Mango: A Component and Association Based Model for Source Data

Version 1.0

IVOA Note 2020-04-22

Author(s)
> François Bonnarel, Gilles Landais, Laurent Michel, Jesus Salgado, Mireille Louys, Marco Molinaro

Editor(s)
> Laurent Michel

Abstract

The Mango model proposes a flexible way to make source data interoperable. It takes into account the huge diversity of source data in term of both format and usage. Each Mango source is defined by an identifier on which one can attach a set of parameters and a set of associated data. Parameters are modeled by MCT both native and extended classes. Their roles are given by UCDs and semantic tags. Associated data can be simple URLs, VO service endpoints or VO data model instances. Their roles are given by semantic tags.

Status of this document

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

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

Contents

Acknowledgments

We would like to thank all those who took the time to present their own use cases (INAF, CDS, CFA, ESAC) on which the model has been built.

We would also like to thank all the people having tested Mango on their own data.

We would like to thank all those who took the time to present their own use cases (INAF, CDS, CFA, ESAC) on which the model has been built. We would also like to thank all the people having tested Mango on their own data.

Model Name

This model was initialy named with a very explicit but hard to remember acronym, `CAB-MSD` standing for Component and Association Based Model for Source Data. We decided later to rename it `Mango` with reference to the inside out mango picture used to introduce the model in Groningen. As the tradition requires that such unexpected names are acronyms, let's assume that `Mango` stands for Model Adding Necessaries to Generic Objects

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 (Bradner, 1997).

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 source DM is a long term concern for the DM working group and more generally for the IVOA. In the past years, there were some proposals to design a global model for sources (Salgado and Lemson et al., 2016) of for catalogs (Osuna et al., 2006). Other proposals, more model-agnostic, were focused on the data annotation in VOTables (Demleitner and Ochsenbein et al., 2016) (Derriere, 2016). In this case the goal was no longer to design a source model but to provide a complete description of individual quantities (positions, velocity...). None of these proposals have come to completion

Figure 1: Architecture diagram for this document

The source DM issue resurfaced at the spring 2018 Interop in Victoria during an hands-on session focused on the tools available to work with VO data models and especially with VO-DML. The goal of this session was to annotate data from different origins in order to make them interoperable with each other. The main concern expressed during this session was not related to the tools themselves but to the lack of models for sources. This is a big paradox in the VO world ; source data which represent the basic bricks of the astronomer work, have no model. This paradox can be explained by the fact that sources data are multifaceted. The way of which source data are organized depends on the survey they come from, one the way they have been generated and on the expected use. In a more general way, it depends on the science we want to do with them. This diversity cannot be endorsed by a single model. Having a global source model would lead to a very complex solution not usable in practice.

This standard proposes to overcome this paradox presenting model based on independent components and associated data that can be embedded on demand in a container.

Not designed to describe what a source is nut to help clients to discover and to understand the quantities available for a particular source instance TBC

1.1 Role within the VO Architecture

Fig. 1 shows the role this document plays within the IVOA architecture (Arviset and Gaudet et al., 2010).

2 Use Cases and Requirements

2.1 Use Cases

The use case below have been collected in 2019 they were collected from representative of the missions, archive or tools.

2.1.1 Gaia

- identifier
- reference position
- proper motion
- parallax and distance
- correlation,
- source extension
- radial velocity
- redshift
- luminosity
- date
- multiple detection

2.1.2 Euclid

- identifier
- position
- correlation with Gaia
- photometry (ground + sat),
- morphology
- redshift
- photometric redshift

2.1.3 Exoplanets

- position

- orbit

- different source level(star, planet, moon)

- status and classification

- orbiting system description

2.1.4 Morphologically Complex Structures

- morphology

2.1.5 Chandra

All quantities are time dependant, Dependant on calibration + physical model

- name

- pos

- time

- extension

- PHA

2.1.6 Vizier

All possible measure are present in Vizier in addition to lots of associated data

- pre-existing data

- grouping columns

- lots of available metadata

- column name formatting

- one column different frames

2.1.7 Aladin

Quantities possibly plotted or matched with some of other datasets

- position

- time

- flux

- link

- FoV

- column grouping

2.1.8 Xmatch

- Identifier

- position

- proper motion

- photometry

2.1.9 Time Domain

- Identifier: object identifier

- Position: moving object

- Associated products: time series of e.g. a spectrum or an image

- Photometry: for light curves

- Timestamp: Independent axis of the time series

3 Model Overview

The root class of the model is the `Source` class that has only one attribute, the identifier, that is the only quantity common to all Mango instances.

