

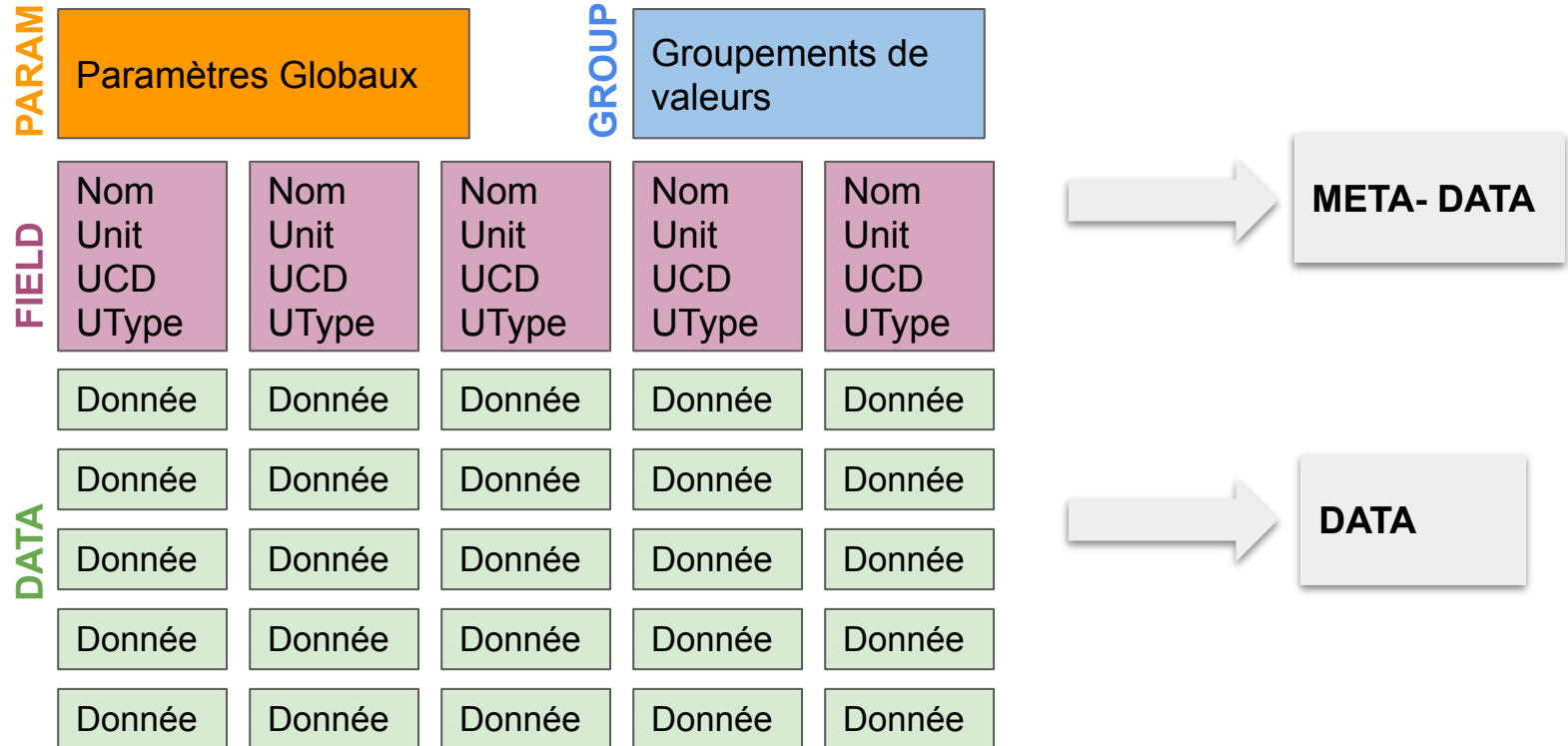
Mapping Syntax

Laurent Michel, Mark Cresitello-Dittmar et al.

