VO-DML/Mapping status update

Omar Laurino
SAO

VO-DML status

- CubeDM, DatasetDM, STC2
  - ⭑ Models follow VO-DML rules
  - ⭑ Feedback from modelers was positive: VO-DML facilitates modeling, can inform decisions, and avoids ad-hoc serialization specs for all formats
  - ⭑ Feedback from modelers informed changes in spec
  - ⭑ Translation scripts for Altova and Modelio UML Modelers
  - ⭑ Consistent reuse of models

VO-DML status

- New WD circulated in late September
    - Document was decoupled from Mapping WD even more by moving some sections to Mapping WD
    - Added introduction and renamed old introduction to Data Integration
    - ReferencableElement —> ReferableElement
    - vodml-id SHOULD be human-readable according to the grammar in Appendix D (?)
    - Model/ModelProxy
        - Model does not extend Package
        - Added version, Removed ivoid (?) in ModelProxy

VO-DML status

- Clarifications:
  - Only models directly used must be imported
  - Uniqueness and mutability of DataType instances
  - VO-UML attribute notation
  - Collection/Composition and Reference sections (navigation of instances, life cycles)
  - Subsets/Redefines

Mapping WD

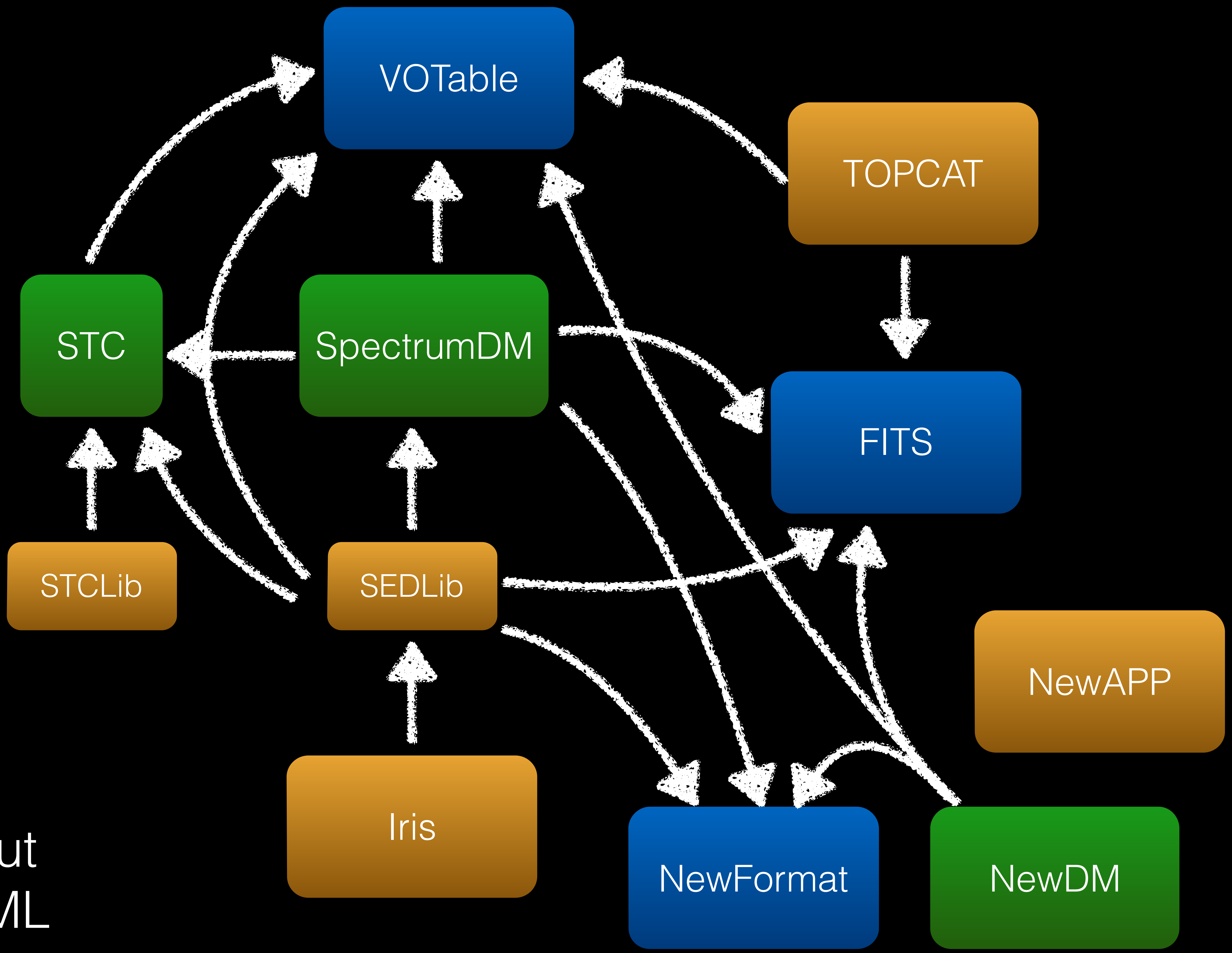
- New version in progress
  - changed title to “Mapping Data Model Instances to VOTable”
  - depends on possible changes to VOTable
  - feedback from Cube, Dataset, STC2 to be folded in:
    - Reference + Reference to external instances
    - Composition/Collection
  - might define some advanced ORM mappings
- Need to verify consensus on globally unique “prefixes”

Implementations

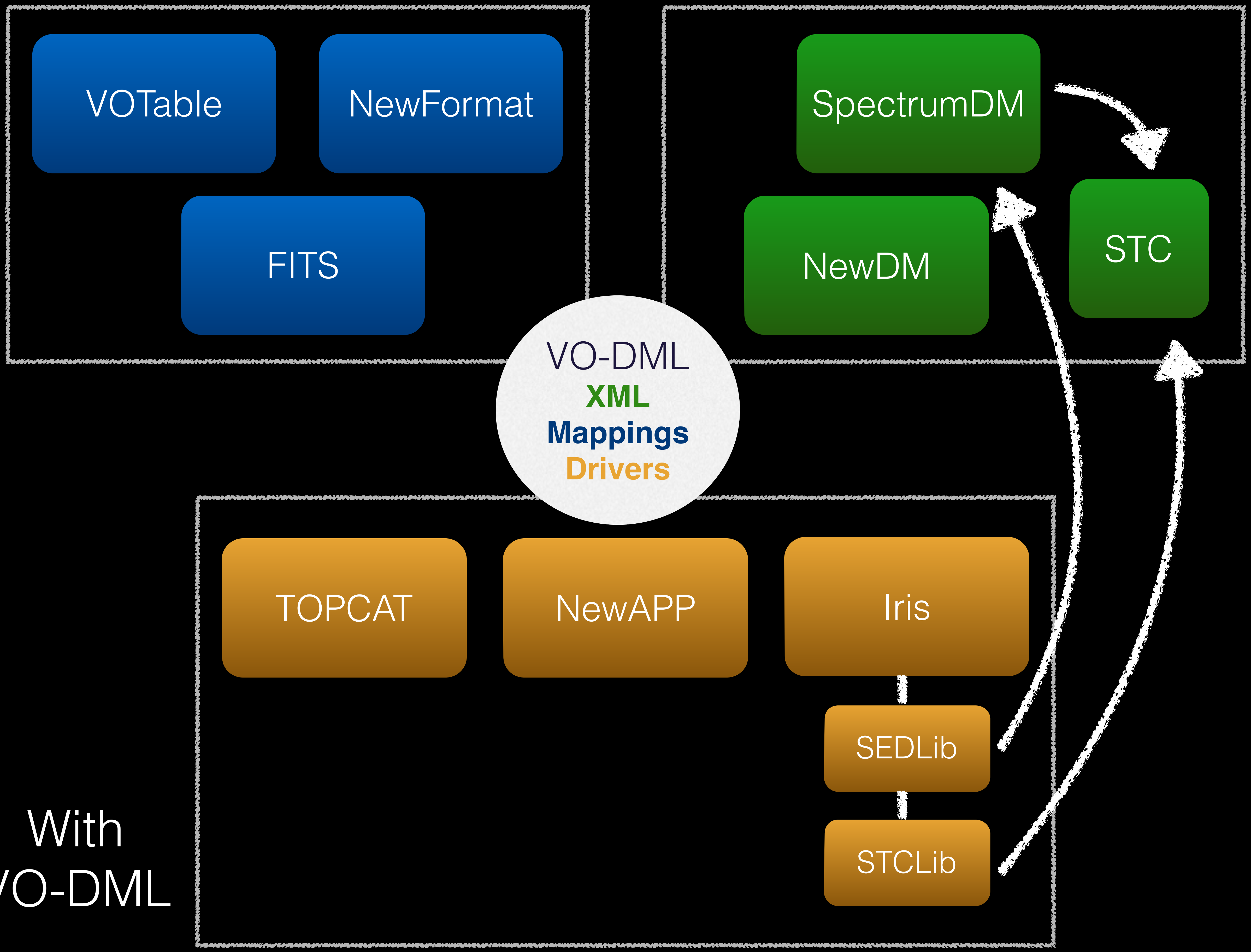
- VO-DML Mapper (Hawaii)
- Eclipse Domain Specific Language (Madrid)
- Python implementation - ~halfway through (Banff)
- Java (de-)serialization - complete (Heidelberg, TBU)
- VO-DML/XML to HTML, TAP Schema (Heidelberg)
- XMI —> VO-DML/XML translation scripts for Modelio, Altova, Magic Draw
- (De-)serialization tutorial

Some lessons learned

- Continuous communication is key:
  - Modeling for interoperability is no simple feat
  - Critical issues must be identified early
  - Iterations must be much shorter than 6 months
  - Personal interaction works best
  - Common repository most useful (please use volute :))
- Consistency matters:
  - Models reuse makes standards light, interoperable, unambiguous
  - Consistent models are easier to read (by both software and wetware)
- Separation of Modeling/Serialization makes life easier for everyone, and is key for interoperability



Without
VO-DML


VOTable
NewFormat
FITS
SpectrumDM
NewDM
STC
VO-DML
XML
Mappings
Drivers
TOPCAT
NewAPP
Iris
SEDLib
STCLib

With
VO-DML

Mapping Data Model instances to VOTable

A Tutorial

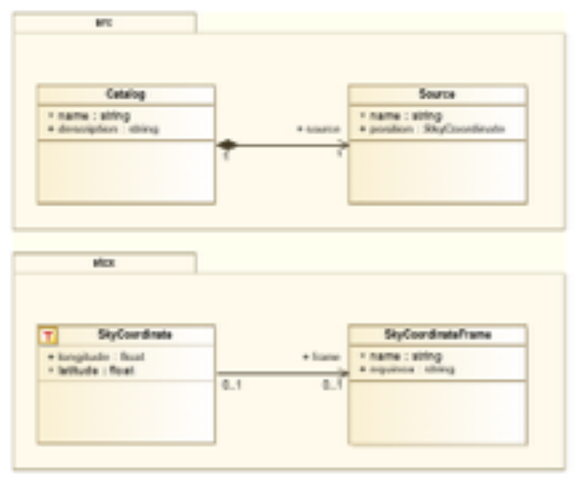
Disclaimer

The following examples assume the current Mapping Document, i.e. the use of @utype.

