

Model Instances in Votables Version 1.0

IVOA Working Draft 2020-08-18

Working group

DM

This version

http://www.ivoa.net/documents/vodml-instance-vot/20200818

Latest version

http://www.ivoa.net/documents/vodml-instance-vot

Previous versions

This is the first public release

Author(s)

François Bonnarel, Gilles Landais, Laurent Michel, Jesus Salgado Editor(s)

Laurent Michel

Abstract

Vodml-instance-vot proposes a syntax to map VOTable data on any model serialized in VO-DML. Vodml-instance-vot annotations are grouped in a single XML block located in the VOTable head. The annotation block allows to easily reconstruct the model structure. It designed in a way that the block can be reused on different data sets in order to facilitate the annotation process. Vodml-instance-vot is enable to join data from different tables

Status of this document

This is an IVOA Working Draft for review by IVOA members and other interested parties. It is a draft document and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use IVOA Working Drafts as reference materials or to cite them as other than "work in progress".

A list of current IVOA Recommendations and other technical documents can be found at http://www.ivoa.net/documents/.

Contents

1	\mathbf{Intr}	oduction	3
	1.1	Role within the VO Architecture	4
2	Use	Cases and Requirement	4
	2.1	Use Cases	4
		2.1.1 Client Side	4
		2.1.2 Server Side	5
	2.2	Requirements	6
3	Syn	tax	7
	3.1	Mapping Block Structure	7
	3.2	MODELS	8
	3.3	GLOBALS	9
	3.4	TABLE_MAPPING	10
	3.5	INSTANCE	11
	3.6	ATTRIBUTE	13
	3.7	COLLECTION	15
	3.8	TABLE_ROW_TEMPLATE	17
	3.9	FILTER	18
	3.10	JOIN	20
	3.11	GROUPBY	22
	3.12	Shortcuts	24
		3.12.1 SC_REALQUANTITY	24
		3.12.2 SC_INTQUANTITY	24
\mathbf{A}	Cha	nges from Previous Versions	26

Acknowledgments

 $CDS/TDIG/SourceDM\ contributors$

Conformance-related definitions

The words "MUST", "SHALL", "SHOULD", "MAY", "RECOMMENDED", and "OPTIONAL" (in upper or lower case) used in this document are to be interpreted as described in IETF standard RFC2119 (?).

The Virtual Observatory (VO) is a general term for a collection of federated resources that can be used to conduct astronomical research, education, and outreach. The International Virtual Observatory Alliance (IVOA) is a global collaboration of separately funded projects to develop standards and infrastructure that enable VO applications.

1 Introduction

The first purpose of a model is to provide, for a particular domain, a formal description of the relevant quantities and of the way they are connected together. This documentary role facilitates the communication between the stack-holders and thus the design of interoperability protocols.

At data level, interoperability consists in arranging searched data in a way that a client can understand them without taking care of their origin. So that, the same code can process and compare data coming from different sources. That way to arrange data is given by the model.

This is not done by default with VOtables because VOTables are containers. The VOTable schema cannot say how data are mapped on a given model or whether they match any model at all. This is not an issue for simple protocol responses (ref) because the VOTable structure is defined by the protocol itself but this is however a big issue for VOTables containing native data such as Vizier or TAP query responses.

The challenge here is to bind native data with a given model in a way that a model aware software can see them as model instances while maintaining the possibility to access them in their original forms.

This is partially done with UTypes which may connect FIELDs or PARAMs with model leaves in the case of simple tree-views of the model. Unfortunately, there is no more unique way to build and parse UTypes in the context of more complex the models. This occurs when e.g the same class is used in different location of the model or when the model contains loops. It is also not possible to refer data from different tables with UTypes.

The landscape has dramatically changed in 2016 when VODML (ref) became a recommendation. VODML is a meta-model that gives a standard way to design VO models and to make them machine-readable. In VODML, model leaves are no longer identified by a simple string like UTypes do but by a certain role played in a given location in the model hierarchy. The consequence is that any annotation mechanism based on VODML must preserve

the model hierarchy to save the role played by any components. In this context, it might be easy to re-construct model instances from the annotations.

