

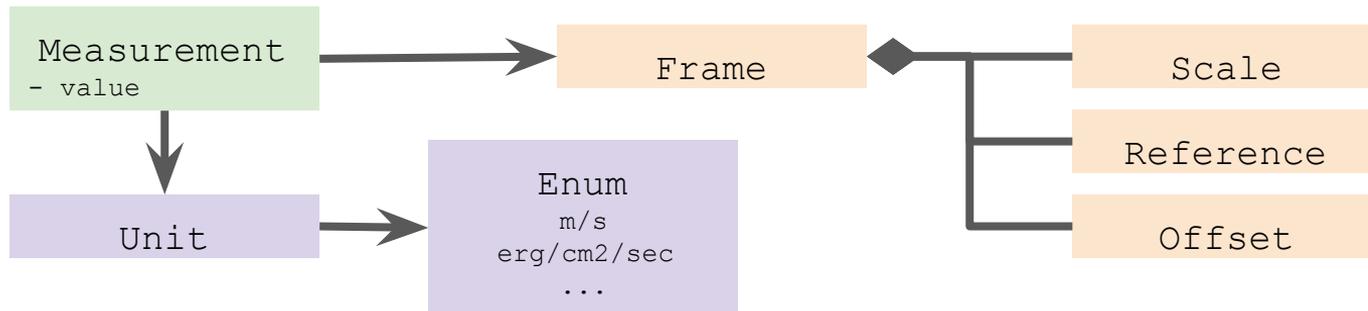
Modèles et VOTables

Une histoire qui n'en finit pas

L. Michel

What Are Data Models For

- **Formal** description of the quantities used by the experts in a domain
- **What does the human knowledge (common sense) say:**
 - A `measurement` is a value with a given `unit` that is valid in a given `frame`
- **The Model gives a formal representation of that knowledge**
 - The model defines the quantity classes, the names, the vocabulary and
.... The relationships between those elements



What Are VO Data Models Used For: Just ask ChatGPT

Why do we need to map our data on standard data models in the virtual observatory 

Tl;dr:

Mapping data to standard data models in the virtual observatory helps ensure that data from different sources can be accessed, shared, and compared easily and efficiently. It also allows users to search for, manipulate, and analyze data more quickly and accurately.

What Are VO Data Models For

What could I do with the models that I cannot do now?

- Get a complete representation of data context
 - Provenance
 - Coordinate systems
 - Observation location
- Get an accurate representation of the errors
- Support cross-columns parameters
 - Columns grouping
 - Complex errors (pos + pm + parallax)
- Gather data from multi-table VOTable (sources + detection)
- Exchange model instance with SAMP

The DM and the VOTables

- **Data annotation with UTypes**

- Group items refer to model leaves
- Data response kinked to models by a key mechanism



works well

- **Mapping data on models in query responses**

- Data response comes with a whole description of the model they refer to.
- The client has enough material to build model instance from the data.

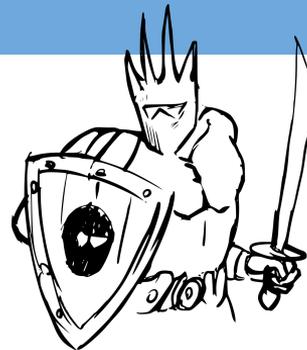


Still stuck for now :-)

La Vie du DM-WG

- **Un champ de bataille permanent**

- Préserver les acquis
- Discussion sur les gains potentiels
- Confrontations de différentes approches incompatibles



- **Un consensus sur les composants à fournir**

- Modèles
- Méthode de sérialisation des modèles
- Moyen de connecter modèles et données
 - Produits intégrés (TS, spectres)
 - Données d'archives
- Code capable de gérer tout cela

Stratégie de Recherche d'un Consensus

- **En 2016, la communauté accepte le principe de VO-DML**
 - Une méthode de sérialisation unique pour tous les modèles
 - VO-DML est un schéma XSD/Schematron
 - Les modèle deviennent utilisables par des outils communs
 - XSLT
 - Générateurs de code
 - Générateurs d'annotations
 - Bibliothèques
 - Permet un réutilisation formelle des modèles