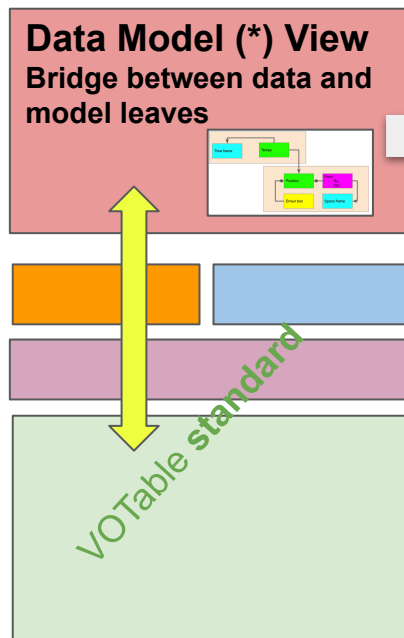
<https://wiki.ivoa.net/twiki/bin/view/IVOA/DataAnnotation>

<https://github.com/ivoa-std/ModelInstanceInVot>

Simplified View on VOTable Content



Providing Tabular Data with a Model View



- **Describe how data are connected to each other**
 - Association flag/error/measure
- **Improve semantic description**
 - Add semantic to data associations
- **Add possibly missing meta-data**
 - Filter descriptions
 - Flag values
- **View not depending on the data provider**

(*) Let's assume we have data models serialized in VODML

Model Mapping Location and Scope

- The model mapping is embedded in one XML Block
 - Located in a `resource[@type=meta]`
 - Scope limited to the parent `resource[@type=results]`
 - Mapped data can be distributed over multiple tables within that `resource`

```
<?xml version="1.0" encoding="UTF-8"?>
<VOTABLE xmlns="http://www.ivoa.net/xml/VOTable/v1.3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  version="1.3">
  <RESOURCE type="results">
    <RESOURCE type="meta">
      <dm-mapping:VODML xmlns:dm-mapping="http://www.ivoa.net/xml/merged-syntax">[]
    </RESOURCE>
    <TABLE name="Results">
      <PARAM ID="_title" name="title" value="TilteReadInParam" datatype="char" arraysize="*" />
      <FIELD ID="_poserr_148" name="oidsaada" datatype="long" ucd="meta.id;meta.main" />
      <DATA>[]
    </TABLE>
    <TABLE name="OtherResults">[]
    <TABLE name="Spectra">[]
  </RESOURCE>
</VOTABLE>
```

Syntax Overview

- **Mapping block structure**

- One **<VODML>** block, the container
- Followed by one **<REPORT>** optional block giving the status of the mapping process
- Followed but one **<MODEL>** element per used model
- Followed by one **<GLOBALS>** element
 - Mapping of quantities that do not relate to table rows
- Followed by one **<TEMPLATES>** element per mapped table
 - Container for the table row mapping

```
<VODML xmlns="http://www.ivoa.net/xml/merged-syntax" >
  <REPORT status="OK">Hand made mapping</REPORT>
  <MODEL name="mango" url="file:/Users/sao/Documents/IVOA/GitHub/ivoa-dm-examples/tmp/Mango-v1.0.vo-dm
  <MODEL name="cube" url="https://volute.g-vo.org/svn/trunk/projects/dm/Cube/vo-dml/Cube-1.0.vo-dml.x
  <MODEL name="ds" url="https://volute.g-vo.org/svn/trunk/projects/dm/DatasetMetadata/vo-dml/Datase
  <MODEL name="meas" url="https://www.ivoa.net/xml/Meas/20200908/Meas-v1.0.vo-dml.xml" />
  <MODEL name="coords" url="https://www.ivoa.net/xml/STC/20200908/Coords-v1.0.vo-dml.xml" />
  <MODEL name="ivoa" url="https://www.ivoa.net/xml/VODML/IVOA-v1.vo-dml.xml" />
  <GLOBALS>
  <TEMPLATES tableref="_PKTable">
  <TEMPLATES tableref="Results">
</VODML>
```