The main concept of VODML mapping is to insert on top of the VOTable an XML block following the model structure and containing references to the actual data. In such a way that in order to build a model instance, a model-aware client only has to make a copy of that structure and to resolve the references. More generic model-unaware clients can just ignore the mapping block. This approach, has been proposed by (GL and OL).

We have tested the syntax originally proposed and it turned out that it was not well suited for archival data or for TAP responses where the annotation process must be automated as much as possible (entirely for TAP).

Vodml-instance-vot is based on the same principles as the original proposal but with a particular attention given to the annotation of archival data by keeping focused on both client needs and easiness of the annotation process. This requires the syntax to be as simple as possible and as flexible as possible to be usable with a wide range of data sets. With Vodml-instance-vot, the model hierarchy is built upon attributes, tuples and arrays (ATTRIBUTE, INSTANCE and COLLECTIONS). Some other elements have been added to guide the parser (FILTER, TABLE ROW TEMPLATE ...)

The connections with the data are setup by XML element attributes, so that the mapping structure just depends on the model but not on the mapped data.

These ideas were first tested in the framework of the TDIG on VOTA-BLEs containing time series provided by different missions such as Gaia or ZWICKI. Then, the syntax has been refined to be used to validate the Mango model on real data.

1.1 Role within the VO Architecture

Fig. 1 shows the role this document plays within the IVOA architecture (?). ???? and so on, LaTeX as you know and love it. ????

2 Use Cases and Requirement

2.1 Use Cases

2.1.1 Client Side

The mapping is self consistent. The role of the mapping is to give the client all information it needs to reconstruct a datastructure similar to a set of instances of the the model. A model-aware client must be able to do this without implementing any code specific to any particular model. The mapping syntax is independent of the model. The structure of the mapped

PDF fallback.

Sorry - your ImageMagick (convert) does not support SVG import. If on Linux, installing librsvg2-bin should remedy this. Otherwise, please commit your SVG and ask the ivoatex creators to do the the conversion.

Figure 1: Architecture diagram for this document

model is given by the arrangement of the mapping elements, not by the elements themselves

Identifying the nature of the content of a VOTable: A client can get it by a quick look at the annotation block.

Measurement discovery: A client wants to discover whether a VOTable contains some peculiar measurements (position, velocity...). The annotation block allows it to get an answer by a quick parsing of the annotation block.

Data set comparison: A client wants to compare different data set (Xmatch, plot). The annotation provides a homogeneous data representation of these data that allows to put them together in a consistent way.

Data set export: A client wants to export (e.g. with SAMP) a model instance in a convenient format (e.g. json). The JSON model instance can be buit from the annotation block and exported to a another party.

2.1.2 Server Side

The server use cases are to make possible the realization of those of the clients for a reasonable cost. The annotation process can represent a significant extra work for the curator team that must be limited as much as possible. To do so the mapping syntax is designed to facilitate the use of templates and components.

3 server types that could annotate data have been identified:

1. Mission data provider: the data annotation can be set once forever for each data product at the design phase.

- 2. Archival data provider: The data annotation must be done for each archived datat set. The curator has a little control, on the data format and he/she has to do his best to match data with the model(s)
- 3. TAP data provider: In case of TAP services, the annotation process is in charge of the TAP server that must dynamically match the queried data with the model quantities for each specific query.

The goal of this version of the specification is to support requirements 1) and 2) with a special attention to make 2) easier. Support of requirement 3) is still an experimental feature at the time this specification is written.

2.2 Requirements

- Shy Annotations: The data mapping must not affect the operation of existing clients.
- Faithful Annotations: The structure of the annotation must be faithful to any VODML compliant model.
- Different Usage Levels
 - The data mapping must be easily ignored by the client.
 - The data mapping must allow clients to easily detect the model on which data are mapped.
 - The data mapping must allow clients to easily get the general metadata (e.g. coordinate systems).
 - The data mapping must allow clients to get full model instances for each table row.

• Easy to Build

- The mapping structure must be independent of the data structure.
- The mapping syntax should make easy the building of both mapping components and templates.

• Complex Data Mapping

- The mapping syntax must support to retrieve data spread over several tables.
- The mapping syntax must be able to filter data rows which are part of a specific instance.
- The mapping syntax must be able to group data rows in an set of instances.