Each quantity or associated data attached to a source is represented by a connector describing the nature of either the quantity or of the associated data.

There is no restriction on the set of connectors attached to one source. There are 2 classes of connectors

Quantity	Ga	Eu	Ex	MCS	Ch	Vi	Al	Xm	TD
identifier	p	P	P		P	P	P		P
position	p	P	P		P	P	P		P
pr. motion	p						P	P	
distance	p					P			
correlation	AD	AD					AD		
extension	P		P	P	P	P	P		
rad. vel.	P					P			
redshift	P	P				P			
phot. rsft		P				P			
luminosity	P	P				P	P	P	P
date	P				P	P			P
detections	AD								
orbit			P						
type			P			P			
status						P			
orb. sys.									
PHA					P				
ass. prd.							AD		AD

Table 1: `P` parameter, `AD` Associated data

Quantity	Model	
identifier	Mango	required for anu instance
position	Meas	
pr. motion	Meas	
distance	Ext Meas	
correlation	Ass data	reference to other Mango instances
extension	Ext Meas	Can be used for FoV, morphlogy or shape
rad. vel.	Ext Meas	
redshift	Ext Meas	
phot. rsft	Ext Meas	
luminosity	Ext Meas	
date	Meas	date or time stamp
detections	Ass data	reference to other Mango instances
orbit	Ext Meas	
type	Ext Meas	
status	Ext Meas	
orb. sys.	Ext Meas	
PHA	Ext Meas	
ass. prd.	Ass data	

Table 2: `Meas` Measue class, `Ext Meas` Measure extension, `Ass data` Mangno associated data

3.1 The parameters

Parameter connectors are used to bind measures with the source. One connector can only reference one measure. The measure itself is modeled by a class of the Measure model or by an extension of this model. Measure extensions are part of Mango. In addition to its class, the nature of a measure is given by a UCD and a semantic tag. The UCD gives the physical meaning of the measure whereas the semantic tag says more about the usage context of the measure. This can be for instance the reduction status (raw, corrected). The vocabulary allowed for the semantic is managed as a VO vocabulary.

3.1.1 Parameter UCD

TBC phys.luminosity vs phot.flux ajoutdr une temperature
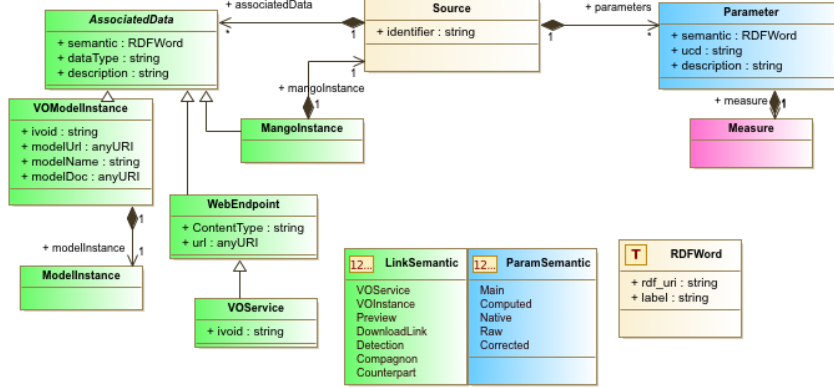
Figure 2: Mango overview

Parameter	Original model	UCDs 1+ first word
Position	Measure	pos
Velocity	Measure	phys.veloc
Proper motion	Measure	pos.pm
Time	Measure	time.epoch
Polarization	Measure	phys.polarization
SphericalSkyPosition	Mango	pos.eq
ObjectType	Mango	src.class
Redshift	Mango	src.redshift
Luminosity	Mango	phot.qqqchose
HardnessRatio	Mango	phot.flux;arith.ratio
Shape	Mango	phys.area
Flag	Mango	meta.code
Orbit	Mango	src.orbital
GenericMeasure	Measure	Appropiate physical UCD

Table 3: UCDs to be set for the supported parameters

3.2 Associated Data

Associated data connectors are used to bind any sort of complex data with the source. One connector can only refer to one dataset. Associated data can be either URIs (VO services or not) or reference to instances of other VO models (Obscore, Provenance...).