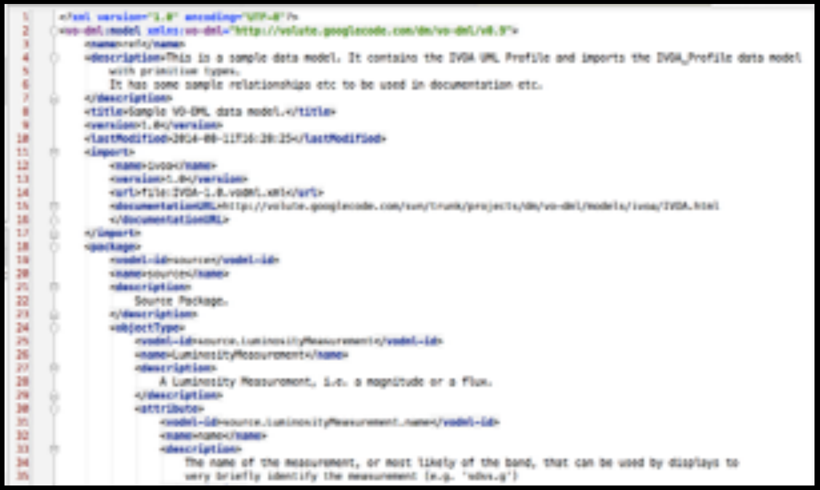
Depending on the conclusions of the VOTable discussion, examples might change **syntactically, but not semantically**, as all options on the table right now are perfectly equivalent.

Data Providers: how to serialize instances

UML



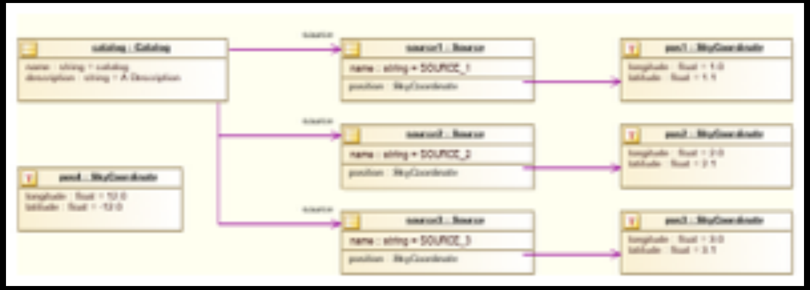
VODML/XML



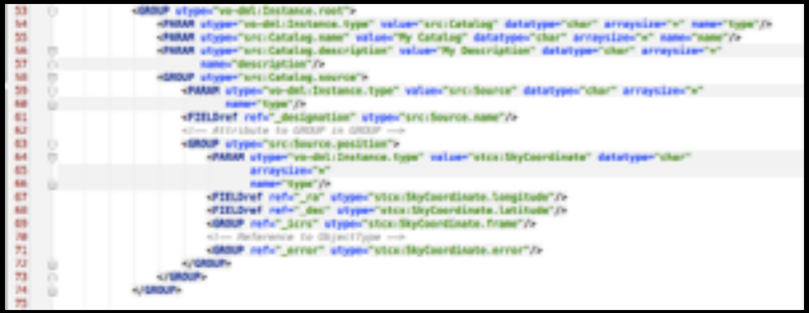
```
<?xml version="1.0" encoding="UTF-8"?>
<vo-dml:model xmlns:vo-dml="http://volute.googlecode.com/dm/vo-dml/v0.9">
    <name>src</name>
    <description>This is a sample data model. It contains the IVOA UML Profile and imports the IVOA_Profile data model
        with primitive types.
        It has some sample relationships etc to be used in documentation etc.
    </description>
    <title>Sample VO-DML data model.</title>
    <version>1.0</version>
    <lastModified>2014-08-11T16:20:25</lastModified>
    <import>
        <name>ivoa</name>
        <version>1.0</version>
        <url>file:IVOA-1.0.vodml.xml</url>
        <documentationURL>http://volute.googlecode.com/svn/trunk/projects/dm/vo-dml/models/ivoa/IVOA.html
        </documentationURL>
    </import>
    <package>
        <vodml-id>source</vodml-id>
        <name>source</name>
        <description>
            Source Package.
        </description>
        <objectType>
            <vodml-id>source.LuminosityMeasurement</vodml-id>
            <name>LuminosityMeasurement</name>
            <description>
                A Luminosity Measurement, i.e. a magnitude or a flux.
            </description>
            <attribute>
                <vodml-id>source.LuminosityMeasurement.name</vodml-id>
                <name>name</name>
                <description>
                    The name of the measurement, or most likely of the band, that can be used by displays to
                    very briefly identify the measurement (e.g. 'sdss.g')
```

Model

UML



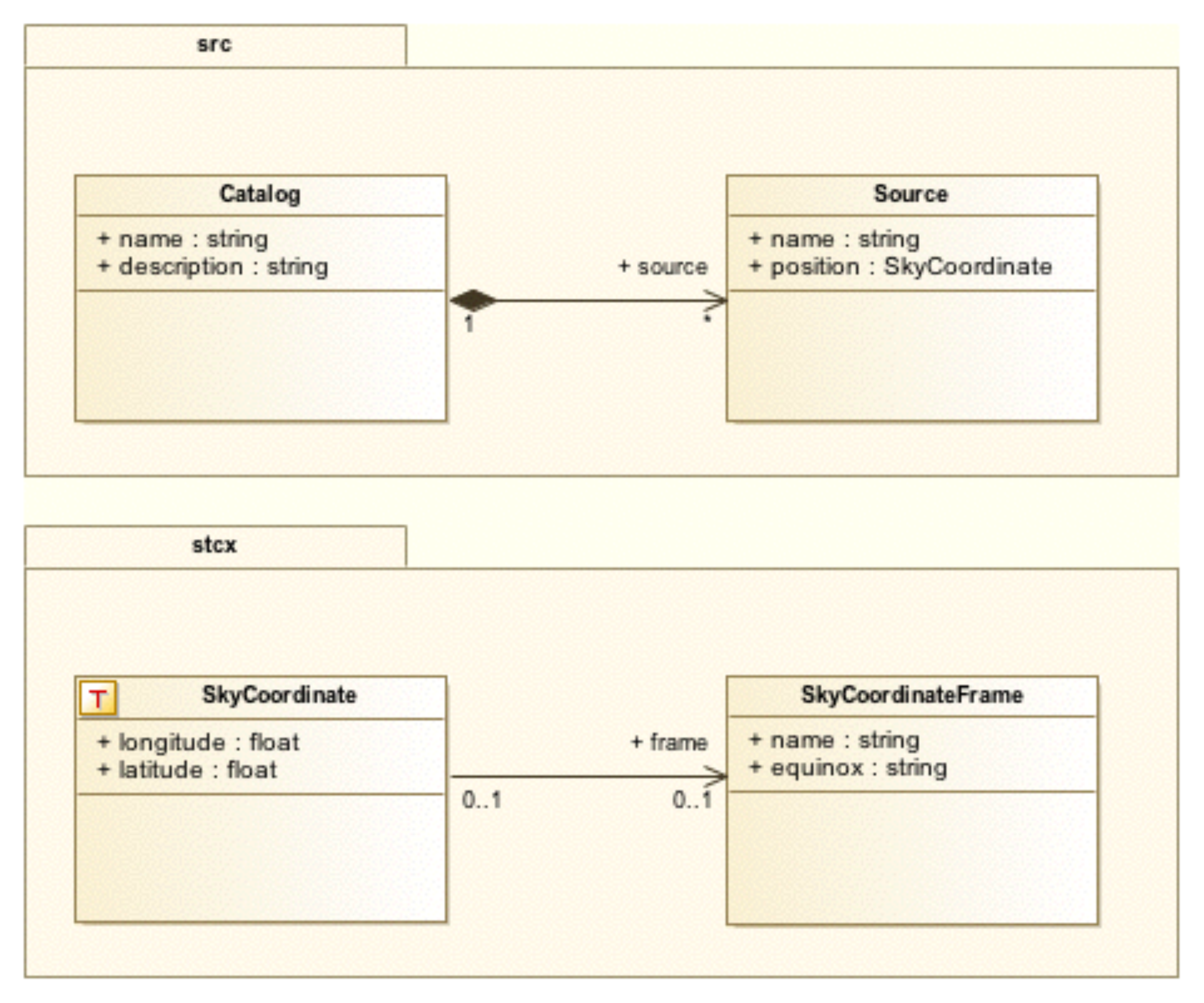
VOTable/Mapping



```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="src:Catalog" datatype="char" arraysize="*" name="type"/>
    <PARAM utype="src:Catalog.name" value="My Catalog" datatype="char" arraysize="*" name="name"/>
    <PARAM utype="src:Catalog.description" value="My Description" datatype="char" arraysize="*"
           name="description"/>
    <GROUP utype="src:Catalog.source">
        <PARAM utype="vo-dml:Instance.type" value="src:Source" datatype="char" arraysize="*"
               name="type"/>
        <FIELDref ref="_designation" utype="src:Source.name"/>
        <!-- Attribute to GROUP in GROUP -->
        <GROUP utype="src:Source.position">
            <PARAM utype="vo-dml:Instance.type" value="stcx:SkyCoordinate" datatype="char"
                   arraysize="*"
                   name="type"/>
            <FIELDref ref="_ra" utype="stcx:SkyCoordinate.longitude"/>
            <FIELDref ref="_dec" utype="stcx:SkyCoordinate.latitude"/>
            <GROUP ref="_icrs" utype="stcx:SkyCoordinate.frame"/>
            <!-- Reference to ObjectType -->
            <GROUP ref="_error" utype="stcx:SkyCoordinate.error"/>
        </GROUP>
    </GROUP>
</GROUP>
```