3 Syntax

The syntax rules specified in this standard allow to build consistent annotations for any model. However, they do not prevent to do foolish things in the same way that a programming language grammar does not protect people against writing irrelevant software. In the following annotation snippets, the values of the XML elements attribute do not refer to any particular model or VOTable. They have been choosen to help readers to figure out their meanings.

3.1 Mapping Block Structure

The mapping block muts be the direct child of VOTABLE. Its scope is the whole VOTable. Its stucture is given below.

Listing 1: INSTANCE bloc example

The mapping construction rules are the same whatever either model or data layout are.

- The mapping is located in a <MODEL_INSTANCE> block, child of <VOTABLE>.
- The mapping elements denote the model structure.
- The <MODEL_INSTANCE> block starts with the list of implemented models.
- The model list is followed the <GLOBALS> block containing data shared by the whole mapping.
- The <GLOBALS> block is followed a sequence of <TABLE_MAPPING>.
- There is one <TABLE_MAPPING> per mapped <TABLE>.

3.2 MODELS

The MODELS blocks contains the list of the models mapped in the block.

- The model list is informative, it is not requested to achieve the parsing.
- The model list can be left empty.
- Models referenced in MODELS are not necessary VO standards, but they
 must be accessible by a VODML URI

```
<MODELS>
   <MODEL>
       <NAME>ivoa</NAME>
       <URL>http://www.ivoa.net/xml/VODML/IVOA-v1.vo-dml.xml
       </URL>
   </MODEL>
   <MODEL>
       <NAME>coords</NAME>
       <URL>http://www.ivoa.net/xml/STC_coords-v1.0.vo-dml.xml
   </URL>
   </MODEL>
   <MODEL>
       <NAME>meas</NAME>
       <URL>http://www.ivoa.net/xml/STC_meas-v1.0.vo-dml.xml
       </UR.L.>
   </MODEL>
</MODELS>
```

Listing 2: GLOBALS block example

MODELS, MODEL, NAME and URL have no attributes.

3.3 GLOBALS

Contains INSTANCEs that can be used everywhere in the MODEL_INSTANCE.

- INSTANCEs children of GLOBALS should have an QID attribute so that they can be referenced from other instances.
- The role of the GLOBALSs children, (INSTANCEs by construction), must be ignored although being mandatory.
- References within GLOBALSs sub-elements to VOTable data (FIELD of PARAM) must be searched in all tables. They must be resolved by the first occurrence matching the reference found.
- GLOBALS has no attributes.

Listing 3: GLOBALS block example

Child	Role
INSTANCE	Model instances with a scope covering the whole VOTable .

Table 1: Allowed GLOBALS children

3.4 TABLE MAPPING

TABLE_MAPPING blocks contain the mapping statements of the data contained in one TABLE.

- There is one TABLE_MAPPING block for each mapped TABLE in the VOTAble.
- A TABLE cannot be referenced by more than one TABLE_MAPPING element.
- The table related to a TABLE_MAPPING is identified by the @tableref attribute. It must be first resolved against the TABLE identifier (@ID) and then against the table name.

</TABLE_MAPPING>

Listing 4: TABLE MAPPING block example

Child	Role
INSTANCE	Mapping of an object or a data type.
TABLE_ROW_TEMPLATE	One instance must be built for each table row. The structure of those instances is given by the TA-BLE_ROW_TEMPLATE children
COLLECTION	Mapping of an object collection

Table 2: Valid TABLE_MAPPING children

Attribute	Role
@tableref	The @ID or the @name of the mapped table

Table 3: TABLE_MAPPING attributes

@tableref	Pattern
MAND	Always mandatory

Table 4: Valid attribute patterns for TABLE_MAPPING

3.5 INSTANCE

Mapping for either object types or a datatype instances.

```
TODO:
put a better rexample
```

Listing 5: INSTANCE block example

Child	Role
INSTANCE	Another embedded instance .
ATTRIBUTE	Primitive attribute .
COLLECTION	Set of items

Table 5: Valid INSTANCE children

Attribute	Role		
@dmrole	VODML role.		
	Can be empty if located in GLOBALS or in TA-		
	BLE_ROW_INSTANCE		
@dmtype	VODML type of the instance.		
	Must never be empty		
@dmref	Reference to another instance in the mapping block.		
	Must never be empty		
@ID	Unique identifier of the instance.		
	Must never be empty		

Table 6: INSTANCE attributes

@dmrole	@dmref	@dmtype	@ID	Pattern
MAND		MAND	OPT	Instance of a certain type playing a certain role.
MAND	MAND			Reference to another instance. No allowed children in this case.