Mapping table rows with TEMPLATES

```
<TEMPLATES tableref="Spectra">
  <INSTANCE dmttype="test:Spectrum">
    <ATTRIBUTE dmrole="test:detection.num" dmttype="ivoa:string1" ref=" _spc_148" />
    <ATTRIBUTE dmrole="test:detection.id" dmttype="ivoa:real" ref=" _foreign_spectra" />
  </INSTANCE>
</TEMPLATES>
</VODML>

RESOURCE>

TABLE name="Spectra">
  <FIELD ID=" _spc_148" name="oidsaada" datatype="char" arraysize='*' ucd="meta.id;meta.main" />
  <FIELD ID=" _foreign_spectra" name="ele" datatype="long" ucd="meta.id;meta.main" />
  <DATA>
    <TABLEDATA>
      <TR>
        <TD>Spectrum 11</TD>
        <TD>1</TD>
      </TR>
      <TR>
        <TD>Spectrum 12</TD>
        <TD>1</TD>
      </TR>
    </TABLEDATA>
  </TABLE>
</TABLE>
```

The diagram illustrates the mapping between a VODML template and a table. The template instance (highlighted in a red box) defines two attributes: `test:detection.num` with `ref=" _spc_148"` and `test:detection.id` with `ref=" _foreign_spectra"`. The table (highlighted in a red box) has two fields: `oidsaada` with `ID=" _spc_148"` and `ele` with `ID=" _foreign_spectra"`. Red arrows indicate the mapping from the `ref` attributes in the template to the `ID` attributes in the table. An orange arrow points from the template box to the explanatory text box on the right.

TEMPLATE[@tableref=Spectra] maps the data of **TABLE**[@name=Spectra]

Each row of that table can be interpreted an instance of type `test:Spectrum`

The connection between model leaves and table columns is achieved by the **ATTRIBUTE**[@ref]=**FIELD**[@ID] condition

Syntax Overview

- A basic object instance

```
<INSTANCE dmrole="coords:PhysicalCoordSys.frame" dmtype="coords:SpaceFrame">  
  <ATTRIBUTE dmrole="coords:SpaceFrame.spaceRefFrame" dmtype="ivoa:string" value="ICRS"/>  
  <ATTRIBUTE dmrole="coords:SpaceFrame.equinox" dmtype="coords:Epoch" value="J2015.5"/>  
</INSTANCE>
```

Role played by this instance in the context of the parent object
not displayed here

- A more complex object instance

```
<INSTANCE dmid="_ts_data" dmrole="" dmtype="cube:NDPoint">  
  <COLLECTION dmrole="cube:NDPoint.observable">  
    <INSTANCE dmtype="cube:Observable">  
      <ATTRIBUTE dmrole="cube:DataAxis.dependent" dmtype="ivoa:boolean" value="true"/>  
      <INSTANCE dmrole="cube:MeasurementAxis.measure" dmtype="meas:Time">  
        <INSTANCE dmrole="meas:Time.coord" dmtype="coords:MJD">  
          <ATTRIBUTE dmrole="coords:MJD.date" dmtype="ivoa:real" ref="_obstime"/>  
          <REFERENCE dmrole="coords:Coordinate.coordSys" dmref="timesys"/>  
        </INSTANCE>  
      </INSTANCE>  
    </INSTANCE>  
    <INSTANCE dmtype="cube:Observable">.  
    <INSTANCE dmtype="cube:Observable">.  
  </COLLECTION>  
</INSTANCE>
```

Attributes set with literal values

Attribute set with the values of the **FIELD**[@ID=_obstime]

The coordinate system is mapped by the element **INSTANCE**[@dmid=_timesys] and located somewhere else in the mapping block, likely in **GLOBALS**