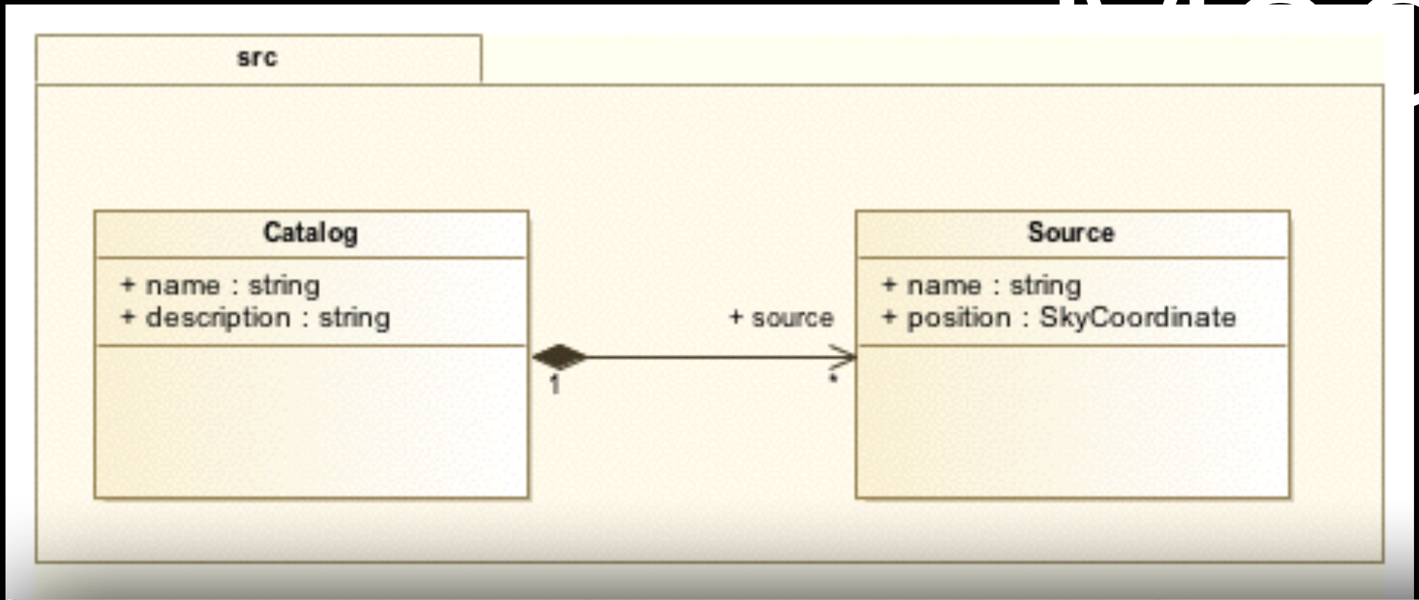
Instances

Model

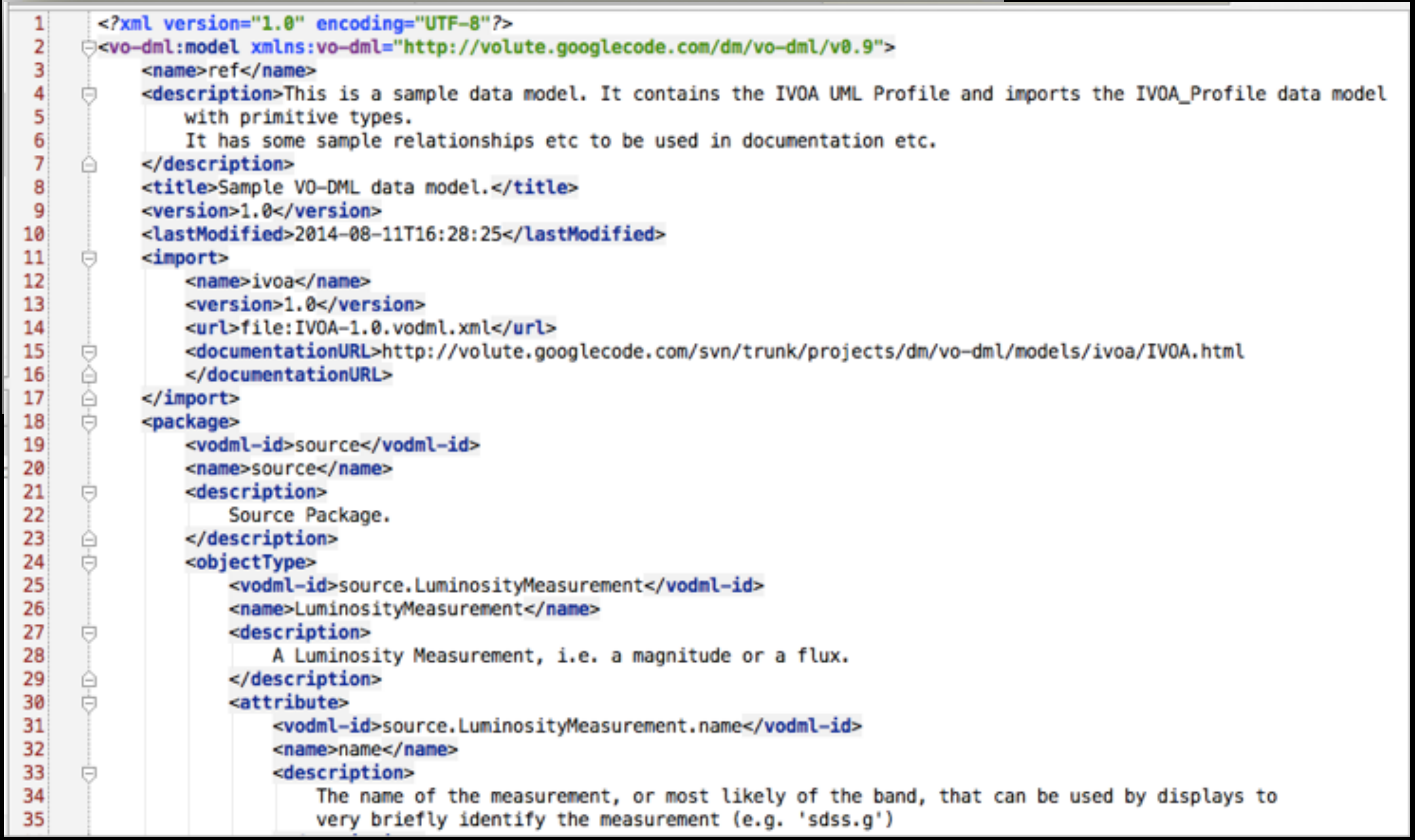


UML
Class Diagram

Model



A VO-DML/XML document unambiguously describes a Data Model and defines portable identifiers for each element in the DM



```
<?xml version="1.0" encoding="UTF-8"?>
<vo-dml:model xmlns:vo-dml="http://volute.googlecode.com/dm/vo-dml/v0.9">
    <name>ref</name>
    <description>This is a sample data model. It contains the IVOA UML Profile and imports the IVOA_Profile data model
        with primitive types.
        It has some sample relationships etc to be used in documentation etc.
    </description>
    <title>Sample VO-DML data model.</title>
    <version>1.0</version>
    <lastModified>2014-08-11T16:28:25</lastModified>
    <import>
        <name>ivoa</name>
        <version>1.0</version>
        <url>file:IVOA-1.0.vodml.xml</url>
        <documentationURL>http://volute.googlecode.com/svn/trunk/projects/dm/vo-dml/models/ivoa/IVOA.html
        </documentationURL>
    </import>
    <package>
        <vodml-id>source</vodml-id>
        <name>source</name>
        <description>
            Source Package.
        </description>
        <objectType>
            <vodml-id>source.LuminosityMeasurement</vodml-id>
            <name>LuminosityMeasurement</name>
            <description>
                A Luminosity Measurement, i.e. a magnitude or a flux.
            </description>
            <attribute>
                <vodml-id>source.LuminosityMeasurement.name</vodml-id>
                <name>name</name>
                <description>
                    The name of the measurement, or most likely of the band, that can be used by displays to
                    very briefly identify the measurement (e.g. 'sdss.g')
```