Table 7: Valid attribute patterns for INSTANCE

3.6 ATTRIBUTE

Mapping statement for primitive attributes.

- ATTRIBUTEs are the model leaves.
- ATTRIBUTE values can either be set by reference on tabke data or by literal values.
- ATTRIBUTEs have no children.

Listing 6: ATTRIBUTE examples

Attribute	Role	
@dmrole	VODML role of the attribute.	
	Must never be empty	
@dmtype	VODML type of the instance attribute.	
	Must never be empty	
@value	Literal value of the instance attribute.	
	If ATTRIBUTE has also a @ref, @ref MUST be resolved first.	
	ATTRIBUTE @val must be taken when @ref cannot be resolved	
@ref	Reference of the data element (FIELD or PARAM).	
	Must refer to an element of the TABLE referenced by the current	
	TABLE_MAPPING	
	The client MUST first look for a FIELD matching @ref.	
	If no FIELD is found, it must look for a PARAM	

Table 8: ATTRIBUTE attributes

@dmrole	@dmtype	@ref	@value	Pattern
MAND	MAND	MAND	OPT	The instance attribute must take the value pointed by <code>@ref.</code> If the reference cannot be resolved, the attribute takes the value of <code>@val</code> if present. It is considered as not set otherwise.
MAND	MAND		MAND	The attribute takes the value of @val.

Table 9: Valid attribute patterns for ATTRIBUTE

3.7 COLLECTION

Mapping statement fort sets of either instances or collections.

- A COLLECTION can contain a fixed set of instances or collections. In this case, each element must be mapped individually. Elements can be either INSTANCEs or COLLECTIONS.
- A COLLECTION can contain an unbounded set of instances, one per selected table row. In this case, all items have the same type and thus the same mapping. They can be set with local table data or with data from a joint table.

The example below show up a a fixed length COLLECTION.

Listing γ : COLLECTION example

Child	Role	
INSTANCE	Mapping of one collection item. A collection can embed multiple INSTANCEs	
COLLECTION	Mapping of one collection item. A collection can embed multiple COLLECTIONs	
TABLE_ROW_TEMPLATE. The collection is populated with with one instance per of the current table. When present, this element mutthe only child.		
JOIN The collection is populated with data read from anot table. When present, this element must be the only characteristics.		

Table 10: Valid COLLECTION children

Attribute	Role	
@dmrole	Role played by the collection (VODML relation name usu-	
	ally). Cannot be empty.	
@size	Collection size. This attribute is not necessary to parse	
	the mapping block.	

Table 11: Valid attributes for COLLECTION

3.8 TABLE_ROW_TEMPLATE

This element indicates that one element must be added to the host COLLECTION for each table row.

- The row mapping is given by the INSTANCE child.
- One and only one INSTANCE can be mapped per row. This makes senses since collection elements cannot be made with more than one instance.
- TABLE_ROW_TEMPLATE has no attributes.

@dmrole	@size	Role
MAND	OPT	Role played by the collection (VODML relation name usually). Cannot be empty

Table 12: Valid attribute patterns for COLLECTION

Listing 8: TABLE ROW TEMPLATE examples

Child	Role
INSTANCE	Mapping to be applied to table row

Table 13: Supported TABLE_ROW_TEMPLATE children

3.9 FILTER

This element filters the table rows that are to mapped.

- The filtering condition is based on the equality of a column value with the filter value.
- The mapping specification does not specify the way to deal with data types.

In the example below::

- The light curve will be populated with table rows mapped by the INSTANCE of type test:photometric.point
- Each of these rows must have the value of the column phot_filter_name equals to G.