Using References

```
<INSTANCE dmtpe="sed:Sed">
  <INSTANCE dmtpe="sed:Source" dmrole="sed:Sed.source">
    <ATTRIBUTE dmtpe="ivoa:string" dmrole="sed:Source.name">
      ...
    <INSTANCE dmtpe="sed:Position" dmrole="sed:Source.position">
      <ATTRIBUTE dmtpe="ivoa:real" dmrole="sed:Position.longitude" ref="ra" />
      <ATTRIBUTE dmtpe="ivoa:real" dmrole="sed:Position.latitude" ref="dec" />
      <ATTRIBUTE dmtpe="ivoa:vounit" dmrole="sed:Position.unit" value="deg" />
    </INSTANCE>
  </INSTANCE>

  <COLLECTION dmrole="sed:Sed.points">
    <INSTANCE dmtpe="sed:PhotPoint">
      <ATTRIBUTE dmtpe="ivoa:real" dmrole="sed:PhotPoint.flux" ref="flux"
        unit="erg/s/cm^2" />
      <REFERENCE dmrole="sed:PhotPoint.photcal" dmref="_XMMSL2_EB6" />
    </INSTANCE>

    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    <INSTANCE dmtpe="sed:PhotPoint">
    </COLLECTION>
  </INSTANCE>
```

Each table row contains one SED
Toy model

The connection between photometric points
and calibration is achieved by the
REFERENCE [@dmref] = **INSTANCE** [@dmid]
condition

```
<GLOBALS>
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB6">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB7">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB6">

  <INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_H">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_J">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_Ks">

  <INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W1">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W2">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W3">
  <INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W4">

  <INSTANCE dmtpe="photdm:PhotCal" dmid="_GAIA2_G">
</GLOBALS>
```


A Very Complex Object

<GLOBALS>

```
<INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB8">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB7">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_XMMSL2_EB6">
```

```
<INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_H">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_J">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_Ks">
```

```
<INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W1">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W2">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W3">
<INSTANCE dmtpe="photdm:PhotCal" dmid="_WISE_W4">
```

```
<INSTANCE dmtpe="photdm:PhotCal" dmid="_GAIA2_G">
```

</GLOBALS>

```
<INSTANCE dmtpe="photdm:PhotCal" dmid="_2MASS_H">
<ATTRIBUTE dmrole="photdm:PhotCal.identifier" dmtpe="ivoa:string"
value="2MASS/2MASS.H/Vega" />
<!-- Magnitude System -->
<INSTANCE dmtpe="photdm:MagnitudeSystem" dmrole="photdm:PhotCal.magnitudeSystem">
<ATTRIBUTE dmrole="photdm:MagnitudeSystem.type" dmtpe="photdm:TypeOfMagSystem"
value="Vega" />
<ATTRIBUTE dmrole="photdm:MagnitudeSystem.referenceSpectrum" dmtpe="ivoa:anyURI"
value="http://svo2.cab.inta-csic.es/theory/fps/morefiles/vega.dat" />
</INSTANCE>

<!-- Zero Point -->
<INSTANCE dmtpe="photdm:PogsonZeroPoint" dmrole="photdm:PhotCal.zeroPoint">
<ATTRIBUTE dmrole="photdm:ZeroPoint.type" dmtpe="ivoa:string" value="Pogson" />
<ATTRIBUTE dmrole="photdm:ZeroPoint.referenceMagnitudeValue" dmtpe="ivoa:real" value="0" />

<INSTANCE dmtpe="photdm:Flux" dmrole="photdm:ZeroPoint.flux">
<ATTRIBUTE dmrole="photdm:Flux.ucd" dmtpe="ivoa:UCD" value="phot.flux;meta.modelled" />
<ATTRIBUTE dmrole="photdm:Flux.unitexpression" dmtpe="ivoa:Unit" value="Jy" />
<ATTRIBUTE dmrole="photdm:Flux.value" dmtpe="ivoa:real" value="1024.0" />
</INSTANCE>

<!-- Filter -->
<INSTANCE dmtpe="photdm:PhotometryFilter" dmrole="photdm:PhotCal.photometryFilter">
<ATTRIBUTE dmrole="photdm:PhotometryFilter.fpsidentifier" dmtpe="ivoa:string"
value="ivo://svo/fps" />
<ATTRIBUTE dmrole="photdm:PhotometryFilter.identifier" dmtpe="ivoa:string"
value="2MASS/2MASS.H" />
<ATTRIBUTE dmrole="photdm:PhotometryFilter.name" dmtpe="ivoa:string" value="2MASS H" />
<ATTRIBUTE dmrole="photdm:PhotometryFilter.description" dmtpe="ivoa:string"
value="2MASS H" />
<ATTRIBUTE dmrole="photdm:PhotometryFilter.bandname" dmtpe="ivoa:string" value="H" />

<!-- Spectral Location -->
<INSTANCE dmtpe="photdm:SpectralLocation"
dmrole="photdm:PhotometryFilter.spectralLocation">
<ATTRIBUTE dmrole="photdm:SpectralLocation.ucd" dmtpe="ivoa:UCD"
value="em.wl.effective" />
<ATTRIBUTE dmrole="photdm:SpectralLocation.unitexpression" dmtpe="ivoa:Unit"
value="Angstrom" />
<ATTRIBUTE dmrole="photdm:SpectralLocation.value" dmtpe="ivoa:real" value="16620.0" />
</INSTANCE>

<!-- Band width -->
<INSTANCE dmtpe="photdm:Bandwidth" dmrole="photdm:PhotometryFilter.spectralLocation">
<ATTRIBUTE dmrole="photdm:Bandwidth.ucd" dmtpe="ivoa:UCD"
value="instr.bandwidth;stat.fwhm" />
<ATTRIBUTE dmrole="photdm:Bandwidth.unitexpression" dmtpe="ivoa:Unit" value="Angstrom" />
<ATTRIBUTE dmrole="photdm:Bandwidth.extent" dmtpe="ivoa:real" value="2509.4034987068" />
<ATTRIBUTE dmrole="photdm:Bandwidth.start" dmtpe="ivoa:real" value="14787.378640179" />
<ATTRIBUTE dmrole="photdm:Bandwidth.stop" dmtpe="ivoa:real" value="18231.020407164" />
</INSTANCE>

<!-- Transmission Curve -->
<INSTANCE dmtpe="photdm:TransmissionCurve" dmrole="photdm:Bandwidth.transmissionCurve">
<INSTANCE dmtpe="photdm:Access" dmrole="photdm:TransmissionCurve.access">
<ATTRIBUTE dmrole="photdm:Access.reference" dmtpe="ivoa:anyURI"
value="http://svo2.cab.inta-csic.es/theory/fps/fps.php?PhotCalID=2MASS/2MASS.H/Vega" />
<ATTRIBUTE dmrole="photdm:Access.size" dmtpe="ivoa:integer" value="-1" />
<ATTRIBUTE dmrole="photdm:Access.format" dmtpe="ivoa:string" value="VOTable" />
</INSTANCE>
</INSTANCE>
</INSTANCE>
```

