Data modeling in the Virtual Observatory Framework

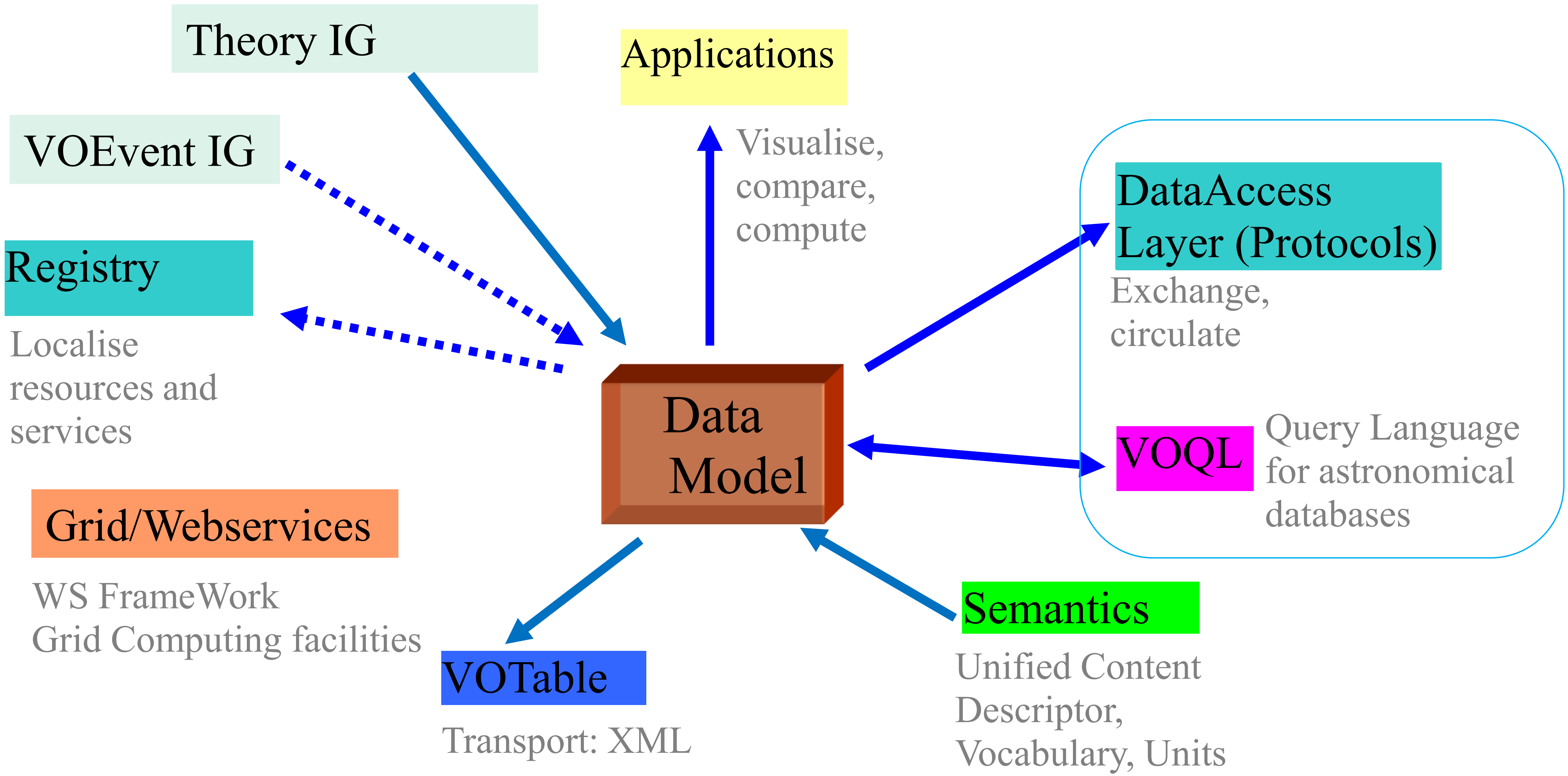
Mireille Louys
CDS, Strasbourg Observatory
ICube Laboratory, Strasbourg University, F