Attaching VO model instances to sources does not mean that these models must be imported. The merge between a Mango instance and one associated instance is operated by the data annotation process. The indication given by Mango tell the client how to interpret the embedded instance

11
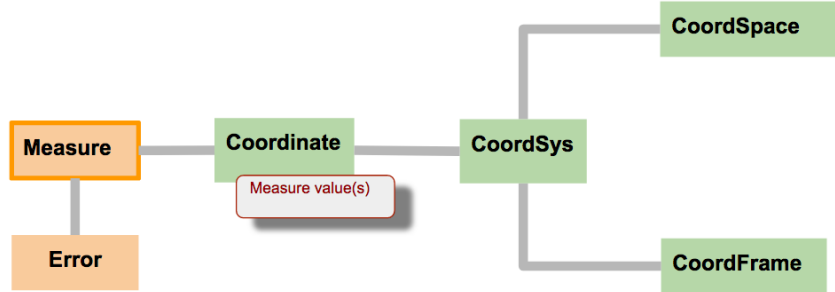
Figure 3: STC pattern (simplified view)

3.3 STC Extension

All STC measures are built upon the same pattern (see fig 3). The value(s) `Measure` object is carried by a `Coordinate` object. The coordinate is attached to a system (class `CoordSys`). The system has 2 components 1) the space (class `CoordSpace`) that describes the axis and 2) the frame `CoordFrame`). All Mango parameters are based on this pattern. Native STC classes are used when possible. Others parameters are built by extending STC classes . Figure 3) shows the extended STC classes being part of Mango

Figure 4: STC extension diagram

4 Model: mango

Data model based oon components and data association for source data

4.1 AssociatedData (Abstract)

Abstract reference to a particular dataset associated to the Source. This class is used to specify the type of the dataset as well as its role.

4.1.1 AssociatedData.semantic

vodml-id: AssociatedData.semantic
type: mango:stcextend.RDFWord
multiplicity: 1
Reference to a semantic concept giving the nature of the associated data. As long as the vocabulary is not set, the possible values of this attribute are given by the LinkSemantic enumeration.

4.1.2 AssociatedData.dataType

vodml-id: AssociatedData.dataType
type: ivoa:string
multiplicity: 1
Type of the associated data (not defined yet)

4.2 HardnessRatioSys

TODO : Missing description : please, update your UML model asap.
subset
    role: coords:PhysicalCoordSys.frame
    type: mango:HRFrame

4.3 MangoInstance

Reference to another CAB-MSD instance that is part of the associated data.

4.3.1 MangoInstance.mangoInstance

vodml-id: MangoInstance.mangoInstance
type: mango:Source
multiplicity: 1
Composition link pointing on one cab_msd instance associated with the source.

4.4 ModelInstance

Placeholder for the mapping of the model instance

4.5 Parameter

Reference to a particular measure of the Source. This class is used to specify the type of the measure as well as its role.

constraint

    detail: Parameter.One association at the time

4.5.1 Parameter.semantic

vodml-id: Parameter.semantic
type: mango:stcextend.RDFWord
multiplicity: 1

Reference to a semantic concept giving the nature of the parameter As long as the vocabulary is not set, the possible values of this attribute are given by the ParamSemantic enumeration.

4.5.2 Parameter.ucd

vodml-id: Parameter.ucd
type: ivoa:string
multiplicity: 1

UCD1+ giving the type of the physical measure

4.5.3 Parameter.measure

vodml-id: Parameter.measure
type: meas:Measure
multiplicity: 1

Composition link pointing to the meas:Measure instance

4.6 PhotometryCoordSys

TODO : Missing description : please, update your UML model asap.

subset

    role: coords:PhysicalCoordSys.frame
    type: mango:PhotFilter

4.7 Source

Root class of the model. CAB_MSF instance are meant ot be Source instances. A source ha an identifier and to sets of hooks: one for the parameters and one for the associated data.