Table Join (advanced feature)

```
<TEMPLATES tableref="_PKTable">
  <INSTANCE dmid="_TimeSeries" dmrole="" dmtype="cube:SparseCube">
    <REFERENCE dmrole="cube:DataProduct.dataset" sourceref="_Datasets">
      <FOREIGN_KEY ref="_pksrcid"/>
    </REFERENCE>
    <COLLECTION dmrole="cube:SparseCube.data">
      <JOIN dmref="_ts_data">
        <WHERE foreignkey="_srcid" primarykey="_pksrcid" />
        <WHERE foreignkey="_band" primarykey="_pkband" />
      </JOIN>
    </COLLECTION>
  </INSTANCE>
</TEMPLATES>
<TEMPLATES tableret="Results">
  <INSTANCE dmid="_ts_data" dmrole="" dmtype="cube:NDPoint">
    <COLLECTION dmrole="cube:NDPoint.observable">
      <INSTANCE dmtype="cube:Observable">
        <ATTRIBUTE dmrole="cube:DataAxis.dependent" dmtype="ivoa:boolean" value="False"/>
      </INSTANCE>
    </COLLECTION>
  </INSTANCE>
</TEMPLATES>
```

The `cube:SparseCube.data` is populated with joined data

- That are instances of the table `Results[@dmid=_ts_data]`
- That use both columns `_band` and `_srcid` of table `Results` as foreign keys
- And both columns `_pkband` and `_pksrcid` of table `PKTable` as primary keys