stc:SkyCoordinate

stc:SkyCoordinate.longitude

stc:SkyCoordinate.latitude

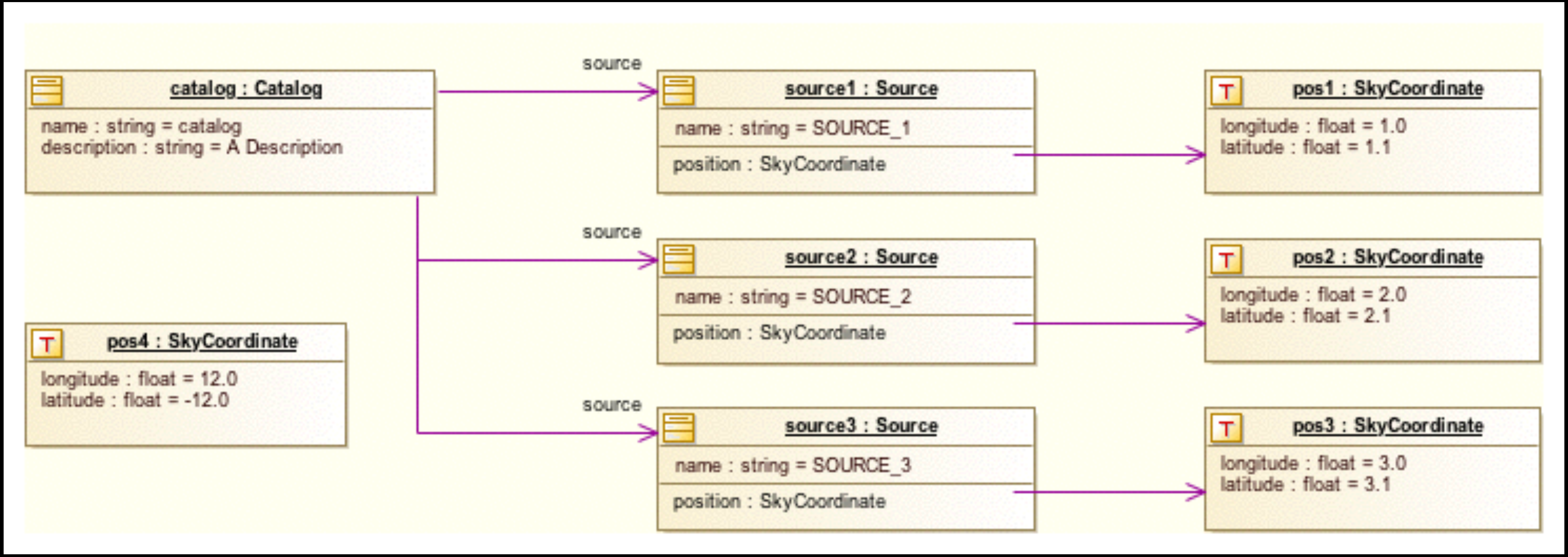
stc:SkyCoordinate.frame

src:Source

src:Source.name

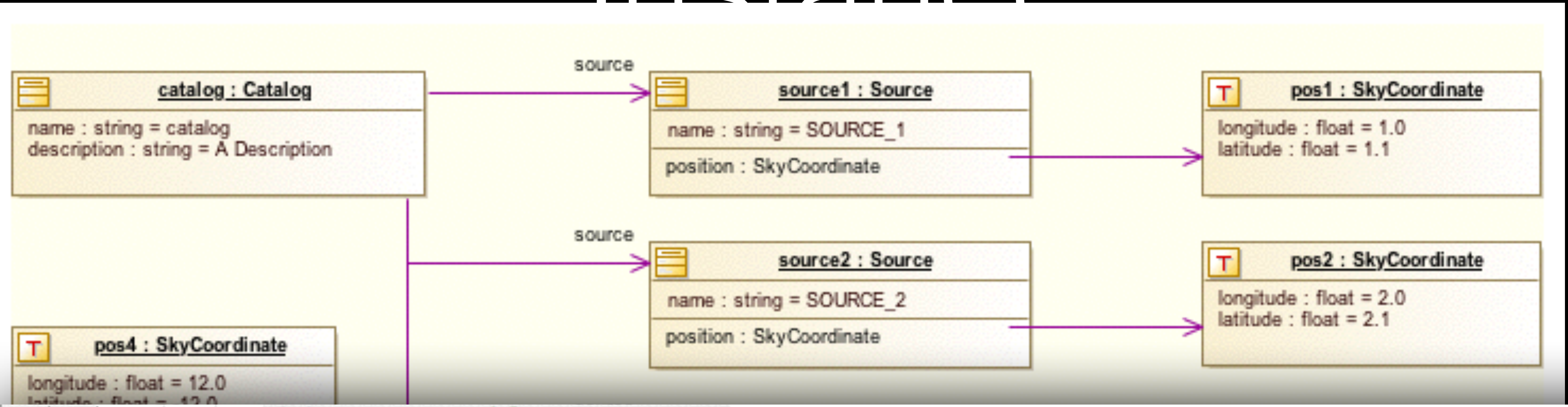
src:Source.position

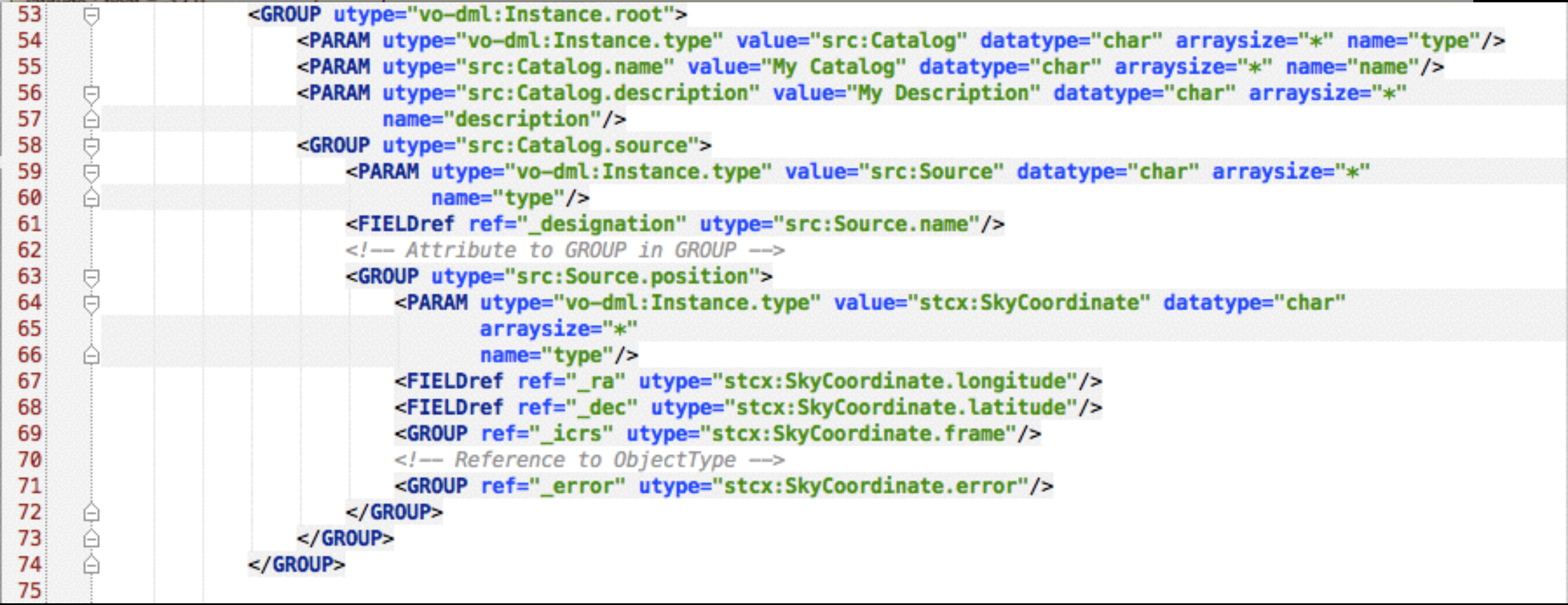
Instance



UML
Object Diagram

Instance



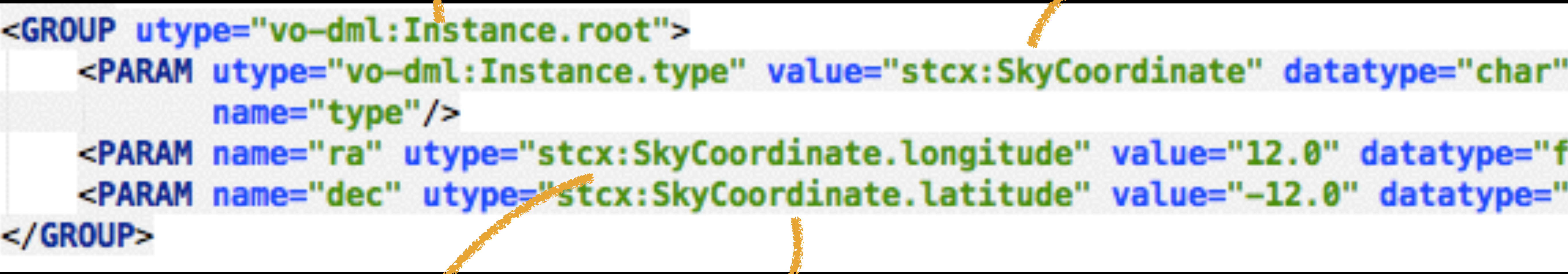


```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="src:Catalog" datatype="char" arraysize="*" name="type"/>
    <PARAM utype="src:Catalog.name" value="My Catalog" datatype="char" arraysize="*" name="name"/>
    <PARAM utype="src:Catalog.description" value="My Description" datatype="char" arraysize="*"
           name="description"/>
    <GROUP utype="src:Catalog.source">
        <PARAM utype="vo-dml:Instance.type" value="src:Source" datatype="char" arraysize="*"
               name="type"/>
        <FIELDref ref="_designation" utype="src:Source.name"/>
        <!-- Attribute to GROUP in GROUP -->
        <GROUP utype="src:Source.position">
            <PARAM utype="vo-dml:Instance.type" value="stcx:SkyCoordinate" datatype="char"
                   arraysize="*"
                   name="type"/>
            <FIELDref ref="_ra" utype="stcx:SkyCoordinate.longitude"/>
            <FIELDref ref="_dec" utype="stcx:SkyCoordinate.latitude"/>
            <GROUP ref="_icrs" utype="stcx:SkyCoordinate.frame"/>
            <!-- Reference to ObjectType -->
            <GROUP ref="_error" utype="stcx:SkyCoordinate.error"/>
        </GROUP>
    </GROUP>
</GROUP>
```