4.7.1 Source.identifier

vodml-id: Source.identifier
type: ivoa:string
multiplicity: 1
Unique identifier for a Source. The uniqness of that identifier is not managed by the model. The format is free.

4.7.2 Source.associatedData

vodml-id: Source.associatedData
type: mango:AssociatedData
multiplicity: 0..*
Composition link pointing on all data associated with the source.

4.7.3 Source.parameters

vodml-id: Source.parameters
type: mango:Parameter
multiplicity: 0..*
Composition link pointing on all parameters attached to the source.

4.8 Temperature

TODO : Missing description : please, update your UML model asap.

4.8.1 Temperature.coord

vodml-id: Temperature.coord
type: ivoa:string
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

4.9 VOModelInstance

Reference to a VO model instance that is part of the associated data.

4.9.1 VOModelInstance.ivoid

vodml-id: VOModelInstance.ivoid
type: ivoa:string
multiplicity: 1
VO-DML id of the referenced model

4.9.2 VOModelInstance.modelUrl

vodml-id: VOModelInstance.modelUrl
type: ivoa:anyURI
multiplicity: 1
URL on the VO-DML model

4.9.3 VOModelInstance.modelName

vodml-id: VOModelInstance.modelName
type: ivoa:string
multiplicity: 1
Name of the referenced model

4.9.4 VOModelInstance.modelDoc

vodml-id: VOModelInstance.modelDoc
type: ivoa:anyURI
multiplicity: 1
Documentation URL of the model

4.9.5 VOModelInstance.modelInstance

vodml-id: VOModelInstance.modelInstance
type: mango:ModelInstance
multiplicity: 1
Composition link pointing on one VO instance instance associated with the source.

4.10 VOService

Class for associated data referenced by an URL that is a VO service

4.10.1 VOService.ivoid

vodml-id: VOService.ivoid
type: ivoa:string
multiplicity: 1
IVOA id attached to the URI

4.11   WebEndpoint

Class for associated data referenced by an URL

4.11.1   WebEndpoint.ContentType

vodml-id: WebEndpoint.ContentType
type: ivoa:string
multiplicity: 1
Mime type of the URL

4.11.2   WebEndpoint.url

vodml-id: WebEndpoint.url
type: ivoa:anyURI
multiplicity: 1
Web endpoint

4.12   LinkSemantic

Literal enumeration of the possible values for the associated data semantic.
This stands for an example before we have defined a vocabulary.
<u>Enumeration Literals</u>

VOService : **vodml-id:** LinkSemantic.VOService
  description: Data returned by a VO service

VOInstance : **vodml-id:** LinkSemantic.VOInstance
  description: Data Serialized in a VO model

Preview : **vodml-id:** LinkSemantic.Preview
  description: data preview

DownloadLink : **vodml-id:** LinkSemantic.DownloadLink
  description: Data download link

Detection : **vodml-id:** LinkSemantic.Detection
  description: Particular detection

Compagnon : **vodml-id:** LinkSemantic.Compagnon
  description: Compagnon source

Counterpart : **vodml-id:** LinkSemantic.Counterpart
  description: Counter part source

4.13   ParamSemantic

Literal enumeration of the possible values for the parameter semantic. This
stands for an example before we have defined a vocabulary.
<u>Enumeration Literals</u>

Main : **vodml-id:** ParamSemantic.Main
   description: Main measurment

Computed : **vodml-id:** ParamSemantic.Computed
   description: Computed measurement

Native : **vodml-id:** ParamSemantic.Native
   description: Mative measurement

Raw : **vodml-id:** ParamSemantic.Raw
   description: raw measure

Corrected : **vodml-id:** ParamSemantic.Corrected
   description: Corrected measure

5 Package: stcextend

This package contains all object and type classes that has been extended from the Measure and Coordinates models. This extension mechanism is used to add new types of measures while staying whithin the Mes/Coords pattern.