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

A collaborative effort thanks to

Jesus Salgado, Carlos Rodrigo, Pedro Osuna, Mark Allen, Mireille Louys, Jonathan McDowell, Deborah Baines, Jesus Maiz Apellaniz, Evanthia Hatziminaoglou, Sebastien Derriere, Gerard Lemson, Laurent Bourges, Miguel Cervino, Claudio Gheller, Norman Gray, Franck LePetit, Benjamin Ooghe, Rick Wagner, Herve Wozniak, Igor Chilingarian, Marie-Lise Dubernet, Evelyne Roueff, Matteo Guainazzi, Enrique Solano, Joe Mazzarella, Raffaele D'Abrusco, Tamas Budavari, Markus Dolensky, Inga Kamp, Kelly McCusker, Doug Tody, Pavlos Protopapas, Arnold Rots, Randy Thompson, Frank Valdes, Petr Skoda, Bruno Rino, Jim Cant, Omar Laurino, Patrick Dowler, Daniel Durand, Laurent Michel, François Bonnarel, Mark Cresitello-Dittmar, Markus Demleitner, Tom Donaldson, Matthew Graham, Jose Enrique Ruiz, Kristin Riebe, Michèle Sanguillon, Mathieu Servillat, Markus Nullmeier, Catherine Boisson, Claudia Lavalley

IVOA Working Groups / Interactions



Metadata modeling goals in the VO context

Describe metadata for all datasets exchanged in the astronomical community in an homogeneous way.

Sustains the **interoperability** objective of the Virtual Observatory

Based on the requirements:

- **For users :**

  To be able to ask the same question/query to various astronomical data bases, select the results and then retrieve data.

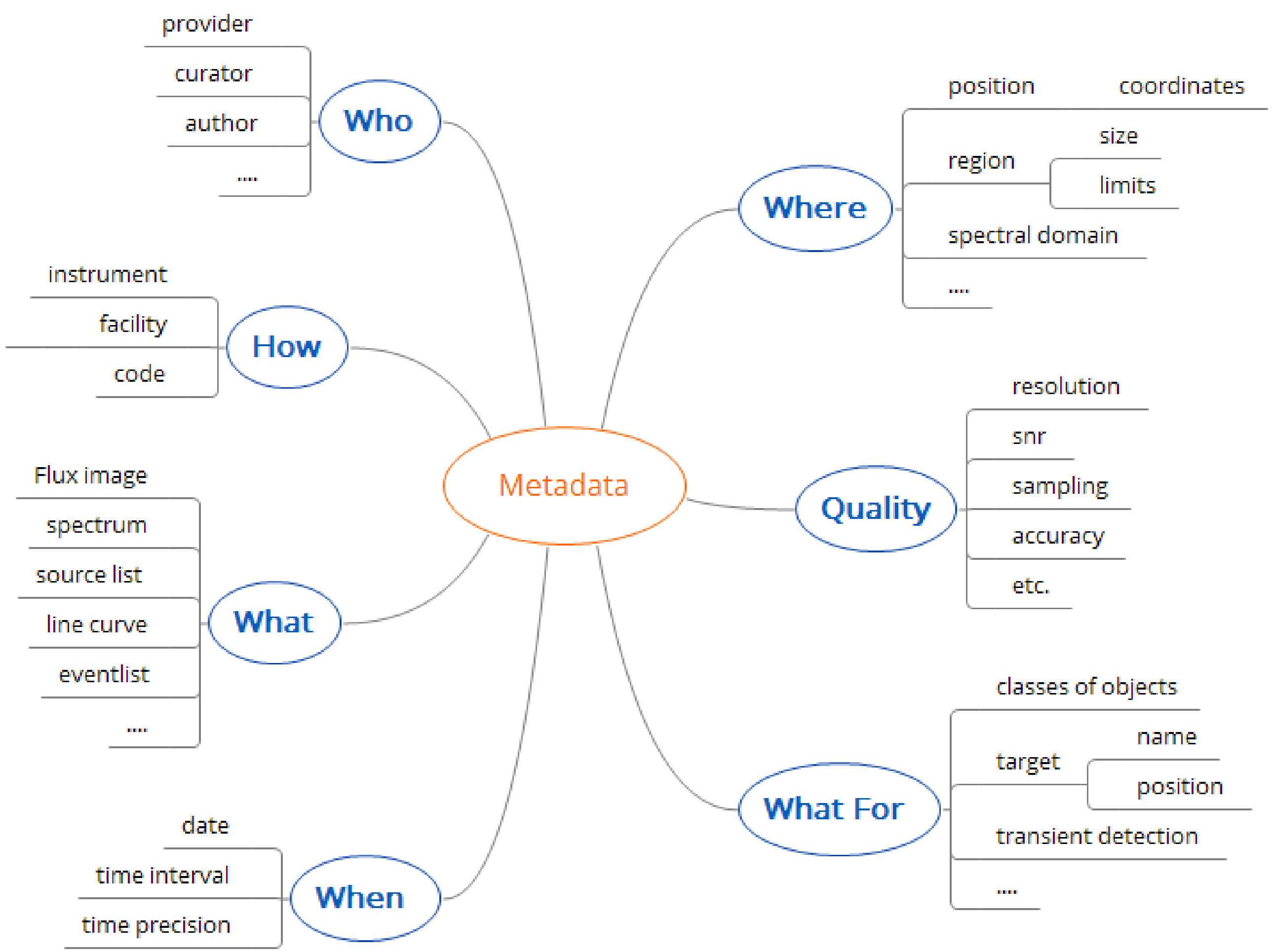
  Search seamlessly for observational data at all wavelengths and for all instruments.

