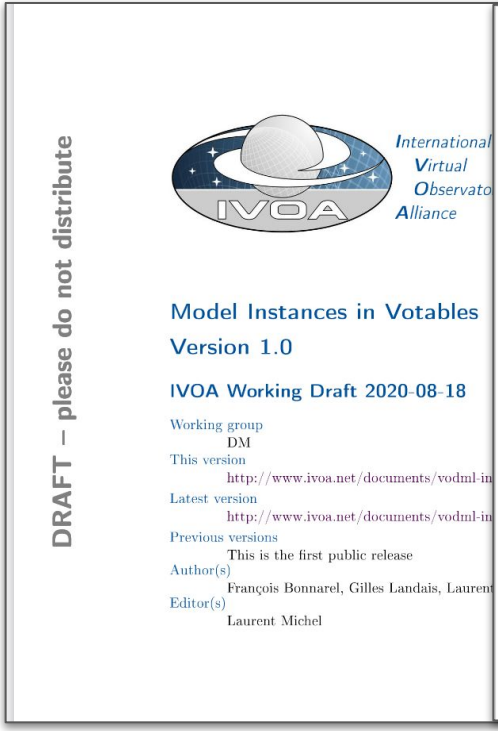
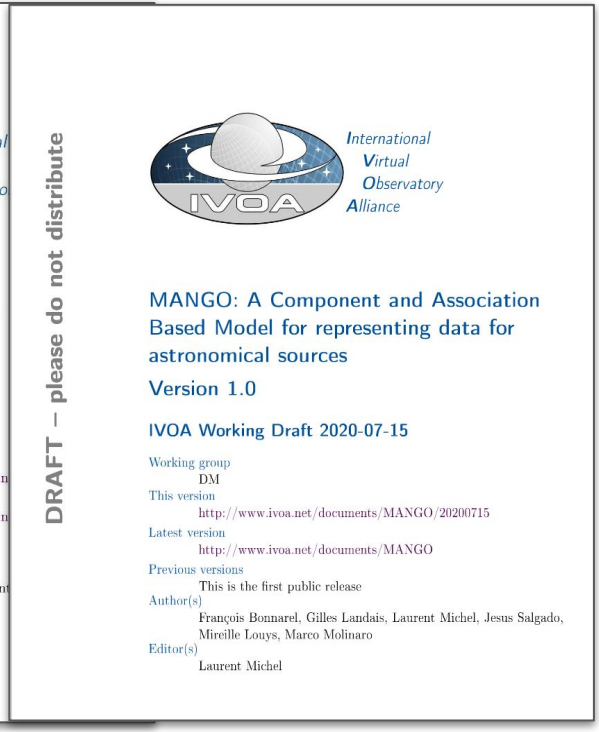
- Could be similar to Obscore (`MangoCore`)
  - Rows: catalog identifiers
  - Columns: MANGO parameters
    - Ranges of simple booleans
- Not easy to to do because parameter sets are open ended
- Must see how to refer to associated data.

WE should have a look at whether there is a way to tag Mango parameters within the `TAP_SCHEMA`.



Possible `ivoa.MangoCore` table

Status and Prospects

https://github.com/ivoa-std/MANGO
https://github.com/ivoa-std/ModelInstanceInVot/

- MANGO
  - UML - Modelio + VO-DML
  - Document in progress
  - Available on GITHub
- Mapping
  - Schema
    - XSD 1.1 ready
    - Lots of unit tests
  - Document in progress
  - Test bench in permanent progress
- The standard is developed along with concrete implementations
  - Slow down or speed up the process
    - Question of point of view