5.1 Flag

Measure to be used for status parameters

5.1.1 Flag.coord

vodml-id: stcextend.Flag.coord
type: mango:stcextend.FlagCoord
multiplicity: 1
Coordinate holding the statsu value

5.2 FlagState

Possible value of a status

5.2.1 FlagState.value

vodml-id: stcextend.FlagState.value
type: ivoa:integer
multiplicity: 1
Status value

5.2.2 FlagState.label

vodml-id: stcextend.FlagState.label
type: ivoa:string
multiplicity: 1
Label attached to that status value

5.3 FlagSys

Coordinate system to be used for statur measures.

5.3.1 FlagSys.statusLabel

vodml-id: stcextend.FlagSys.statusLabel
type: mango:stcextend.FlagState
multiplicity: 0..*
Composition loink to all possible status values for this system

5.4 HRFrame

Hardness ratio frame. Defined by 2 energy bands Eheigh ELow. HR = (Eheigh - Elow)/(Eheigh + Elow) Energy bands are deemed to special photometric filters

5.4.1 HRFrame.low

vodml-id: stcextend.HRFrame.low
type: mango:stcextend.PhotFilter
multiplicity: 1
Low energy band

5.4.2 HRFrame.high

vodml-id: stcextend.HRFrame.high
type: mango:stcextend.PhotFilter
multiplicity: 1
Heigh energy band

5.5 HardnessRatio

TODO : Missing description : please, update your UML model asap.

5.5.1 HardnessRatio.coord

vodml-id: stcextend.HardnessRatio.coord
type: mango:stcextend.HardnessRatioCoord
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

5.6 LonLatCoordSys

TODO : Missing description : please, update your UML model asap.
subset
　　role: coords:PhysicalCoordSys.frame
　　type: coords:SpaceFrame

constraint
    detail: LonLatCoordSys.coordSpace[0]

5.7 LonLatSkyPosition

Measure to used for sky points expressed with a spherical coordinate system

5.7.1 LonLatSkyPosition.coord

vodml-id: stcextend.LonLatSkyPosition.coord
type: mango:stcextend.LonLatPoint
multiplicity: 1
Coordinate of spherical sky position

5.8 ObjectType

TODO : Missing description : please, update your UML model asap.

5.8.1 ObjectType.coord

vodml-id: stcextend.ObjectType.coord
type: mango:stcextend.OrbitCoord
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

5.9 ObjectTypeSys

TODO : Missing description : please, update your UML model asap.

5.10 Orbit

TODO : Missing description : please, update your UML model asap.

5.10.1 Orbit.coord

vodml-id: stcextend.Orbit.coord
type: mango:stcextend.OrbitCoord
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

5.11 PhotFilter

Photometric filter description, compliant with photDM

5.11.1 PhotFilter.name

vodml-id: stcextend.PhotFilter.name
type: ivoa:string
multiplicity: 1
Filter name

5.11.2 PhotFilter.zeroPointFlux

vodml-id: stcextend.PhotFilter.zeroPointFlux
type: ivoa:real
multiplicity: 1
Zero point flux of the filter

5.11.3 PhotFilter.magnitudeSystem

vodml-id: stcextend.PhotFilter.magnitudeSystem
type: ivoa:string
multiplicity: 1
Magnitude system used by the filter

5.11.4 PhotFilter.effectiveWavelength

vodml-id: stcextend.PhotFilter.effectiveWavelength
type: ivoa:real
multiplicity: 1
Effective wavelength of the filter

5.11.5 PhotFilter.unit

vodml-id: stcextend.PhotFilter.unit
type: ivoa:Unit
multiplicity: 1
Wavelength unit used for that filter

5.11.6 PhotFilter.bandWidth

vodml-id: stcextend.PhotFilter.bandWidth
type: ivoa:real
multiplicity: 1
Band width of the filter

5.12 Photometry

TODO : Missing description : please, update your UML model asap.

5.12.1 Photometry.coord

vodml-id: stcextend.Photometry.coord
type: mango:stcextend.PhotometryCoord
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

5.13 Redshift

TODO : Missing description : please, update your UML model asap.

5.13.1 Redshift.coord

vodml-id: stcextend.Redshift.coord
type: mango:stcextend.RedshiftCoord
multiplicity: 1
TODO : Missing description : please, update your UML model asap.

5.14 RedshiftSys

TODO : Missing description : please, update your UML model asap.