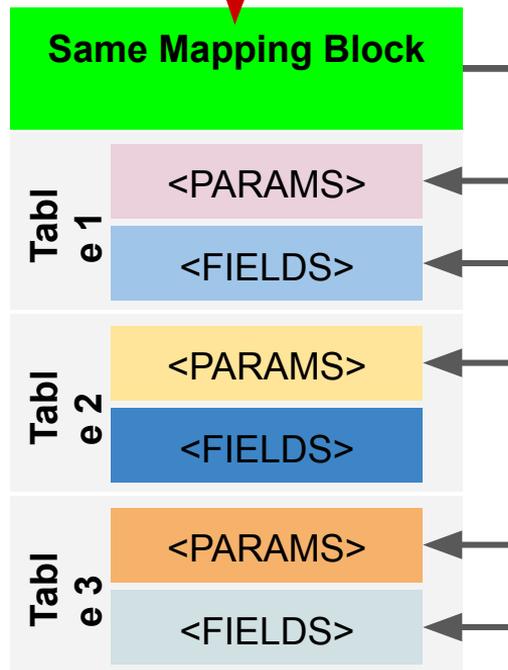
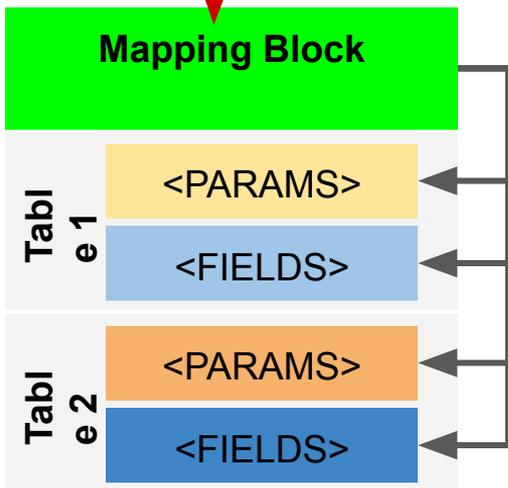
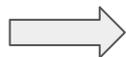
Stratégie de Recherche d'un Consensus

- **En 2017, la communauté accepte les principes de bases pour des annotations**
 - Définir une syntaxe pour connecter données et modèles dans des VOTables
 - Cette syntaxe devait être disposée de manière à ne pas altérer le schéma des VOTables
 - Doit être en XML
 - Doit obéir à un schéma indépendant de tout modèle particulier

Model Aware Client

- Only sees the mapping blocks
- Must be able to retrieve data just by querying the mapping block