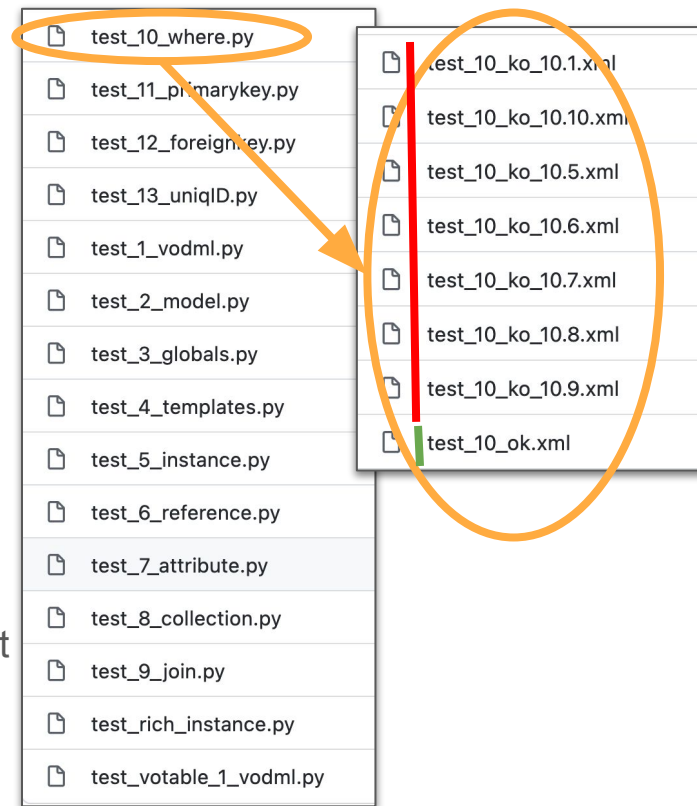
Validation of the Syntax Design

- **Based on many unit tests**

- One test suite per element
 - One snippet file with all matching patterns
 - Some failing snippets
 - Verify that tests fail for the expected reason
- Snippets can be used as a library of standard patterns
 - Discover the syntax (reviewers will thank us)
 - Exercise the annotation

- **Validators**

- All rules are enforced by the schema
 - Any XSD1.1 processor can be used as validator
- Some inconsistencies remain at the discretion of the client
 - Unresolved references
 - Invalid or inconsistent units
 - Mapping not faithful to the schema



Conclusions

- **Document in WD**

- Visit please

<https://wiki.ivoa.net/twiki/bin/view/IVOA/DataAnnotation>

<https://github.com/ivoa-std/ModelInstanceInVot>

- **Rather Satisfied**

- Smooth VOTable integration
- Simple annotations for the simple cases
- Support all use-cases proposed in the workshop
- Working validator

- **Work in progress**

- Provide libraries connected with PyVO
- Find name for the Standard (poll on Slack #dm)
 - For now: ***Model Instance in VOTable*** (short name: ***MIVOT***)

Backup Slides

A Limited Scope for the Model View

- **Modeling Details Ignored**

- Inheritance ignored
- No distinction between association/aggregation/composition
- Data type vs object type

- **Data Hierarchy**

- Complex types (matches the JSON `{...}`)
- Attribute (matches the JSON `key: value`)
- Collection (matches the JSON `[...]`)
- References

Syntax Overview: XML elements

- **Mapping elements**

- Mapping block containers
 - **<VODML>**, **<GLOBALS>** and **<TEMPLATES>**
- Class hierarchy
 - **<INSTANCE>**: complex data type or object type
 - **<ATTRIBUTE>**: simple attribute (atomic value)
 - **<COLLECTION>**: composition of instances or array of attributes
- Data relationships
 - **<REFERENCE>**: Instance reference
 - **<JOIN>**: Data join with either a **<COLLECTION>** or table rows
 - **<WHERE>**: **<JOIN>** condition or table row filter
 - **<PRIMARY_KEY>**: Reference to a value that can be used as primary key for either table rows or **<COLLECTION>** items.
 - **<FOREIGN_KEY>**: Foreign key to be used to resolve a **<REFERENCE>**

VOTable Insertion

- **A strong demand for a shy integration**

- Do not break working things
- Do not bother existing VOTable stakeholders.