5.15 Shape

Measure giving the shape of a source

5.15.1 Shape.coord

vodml-id: stcextend.Shape.coord
type: mango:stcextend.ShapeCoord
multiplicity: 1
String serialization of the source shape

5.16 ShapeSys

Coordinate systen to be used for shape measure

5.16.1 ShapeSys.shapeFrame

vodml-id: stcextend.ShapeSys.shapeFrame
type: mango:stcextend.ShapeFrame
multiplicity: 1
Frame of the shape measure. Gives a enumeration of the supported serializations.

5.17 LonLatPoint

Coordinate of a point on the sky sphere expressed in spherical coordinates.

5.18 FlagCoord

Coordinate of a status Measure

5.19 ShapeCoord

TODO : Missing description : please, update your UML model asap.

5.20 OrbitCoord

TODO : Missing description : please, update your UML model asap.

5.21 ObjectTypeCoord

TODO : Missing description : please, update your UML model asap.

5.22 PhotometryCoord

TODO : Missing description : please, update your UML model asap.

5.23 HardnessRatioCoord

TODO : Missing description : please, update your UML model asap.

5.24 RedshiftCoord

TODO : Missing description : please, update your UML model asap.

5.25 RDFWord

Datatype for vocabulary word. Provides a pointer to the word description and a label.

5.26 ShapeFrame

Enumeration of the possible options to encode a shape in a string.
Enumeration Literals

MOC : **vodml-id:** stcextend.ShapeFrame.MOC
    description: MOC serialization

STCs : **vodml-id:** stcextend.ShapeFrame.STCs
    description: STCs serialization

6 TAP and Mango

This not normative section gives possible tips to save and discover Mango instances in TAP services. We suppose that the TAP service hosts catalogs which sources are Mango instances. These catalogs are named *Mango Catalogs.*

6.1 Storing Mango Catalogs in TAP

For now this section only concerns the parameter. The associated data will be taken into account later.

- One master table for the catalogs with various meta-data out of the Mango scope plus a unique identifier (primary key)

- One master sources table for the source instances with the catalog identifier and a primary key safer than the Mango identifier.

- One table for each supported parameter with a foreign key for the join with the master source table

Although the model of the measures is hierarchical, it should be possible to flatten them in one single table considering that the model structure can be retrieved with the TAP_SCHEMA annotations (TBC)

This schema requires the server to explore all the parameter tables to retrieve whole Mango instances. This process can be speed up by using the *MangoCore* table.

6.2 *MangoCore* Table

The discovery of *Mango Catalogs* can be helped by a *MangoCore* table located in the *schema* schema. As Mango is not dedicated to any specific domain, we cannot define a set of core parameters, but parameters can be flagged as *Core Parameter*. This selection is left at the discretion of the curator. The *MangoCore* table has set of columns per parameter class plus one for the catalog ID. It has one row per stored catalog. Each parameter has at least 2 columns: one with the UCD an one with the *Core* flag. TBC

A Changes from Previous Versions

No previous versions yet.

References

Arviset, C., Gaudet, S. and the IVOA Technical Coordination Group (2010), 'IVOA architecture', IVOA Note.
http://www.ivoa.net/documents/Notes/IVOAArchitecture

Bradner, S. (1997), 'Key words for use in RFCs to indicate requirement levels', RFC 2119.
http://www.ietf.org/rfc/rfc2119.txt

Demleitner, M., Ochsenbein, F., McDowell, J. and Rots, A. (2016), 'Referencing stc in votable', Note 2010-06-18.
http://ivoa.net/documents/Notes/VOTableSTC/

Derriere, S. (2016), 'Referencing stc in votable', Note 2011-05-12.
https://wiki.ivoa.net/internal/IVOA/PhotometryDataModel/
NOTE-PPDMDesc-0.2-20110512.pdf

Osuna, P. et al. (2006), 'Catalog data model', Catalog Data Model Work Package.
https://wiki.ivoa.net/twiki/bin/view/IVOA/IVAODMCatalogsWP

Salgado, J., Lemson, G. and Demleitner, M. (2016), 'Source dm', Source DM 2016-05-10.
https://wiki.ivoa.net/twiki/bin/view/IVOA/SourceDataModel