Simple instance example

Start of a VO-DML instance

Instance's type

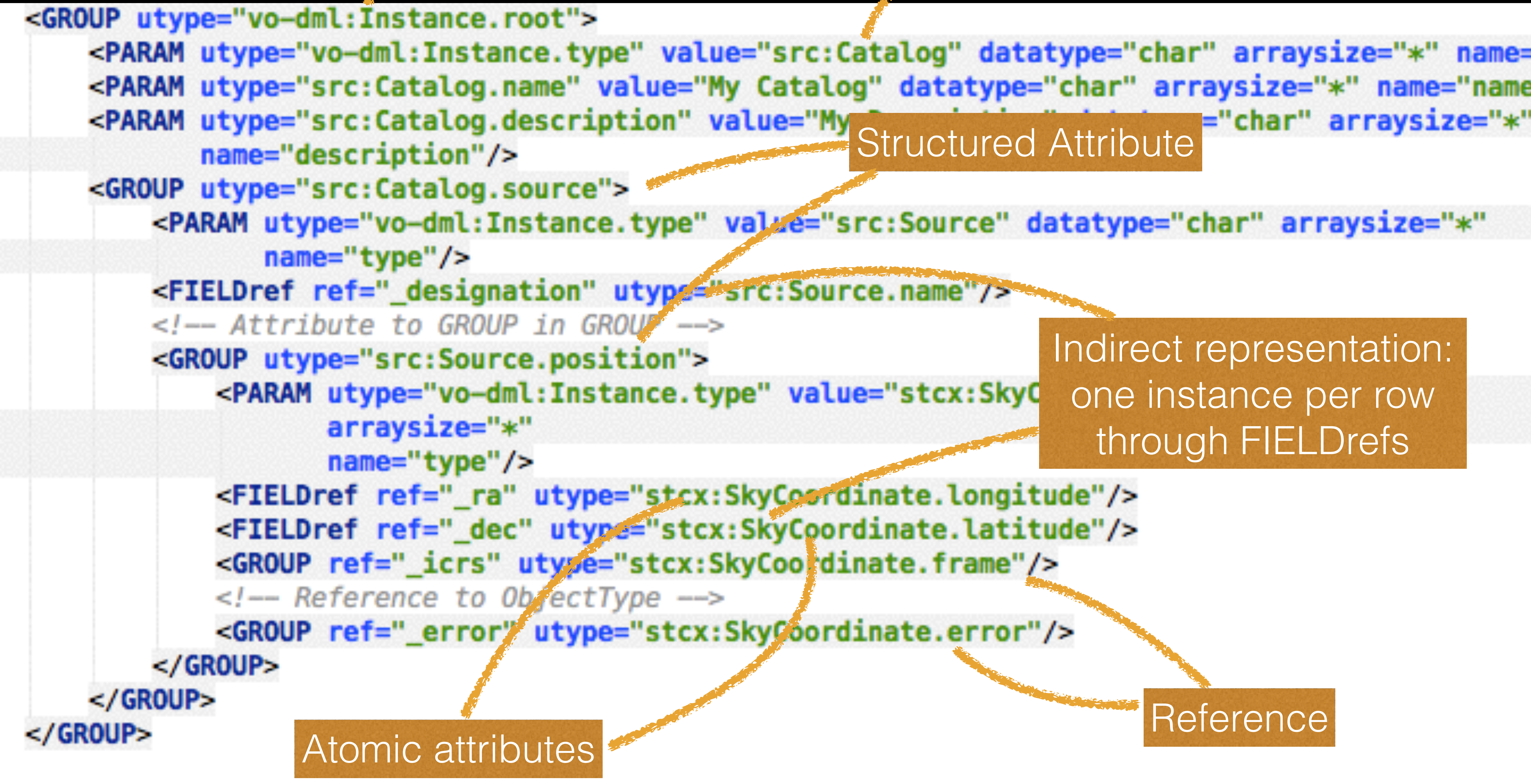


```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="stcx:SkyCoordinate" datatype="char"
           name="type"/>
    <PARAM name="ra" utype="stcx:SkyCoordinate.longitude" value="12.0" datatype="f
    <PARAM name="dec" utype="stcx:SkyCoordinate.latitude" value="-12.0" datatype="
</GROUP>
```

Atomic attributes

Direct representation:
all values defined by PARAMs

Complex Instance example



```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="src:Catalog" datatype="char" arraysize="*" name=
    <PARAM utype="src:Catalog.name" value="My Catalog" datatype="char" arraysize="*" name="name
    <PARAM utype="src:Catalog.description" value="My                    ="char" arraysize="*"
        name="description"/>
    <GROUP utype="src:Catalog.source">
        <PARAM utype="vo-dml:Instance.type" value="src:Source" datatype="char" arraysize="*"
            name="type"/>
        <FIELDref ref="_designation" utype="src:Source.name"/>
        <!-- Attribute to GROUP in GROUP -->
        <GROUP utype="src:Source.position">
            <PARAM utype="vo-dml:Instance.type" value="stcx:SkyC
                arraysize="*"
                name="type"/>
            <FIELDref ref="_ra" utype="stcx:SkyCoordinate.longitude"/>
            <FIELDref ref="_dec" utype="stcx:SkyCoordinate.latitude"/>
            <GROUP ref="_icrs" utype="stcx:SkyCoordinate.frame"/>
            <!-- Reference to ObjectType -->
            <GROUP ref="_error" utype="stcx:SkyCoordinate.error"/>
        </GROUP>
    </GROUP>
</GROUP>
```

Start of a VO-DML instance

Instance's type

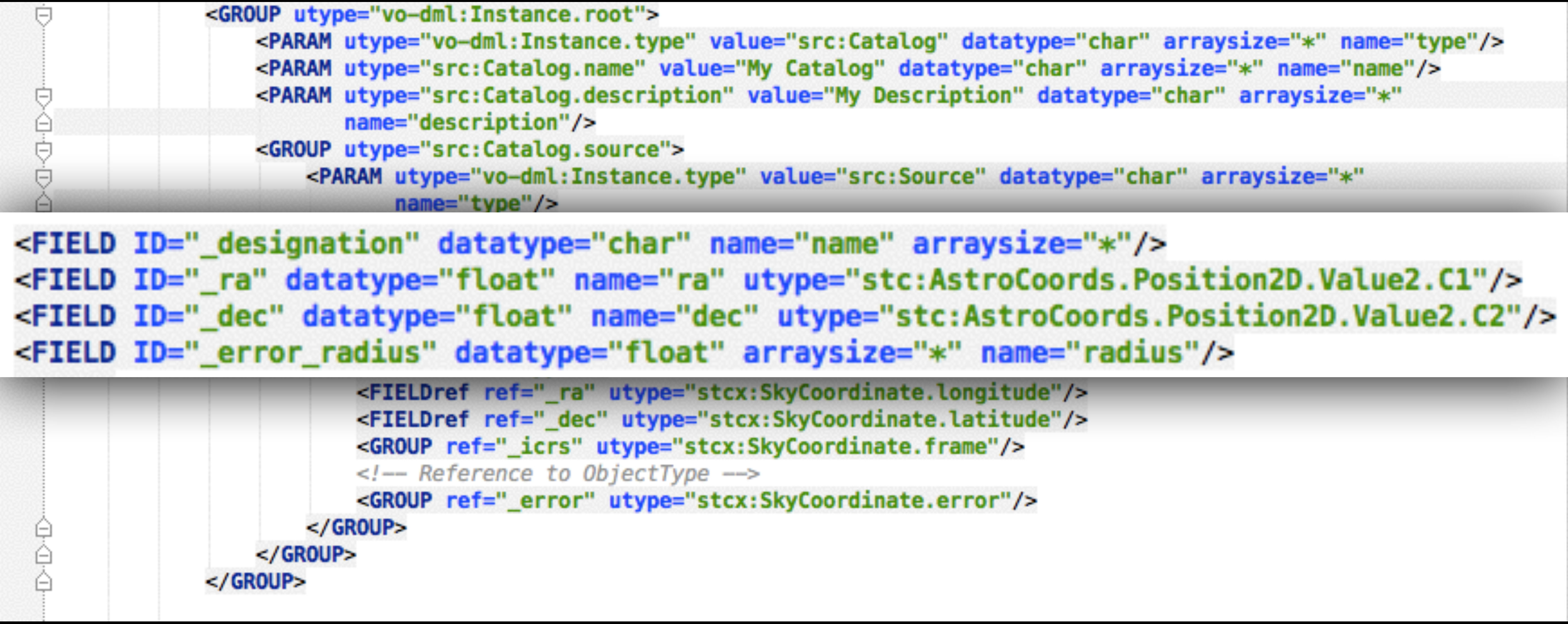
Structured Attribute

Indirect representation: one instance per row through FIELDrefs

Reference

Atomic attributes

Backward Compatibility



```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="src:Catalog" datatype="char" arraysize="*" name="type"/>
    <PARAM utype="src:Catalog.name" value="My Catalog" datatype="char" arraysize="*" name="name"/>
    <PARAM utype="src:Catalog.description" value="My Description" datatype="char" arraysize="*"
           name="description"/>
    <GROUP utype="src:Catalog.source">
        <PARAM utype="vo-dml:Instance.type" value="src:Source" datatype="char" arraysize="*"
               name="type"/>
```

```
<FIELD ID="_designation" datatype="char" name="name" arraysize="*"/>
<FIELD ID="_ra" datatype="float" name="ra" utype="stc:AstroCoords.Position2D.Value2.C1"/>
<FIELD ID="_dec" datatype="float" name="dec" utype="stc:AstroCoords.Position2D.Value2.C2"/>
<FIELD ID="_error_radius" datatype="float" arraysize="*" name="radius"/>
```