- **Encapsulating the mapping block into a <resource>**

- The VOTABLE schema supports resources whose content is not controlled by the schema

- `type=meta`

```
--<!--  
    Suggested Doug Tody, to include new RESOURCE types  
-->  
<xs:any namespace="##other" processContents="lax" minOccurs="0" maxOccurs="unbounded"/>
```

- **2 separate schemas**

- The mapping schema has no connection with the VOTable schema
- Mapped documents must be validated against both schema separately
- To achieve this separation, we took care to use different terms for the content in order to not mislead legacy clients doing XPath-based parsing.

XSD 1.1 in Action

```
<xs:complexType name="PrimaryKey">
  <xs:attribute type="xs:string" name="ref" />
  <xs:attribute type="xs:string" name="dmtype" />
  <xs:attribute type="xs:string" name="value" />
  <xs:assert test="./@ref or ./@value" />
  <xs:assert test="@dmtype != ''" />
  <xs:assert test="if (./@ref) then @ref != '' else true()" />
  <xs:assert test="(./@value and not(./@ref)) or (not(./@value) and ./@ref)" />
</xs:complexType>
```

A **<PrimaryKey>** must have

- a non empty @dmtype
- A @ref or a @value but not both
- @ref must not be empty when present

Example of @attribute patterns

Attribute set with **Laurent** as a literal value

```
<ATTRIBUTE dmrole="test.owner.firstname" dmtype="string" value="Laurent" />
```

Attribute set with the value read from the `_title` table column

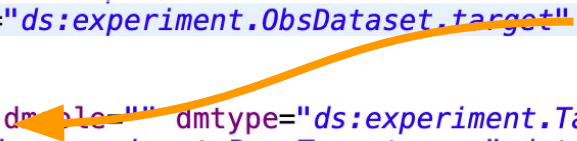
```
<ATTRIBUTE dmrole="test.title" dmtype="string" ref="_title" />
```

Attribute set with a value read in the `_title` if it exists or with **Laurent** as a literal

```
<ATTRIBUTE dmrole="test.title" dmtype="string" ref="_title" value="Default Title"/>
```

Syntax Overview: Reference

```
<COLLECTION dmid="_Datasets" dmrole="">
  <INSTANCE dmid="_ds1" dmrole="" dmtype="ds:experiment.ObsDataset">
    <PRIMARY_KEY dmtype="ivoa:string" value="5813181197970338560"/>
    <ATTRIBUTE dmrole="ds:dataset.Dataset.dataProductType" dmtype="ds:dataset.DataProductType" value="TIMES" />
    <ATTRIBUTE dmrole="ds:dataset.Dataset.dataProductSubtype" dmtype="ivoa:string" value="GAIA Time Series" />
    <ATTRIBUTE dmrole="ds:experiment.ObsDataset.calibLevel" dmtype="ivoa:integer" value="1"/>
    <REFERENCE dmrole="ds:experiment.ObsDataset.target" dmref="_tg1"/>
  </INSTANCE>
</COLLECTION>
<INSTANCE dmid="_tg1" dmrole="" dmtype="ds:experiment.Target">
  <ATTRIBUTE dmrole="ds:experiment.BaseTarget.name" dmtype="ivoa:string" value="5813181197970338560"/>
</INSTANCE>
```



The object

- playing the `ds:experiment.ObsDataset.target` role
- is set with `ds.experiment.Target` instance

Syntax Overview: Join (advanced feature)

- **Flexible JOIN operator**

- Join 2 data collections
 - Static data or **<TABLE>** templates
- Can join **<TABLE>** data together
- Can join **<TABLE>** data with static data collections
- Can join static data collection together
- Support multiple join conditions (**<WHERE>**)

```
<COLLECTION dmid="id5" dmrole="">
  <!-- Test Case 9.6b: sourceref + no dmref + V
  <JOIN sourceref="aaaaa" >
    <WHERE foreignkey="eee" value="ssss" />
    <WHERE foreignkey="fff" value="tttt" />
  </JOIN>
</COLLECTION>
```

sourceref