Listing 9: FILTER examples

Child	Role	
INSTANCE	Mapping to be applied to table rows matching the filter	

Table 14: Valid FILTER children

Attribute	Role	
@ref	Identifier of the column on which the filtering criteria must be applied	
@value	Literal value that is used as filtering criteria	

Table 15: FILTER attribute

3.10 **JOIN**

This element populates the host collection with data taken out from a foreign table and matching the join criteria.

- Each matching row of the foreign table is mapped as one INSTANCE of type test:Detection .
- Self-joins on the local table are allowed.
- The join criteria is based on the equality of the column values. The mapping specification does not specify the way to deal with data types.

```
<TABLE_ROW_TEMPLATE>

<INSTANCE dmrole="primary:point" dmtype="Point">

<ATTRIBUTE dmrole="test:detection.num" dmtype="ivoa:real"

ref="_poserr_148" />

<COLLECTION dmrole="test.detections">

<JOIN tableref="OtherResults" primary="_poserr_148"

foreign="_foreign">

<INSTANCE dmtype="test:Detection">

<ATTRIBUTE dmrole="test:detection.num"</td>
```

@ref	@value	Role
MAND	MAND	All attributes must be set in any case

Table 16: Valid FILTER attribute pattern

Listing 10: JOIN example

Child	Role
INSTANCE	Mapping to be applied to the matching rows.

Table 17: Supported JOIN children

Attribute	Role	
@primary	Column identifier of the primary table used by the join	
@foreign	Column identifier of the foreign table used by the join	
@tableref	ID or name of the foreign table	

Table 18: JOIN attributes

@primary	@foreign	@tableref	Role
MAND	MAND	MAND	All attributes must be set in any case

Table 19: Valid JOIN attribute pattern

3.11 GROUPBY

This element aggregates host table rows in groups which elements have all the same value for a given column.

• Each matching row is mapped as one instance of the INSTANCE child.

In the example below:

- The collection with @dmrole=test.lightcurves will be populated with a set of collections.
- Each of these sub-collections is populated with set of instances mapped by the INSTANCE of test:photometric.point type.
- All INSTANCEs are built wiith rows having all the same values for the column source_name

Listing 11: GROUPBY examples

Child	Role
INSTANCE	Mapping to be applied to the matching rows.

Table 20: Valid GROUPBY children

Attribute	Role	
@ref	Identifier of the column used for the grouping	
@dmrole	Role of the grouped sub-collections	

Table 21: GROUPBY attributes

@ref	@dmrole	Role
MAND	MAND	Must be set with non empty values

Table 22: Valid GROUPBY attribute pattern

3.12 Shortcuts

VODML encourgages people to use the ivoa model for the primitive types. Some of these types have a complex structures that associate units with values. This is the case for the types derived from ivoa:Quantity (ivoa:RealQuantity and ivoa:IntegerQuantity). The XML snippet below shows the regular mapping for a real quantity..

Listing 12: ivoa:RealQuantity example

This block maps a structure that is part of the VODML standards, therefore we can alias it with a compact element.

3.12.1 SC_REALQUANTITY

Shortcut for ivoa:RealQuantity class.

- Can only be used within an INSTANCE
- Using shorcuts requires units to be literals
- Both Oref and Ovalue attribute work the same way as with ATTRIBUTE
- No Odmtype, it is set as ivoa:RealInteger by construction

```
<SC_REALQUANTITY dmrole="coords:PhysicalCoordinate.cval"
    ref="col_id" value="0.0" unit="m/sec" />
        Listing 13: ivoa:RealQuantity example
```

3.12.2 SC INTQUANTITY

Shortcut for ivoa: IntegerQuantity class.

- Can only be used within an INSTANCE
- Using shorcuts requires units to be literals
- Both @ref and @value attribute work the same way as with ATTRIBUTE.
- $\bullet\,$ No ${\tt Qdmtype},$ it is set as ivoa:RealInteger by construction

```
<SC_INTQUANTITY dmrole="coords:PhysicalCoordinate.cval"
    ref="col_id" value="0" unit="m/sec" />
        Listing 14: ivoa:IntegerQuantity example
```

A Changes from Previous Versions

No previous versions yet.