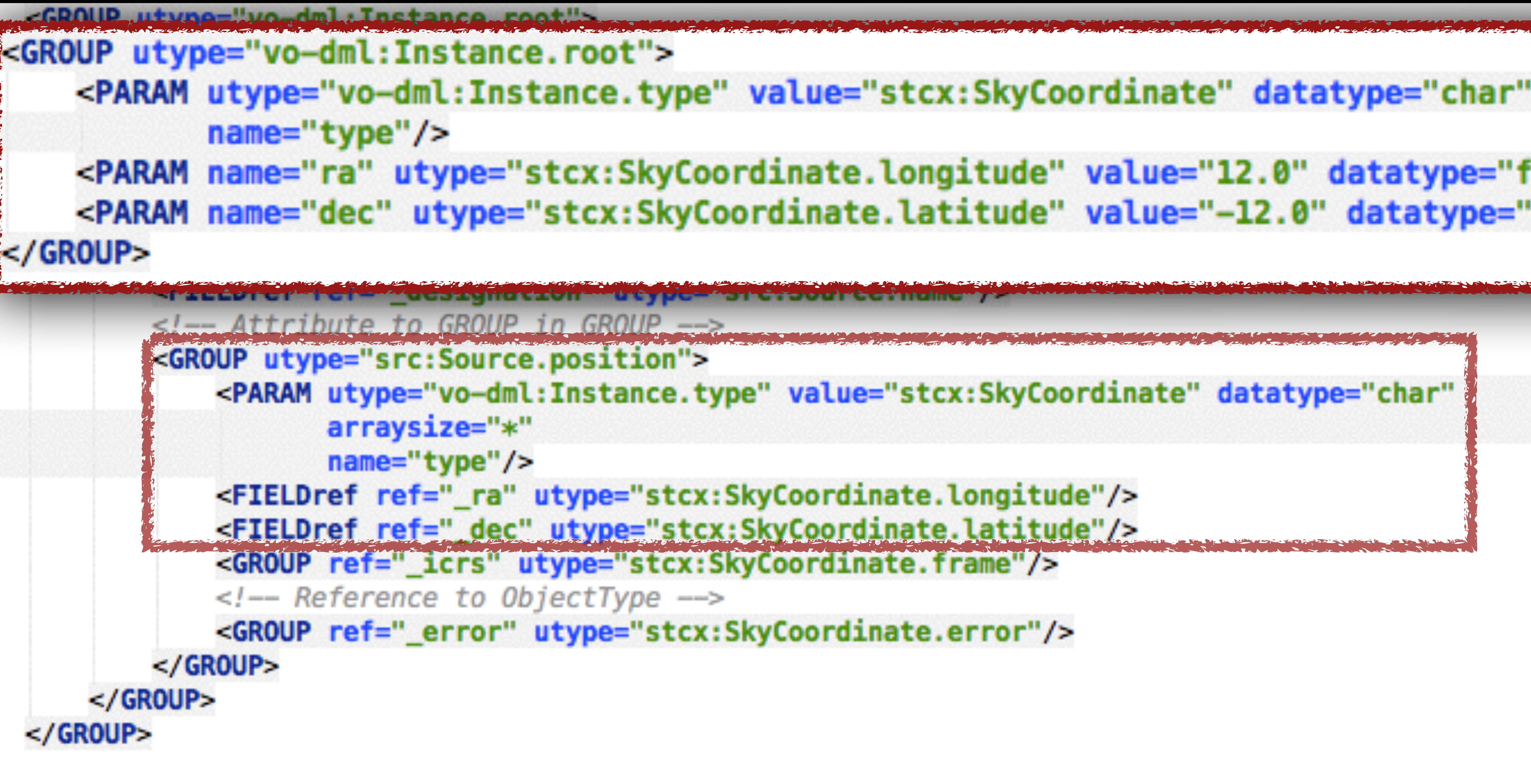
```
                <FIELDref ref="_ra" utype="stcx:SkyCoordinate.longitude"/>
                <FIELDref ref="_dec" utype="stcx:SkyCoordinate.latitude"/>
                <GROUP ref="_icrs" utype="stcx:SkyCoordinate.frame"/>
                <!-- Reference to ObjectType -->
                <GROUP ref="_error" utype="stcx:SkyCoordinate.error"/>
            </GROUP>
        </GROUP>
</GROUP>
```

Accommodates old- and new-style annotations

No name clashes

Multiple views in same serialization

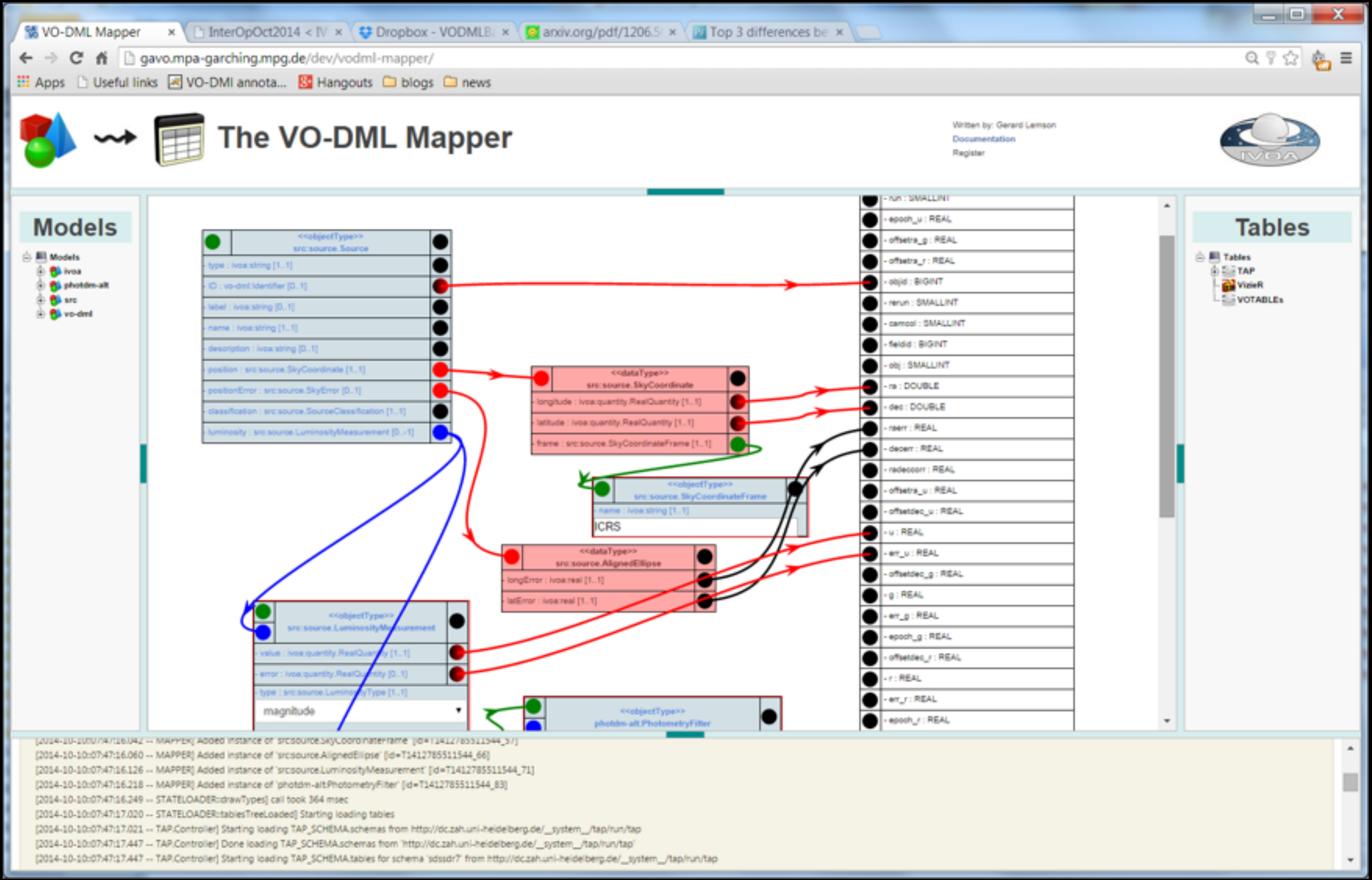
Serialization patterns



```
<GROUP utype="vo-dml:Instance.root">
    <PARAM utype="vo-dml:Instance.type" value="stcx:SkyCoordinate" datatype="char"
           name="type"/>
    <PARAM name="ra" utype="stcx:SkyCoordinate.longitude" value="12.0" datatype="f
    <PARAM name="dec" utype="stcx:SkyCoordinate.latitude" value="-12.0" datatype="
</GROUP>
```

```
<GROUP utype="vo-dml:Instance.root">
        <FIELDref ref="_designation" utype="src:Source.name"/>
        <!-- Attribute to GROUP in GROUP -->
        <GROUP utype="src:Source.position">
            <PARAM utype="vo-dml:Instance.type" value="stcx:SkyCoordinate" datatype="char"
                   arraysize="*"
                   name="type"/>
            <FIELDref ref="_ra" utype="stcx:SkyCoordinate.longitude"/>
            <FIELDref ref="_dec" utype="stcx:SkyCoordinate.latitude"/>
            <GROUP ref="_icrs" utype="stcx:SkyCoordinate.frame"/>
            <!-- Reference to ObjectType -->
            <GROUP ref="_error" utype="stcx:SkyCoordinate.error"/>
        </GROUP>
    </GROUP>
</GROUP>
```

Or Simply Use VODML Mapper

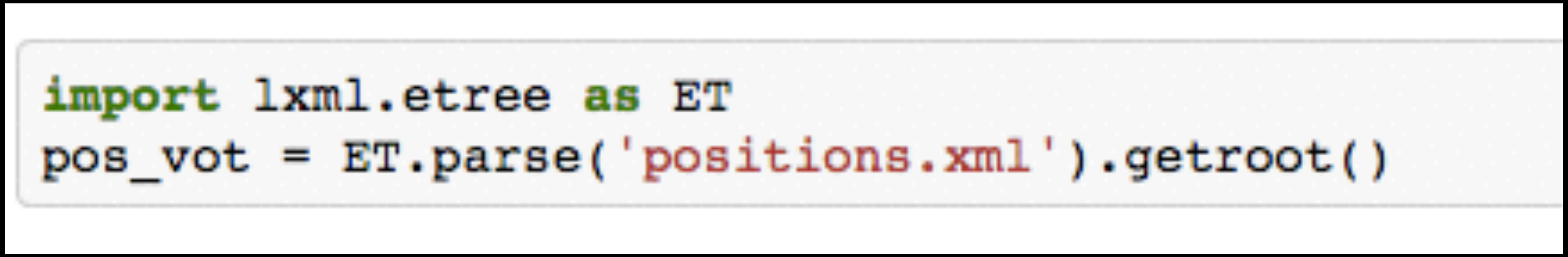


VO-DML Mapper
InterOpOct2014 < IV
Dropbox - VODMLB
arxiv.org/pdf/1206.5
Top 3 differences be
gavo.mpa-garching.mpg.de/dev/vodml-mapper/
Apps
Useful links
VO-DMl annota...
Hangouts
blogs
news
The VO-DML Mapper
Written by: Gerard Lemson
Documentation
Register
IVOA
Models
Models
ivoa
photdm-alt
src
vo-dml
Tables
Tables
TAP
VizieR
VOTABLEs
<<objectType>>
src:source.Source
- type : ivoa:string [1..1]
- ID : vo-dml:Identifier [0..1]
- label : ivoa:string [0..1]
- name : ivoa:string [1..1]
- description : ivoa:string [0..1]
- position : src:source.SkyCoordinate [1..1]
- positionError : src:source.SkyError [0..1]
- classification : src:source.SourceClassification [1..1]
- luminosity : src:source.LuminosityMeasurement [0..-1]
<<dataType>>
src:source.SkyCoordinate
- longitude : ivoa:quantity.RealQuantity [1..1]
- latitude : ivoa:quantity.RealQuantity [1..1]
- frame : src:source.SkyCoordinateFrame [1..1]
<<objectType>>
src:source.SkyCoordinateFrame
- name : ivoa:string [1..1]
ICRS
<<dataType>>
src:source.AlignedEllipse
- longError : ivoa:real [1..1]
- latError : ivoa:real [1..1]
<<objectType>>
src:source.LuminosityMeasurement
- value : ivoa:quantity.RealQuantity [1..1]
- error : ivoa:quantity.RealQuantity [0..1]
- type : src:source.LuminosityType [1..1]
magnitude
<<objectType>>
photdm-alt:PhotometryFilter
- run : SMALLINT
- epoch_u : REAL
- offsetra_g : REAL
- offsetra_r : REAL
- objid : BIGINT
- rerun : SMALLINT
- camcol : SMALLINT
- fieldid : BIGINT
- obj : SMALLINT
- ra : DOUBLE
- dec : DOUBLE
- raerr : REAL
- decerr : REAL
- radeccorr : REAL
- offsetra_u : REAL
- offsetdec_u : REAL
- u : REAL
- err_u : REAL
- offsetdec_g : REAL
- g : REAL
- err_g : REAL
- epoch_g : REAL
- offsetdec_r : REAL
- r : REAL
- err_r : REAL
- epoch_r : REAL