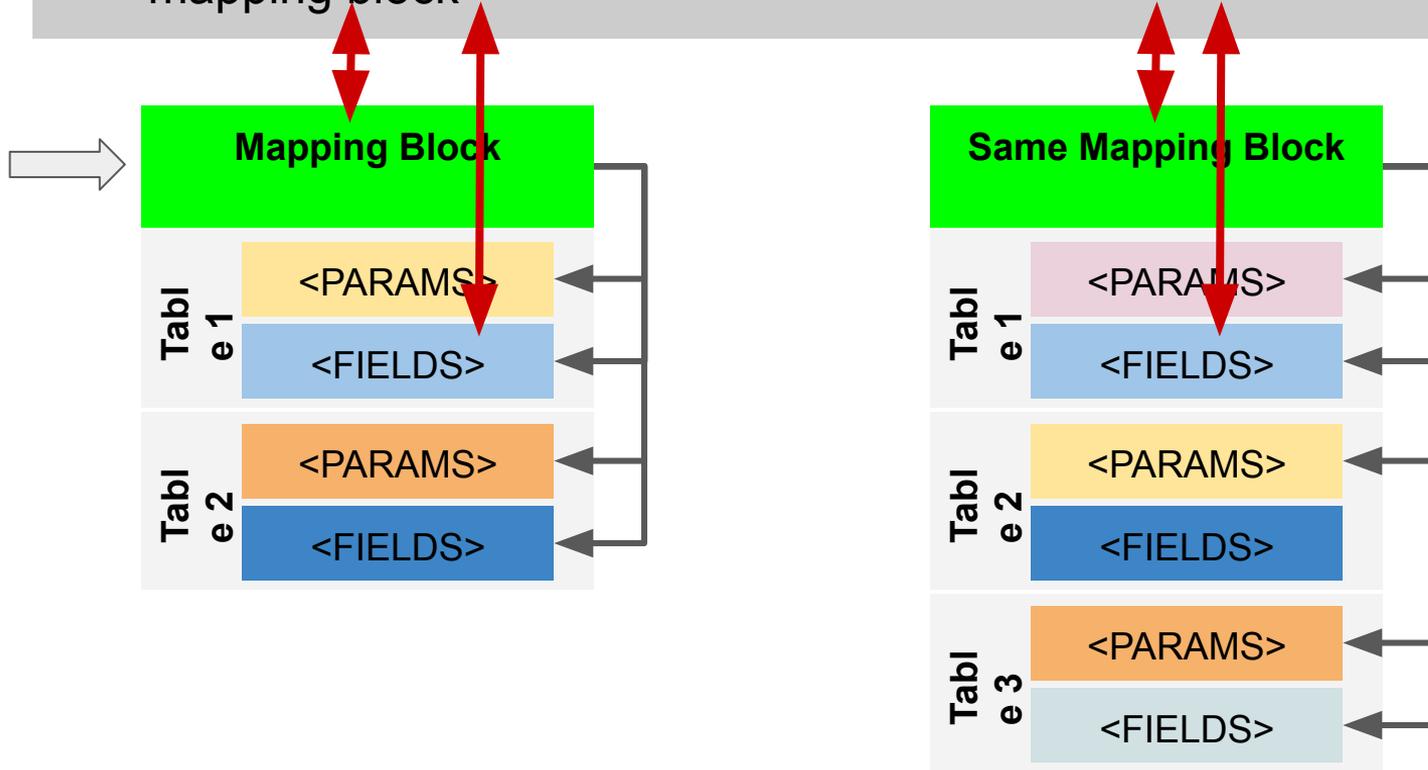
VO-DML
representation
of the model(s)



Model *less* Aware Client

- Sees both data and mapping block
- Refine the data description with information extracted from the mapping block

VO-DML
representation
of the model(s)



Stratégie de Recherche d'un Consensus

- **Tout le travail du WG allait se faire sur cette base**
 - Mise en musique des principes établis par VODML
- **Une feuille de route claire**
 - Définir les modèles qui vont bien
 - Réutiliser l'existant
 - Définir de nouveaux modèles au besoin
 - Définir la syntaxe de mapping
 - Valider avec du code opérationnel

Définir les Modèles

- **Les modèles pour les produits scientifiques**
 - **Cube** (MCD): modèle flexible pour des données complexes telles que des images ou des listes d'événements.
 - Utilise des composants importés d'autres modèles (Measure/Coords)
- **Modèle pour les données d'archive**
 - Difficile à définir car peu de points communs entre les différents jeux de données
 - Essayer de ne pas faire doublons avec les méta-données très riches déjà dans les VOTables.

Définir un modèle pour les données d'archive

Stratégie pour définir les use-cases

2018: Poll to get requirements for parameters describing sources

- A few answers
- The GAIA case among them

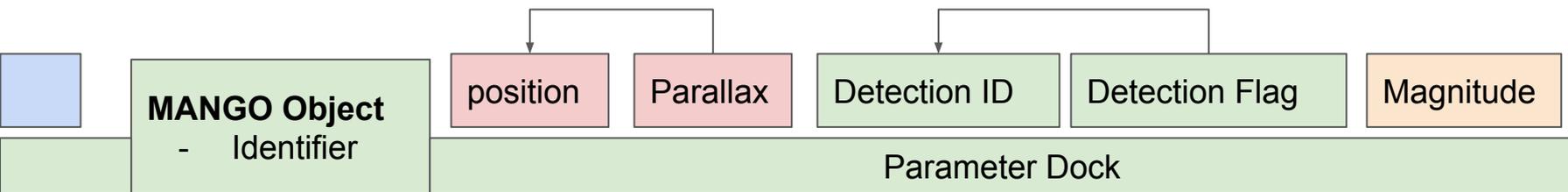
2019: Source DM session in Paris where various people including scientists was invited to present their requirements

- we got a robust set of scientifically relevant use cases
- We got some developers and provider requirements

Un Modèle pour les Données d'Archive

- **Identification des propriétés pas, peu ou mal définies dans les VOTABLEs**
 - Erreurs complexes
 - Systèmes orbitaux
 - Formes complexes
 - Interprétation des statuts et labels
 - Association de mesures entre elles
 - Association de mesure et de drapeaux
 - Définir la sémantique des liens
 - Etat de calibration (photométrie ou autre)
 - Attacher une sémantique aux mesures
- **Limiter au maximum la liste des champs obligatoires**
 - Peu de points communs entre les différents jeux de données
 - Ne mettre que les mesures présentes
 - Ne pas faire doublons avec les méta-données déjà dans les VOTables.