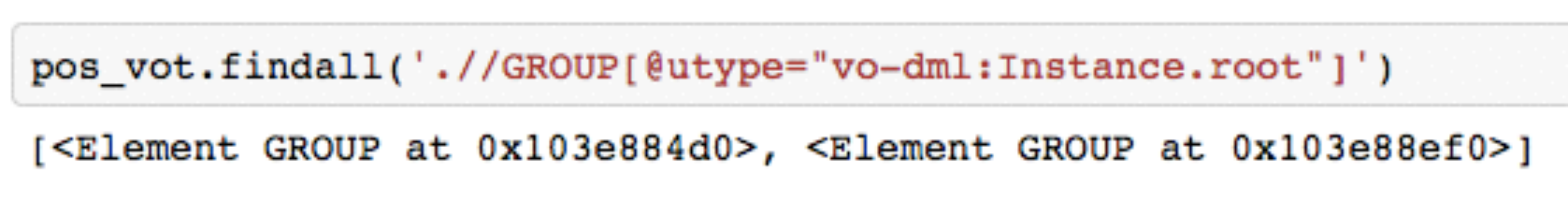
```
[2014-10-10:07:47:16.042 -- MAPPER] Added instance of 'src:source.SkyCoordinateFrame' [id=T1412785511544_57]
[2014-10-10:07:47:16.060 -- MAPPER] Added instance of 'src:source.AlignedEllipse' [id=T1412785511544_66]
[2014-10-10:07:47:16.126 -- MAPPER] Added instance of 'src:source.LuminosityMeasurement' [id=T1412785511544_71]
[2014-10-10:07:47:16.218 -- MAPPER] Added instance of 'photdm-alt:PhotometryFilter' [id=T1412785511544_83]
[2014-10-10:07:47:16.249 -- STATELOADER::drawTypes] call took 364 msec
[2014-10-10:07:47:17.020 -- STATELOADER::tablesTreeLoaded] Starting loading tables
[2014-10-10:07:47:17.021 -- TAP.Controller] Starting loading TAP_SCHEMA.schemas from http://dc.zah.uni-heidelberg.de/__system__/tap/run/tap
[2014-10-10:07:47:17.447 -- TAP.Controller] Done loading TAP_SCHEMA.schemas from 'http://dc.zah.uni-heidelberg.de/__system__/tap/run/tap'
[2014-10-10:07:47:17.447 -- TAP.Controller] Starting loading TAP_SCHEMA.tables for schema 'sdssdr7' from http://dc.zah.uni-heidelberg.de/__system__/tap/run/tap
```

Clients: how to deserialize instances (with DIY XPATH and Python)

Find all VO-DML instances



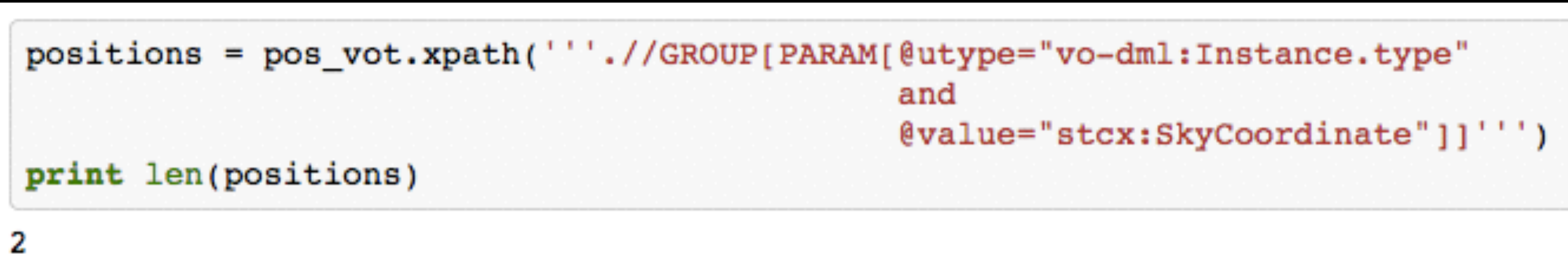
```
import lxml.etree as ET
pos_vot = ET.parse('positions.xml').getroot()
```



```
pos_vot.findall('.//GROUP[@utype="vo-dml:Instance.root"]')
```

```
[<Element GROUP at 0x103e884d0>, <Element GROUP at 0x103e88ef0>]
```

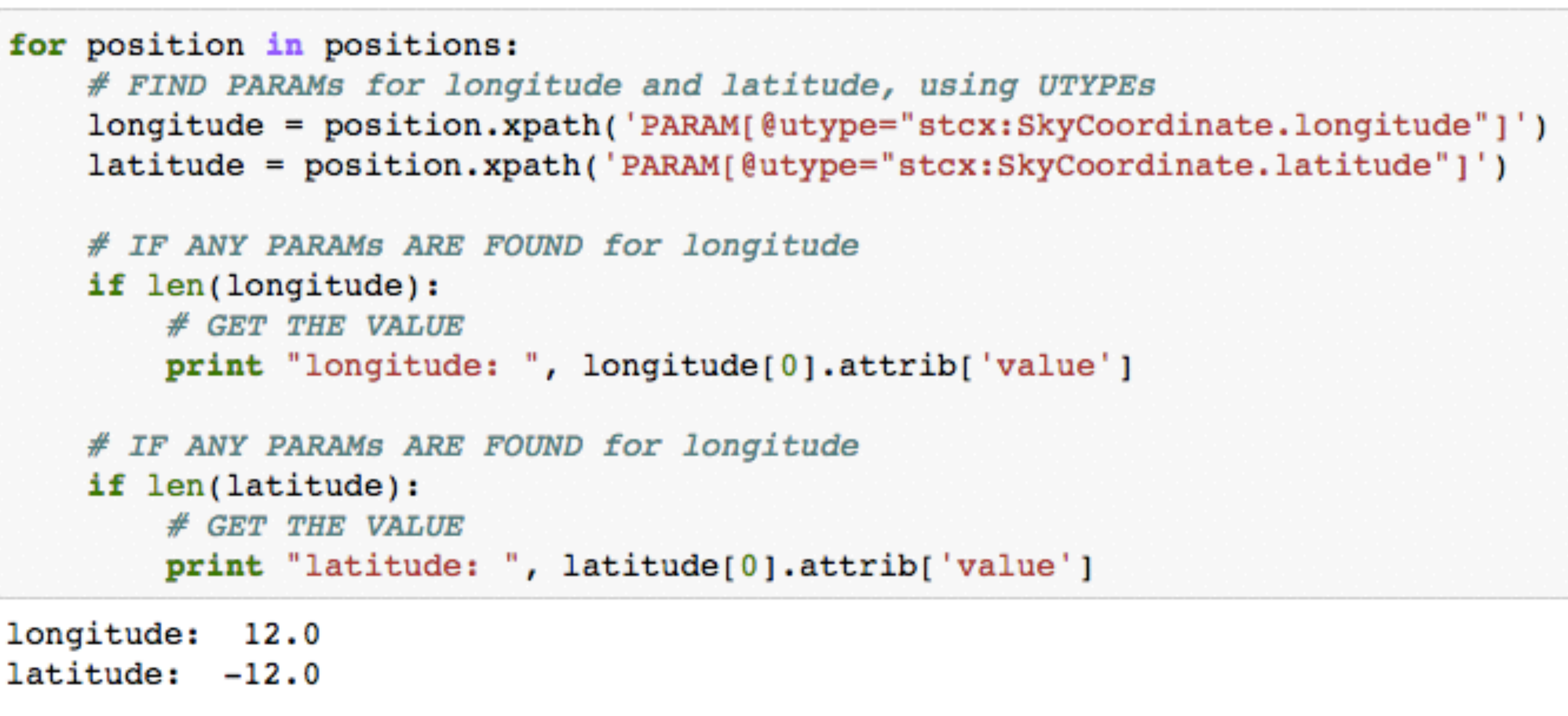
Find all VO-DML instances of a specific type



```
positions = pos_vot.xpath('''.//GROUP[PARAM[@utype="vo-dml:Instance.type"
                                            and
                                            @value="stcx:SkyCoordinate"]]''')
print len(positions)
```

```
2
```

De-serialize Direct Representation



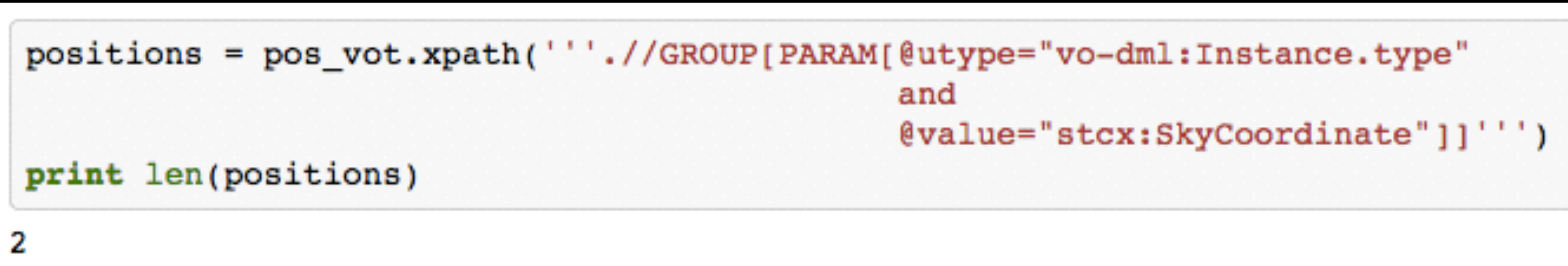
```
for position in positions:
    # FIND PARAMs for longitude and latitude, using UTYPEs
    longitude = position.xpath('PARAM[@utype="stcx:SkyCoordinate.longitude"]')
    latitude = position.xpath('PARAM[@utype="stcx:SkyCoordinate.latitude"]')

    # IF ANY PARAMs ARE FOUND for longitude
    if len(longitude):
        # GET THE VALUE
        print "longitude: ", longitude[0].attrib['value']

    # IF ANY PARAMs ARE FOUND for longitude
    if len(latitude):
        # GET THE VALUE
        print "latitude: ", latitude[0].attrib['value']
```

```
longitude:  12.0
latitude:  -12.0
```

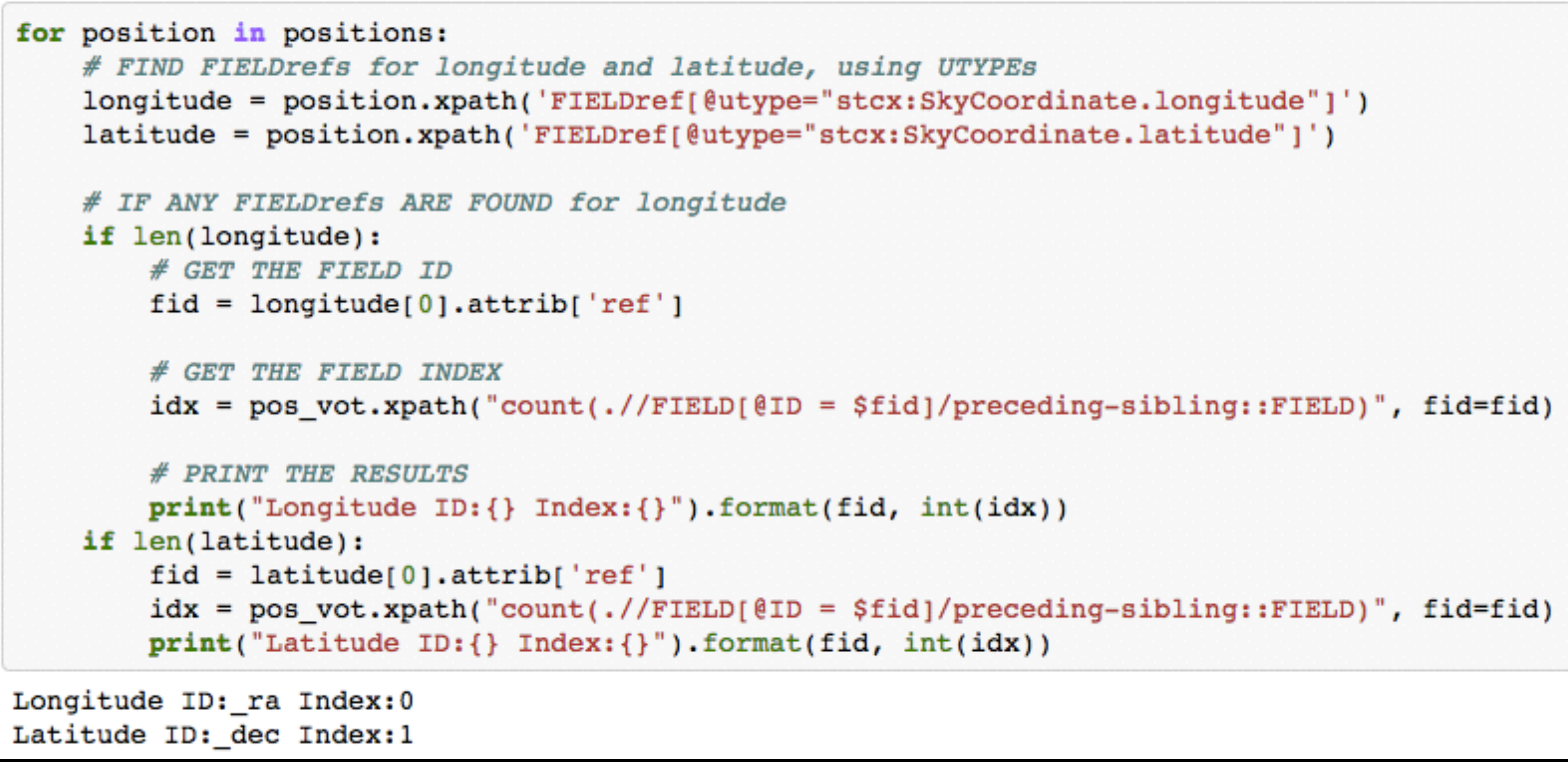
Find all VO-DML instances of a specific type



```
positions = pos_vot.xpath('''.//GROUP[PARAM[@utype="vo-dml:Instance.type"
                                            and
                                            @value="stcx:SkyCoordinate"]]''')
print len(positions)
```

```
2
```

De-serialize Indirect Representation



```
for position in positions:
    # FIND FIELDrefs for longitude and latitude, using UTYPEs
    longitude = position.xpath('FIELDref[@utype="stcx:SkyCoordinate.longitude"]')
    latitude = position.xpath('FIELDref[@utype="stcx:SkyCoordinate.latitude"]')

    # IF ANY FIELDrefs ARE FOUND for longitude
    if len(longitude):
        # GET THE FIELD ID
        fid = longitude[0].attrib['ref']

        # GET THE FIELD INDEX
        idx = pos_vot.xpath("count(.//FIELD[@ID = $fid]/preceding-sibling::FIELD)", fid=fid)

        # PRINT THE RESULTS
        print("Longitude ID:{} Index:{}").format(fid, int(idx))
    if len(latitude):
        fid = latitude[0].attrib['ref']
        idx = pos_vot.xpath("count(.//FIELD[@ID = $fid]/preceding-sibling::FIELD)", fid=fid)
        print("Latitude ID:{} Index:{}").format(fid, int(idx))
```

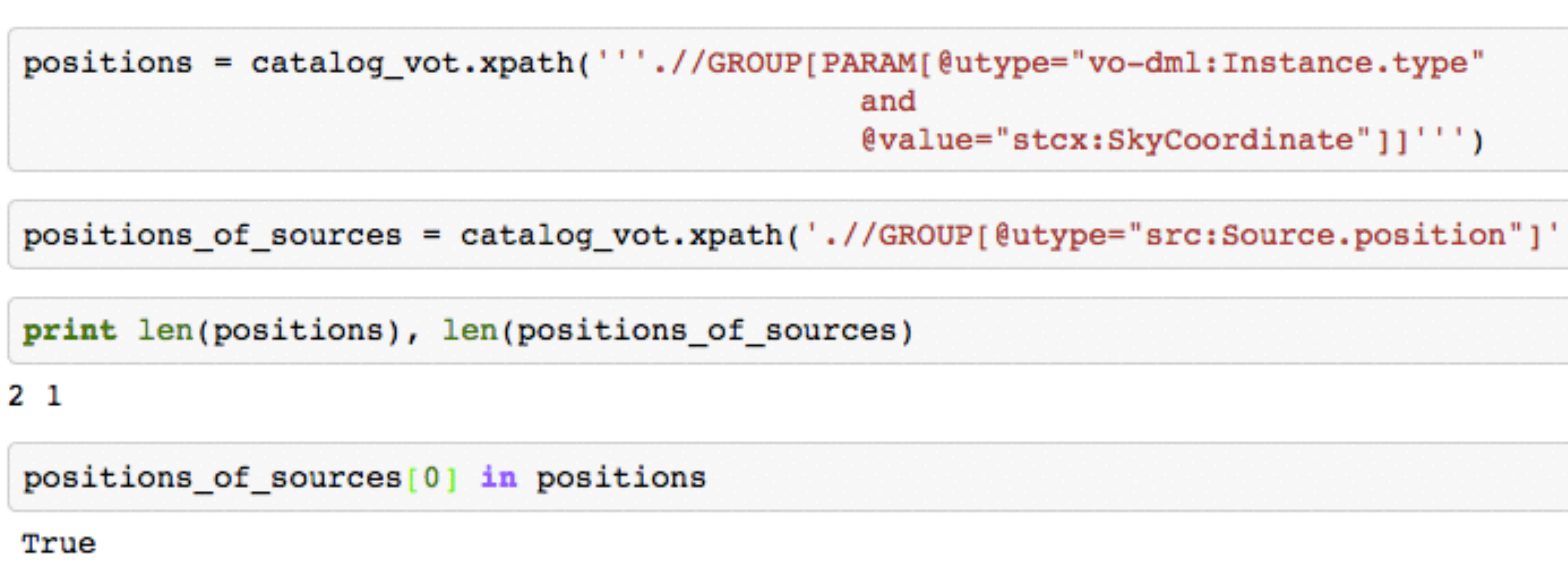
```
Longitude ID:_ra Index:0
Latitude ID:_dec Index:1
```

Contexts

‘The position of a source’

vs

‘a position’



```
positions = catalog_vot.xpath('''.//GROUP[PARAM[@utype="vo-dml:Instance.type"
                                            and
                                            @value="stcx:SkyCoordinate"]]''')
```

```
positions_of_sources = catalog_vot.xpath('.//GROUP[@utype="src:Source.position"]'
```

```
print len(positions), len(positions_of_sources)
```

```
2 1
```

```
positions_of_sources[0] in positions
```

```
True
```

Summary

- Serialization (data providers) is straightforward if:
  - ★ the model makes sense
  - ★ the instances have a clear representation in terms of the model
- Basic serialization patterns are simple and intuitive
- Advanced Object-Relational Mapping patterns are possible if required (multiple tables with foreign keys, etc.)

Summary

- Complete client implementation requires more effort, but:
  - De-serialization strategy is model-agnostic
  - I/O libraries are so straightforward they can be generated by a machine and reused as modules
- Most explicit syntax describes instances for both their role and type
- Custom/old usages allowed without clashes

Tutorial

http://nbviewer.ipython.org/github/olaurino/vo-dml/blob/master/MappingTutorial.ipynb

Thanks!