MANGO: Un modèle d'agrégation de



Meas/Coords

Définition fines des coordonnées spatiales et temporelle

PhotDM

- Calibration photométrique
- Filtres
- Système photométriques

Mango Classes

- Labels
- Couleurs
- TBD

Provenance

Réunir tous les paramètres dans un **Dock** pour pouvoir représenter la source comme un objet unique et non comme un ensemble d'entités indépendantes

Définir une Syntaxe pour les Annotations

- **Une proposition originale (2017)**

- Proche de la structure de VO-DML
- Concept de bloc d'annotations isolé et structuré
 - Possibilité de factoriser des composants utilisés plusieurs fois
- Beaucoup d'éléments XML différents mais pas d'attributs
- Draft en 2018 et puis plus d'activité

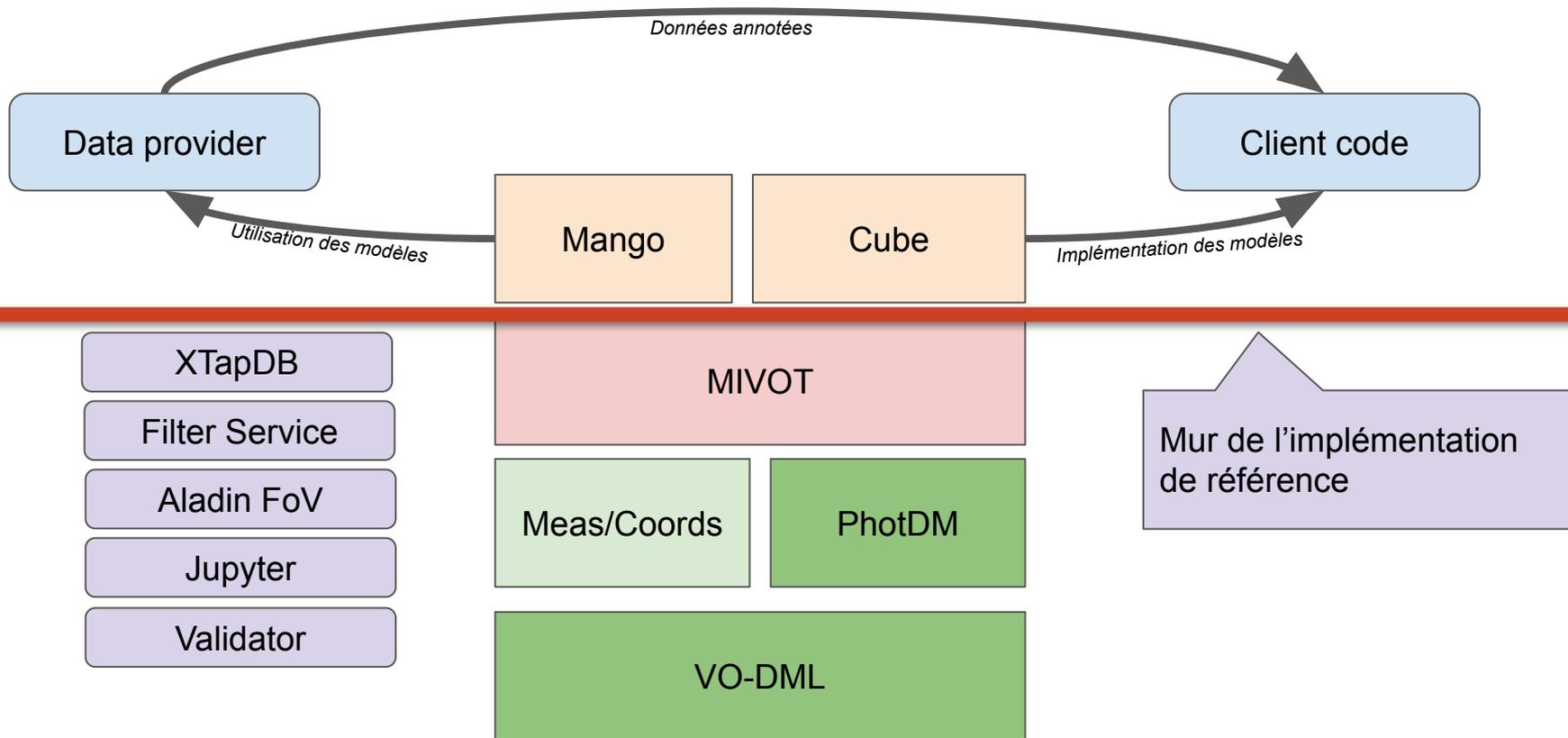
- **Une contre-proposition**

- Même architecture
- Des éléments reprenant la structure du modèle à la JSON (objets, attributs lists), ce dont a besoin le client
- Connection au données via des attributs

- **Une fusion des deux en 2021**

- **Model Instances in VOTables** (en RFC depuis cet automne)

La Situation Aujourd'hui



Ce que nous avons: TAP implémentation

- **Le principe**

- Analyser la requête pour voir quelles sont les colonnes qui peuvent être mappées
- Pour chaque classe mappable: ajouter le bloc XML correspondant dans le BLOC MIVOT de la VOTABLE

- **Réalisations**

- Un BoF ADASS 2021, “TAP and the Data Models”
- Un service XTapDB fonctionnant pour deux tables (extension VOLLT)

<https://xcatdb.unistra.fr/xtapdb>

Ce que nous avons: Autre service

- **SVO fFilter Profile Service (PhotDM)**

- Retourne des calibrations photométriques au format MIVOT
- Directement insérables dans une VOTable

<http://svo2.cab.inta-csic.es/svo/theory/fps/fpsmivot.php?PhotCallID=2MASS/2MASS.H/Vega>

- **Foot print service**

- Draw a Footprint
- Save it as MIVOT instance
- Display it in Aladin Lite

<https://saada.unistra.fr/fov/>

Ce que nous avons: Coté client

Un interpréteur PYTHON

- Capable de traiter tous les motifs permis par MIVOT
 - Mapping simples
 - Jointures...
- Différentes sorties
 - Bloc XML
 - Messages JSON
 - Objets AstroPY
- Utilisé pour nos différentes implémentations
 - Notebooks
 - Fork AstroPY PyVO
 - validateur

https://github.com/ivoa/modelinstanceinvot-code/tree/package/mivot_code

Ce que nous avons: Coté validateur

Un validateur PYTHON

- Valide la syntaxe
 - XSD 1.1
- Valide la structure du modèle instance par instance
 - Les éléments présents dans les annotations existent-ils vraiment dans le modèle?
- Peut être utilisé pour créer des blocs MIVOT valides
 - Donne moi un data type et je te donnerai le bloc MIVOT correspondant et prêt à l'emploi

<https://github.com/ivoa/mivot-validator>

Ce qui manque

- **De l'huile de coude**

- Amener nos démonstrateurs à un niveau de qualité proche de la mise en production
 - Surtout la partie Astropy/PyVO
- Tout contributeur/trice est bienvenu

- **De la doc sur les implémentations réalisées**

- Ou?

- **Des données fournies par des tiers**

- Montrer comment les annoter
- Montrer comment consommer les annotations
- Alimenter des tutoriaux

Le problème de validation

- **Nous sommes bloqués sur un problème de validation**
 - Normalement, il faut deux implémentations de référence pour un standard
- **Le problème est que nous ne voulons pas valider un standard mais une pile de 4 étages.**
 - Pour valider l'ensemble il nous faut CUBE et/ou Mango
 - Pour valider Cube et/ou MANGO il nous faut MIVOT
 - Personne ne va s'investir dans l'implémentation tant qu'il n'y a pas de valeur ajoutée.
 - La valeur ajoutée peut venir un peu de PhotDM mais beaucoup de MANGO/CUBE