- **For data providers :**

  Expose and distribute more data with a standardized metadata description.

  In core in the DB structure or as an interoperability layer on existing architecture (DB view).

What is metadata in our context ?



- Organize
- Describe at the appropriate level of details
- Each use-case sets its own quality requirements and criteria
- Describe all properties on each physical axis of the data: spatial, temporal, spectral, polarization, etc.

Modeling Strategy

Following the principle of Object Oriented Design :

- Describe the responsibilities and properties of the metadata involved for data produced by an observation or a simulation process.
- A dialog between data providers and users who came up with :
  - **Root concepts** qualifying astronomical data
  - A **vocabulary** (list of terms) based on existing practises (FITS, bibliographic services, archives contents, interviews from astronomers and data providers ..)
  - **logical structure** showing the dependencies and relations between all pieces of metadata

→ A general schema for metadata representation built from commonalities and specificities used at various data providers archives

General building blocks models

Metadata features	Data model name&version	Year	Status
Astronomical Space Time Coordinates	STC v 1.33	2007	REC
	STC v 2.0	2016	WD
Physical axis description and properties	Characterization v1.13	2008	REC

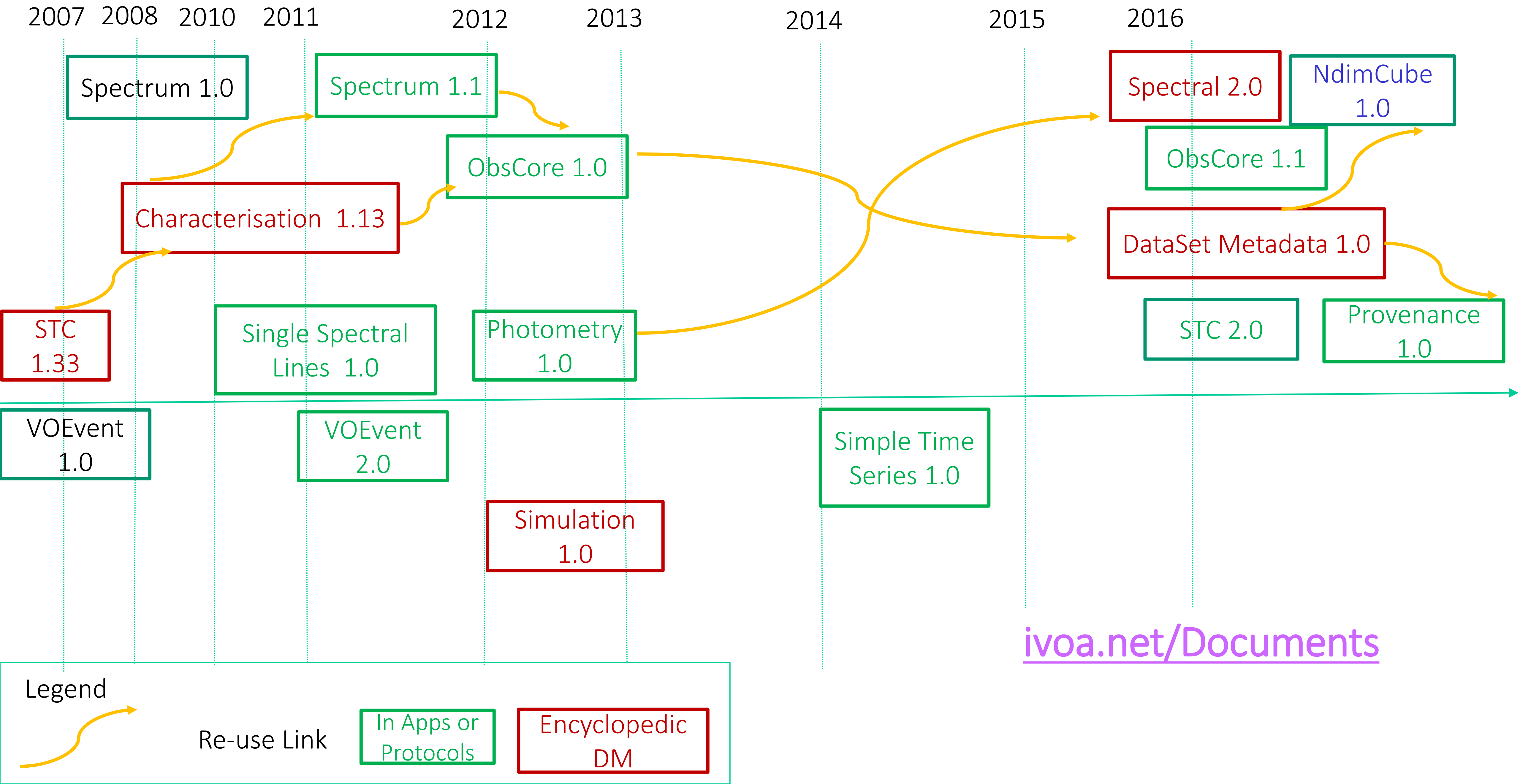
REC : IVOA Recommendation   PR: Proposed Recommendation   WD: Working Draft

Observational Metadata

Metadata features	Data model name &version	Year	Status	Protocol	Applications
Spectral Line Transitions	Simple Spectral Line	2010	REC	SLAP	VOSpec, SPLAT-VO
1D Spectrum, Light Curves	Spectrum v1.0	2007	REC	SSA 1.0	
1D Spectrum, Light Curves	Spectrum v1.1	2011	REC	SSA 1.1	SPLAT-VO, IRIS
SED, Photometric Points, Time series, Multi-segment 1D spectrum	Spectral v2.0	2014	PR	SSA 1.1	SPLAT
Observational dataset Core Components	ObsCore v1.0	2011	REC	TAP 1.0, SIAv2	TAPHandle, TOPCAT
(All data products for global discovery)	ObsCore v1.1	2016	PR	TAP 1.0	TAPHandle, TOPCAT
Photometric calibration	Photometry v1.0	2013	REC		SPLAT-VO, CDS Sed browser, SVO Filter Profile Service
All observation datasets (Fine grain description)	DataSet Metadata v1.0	2015	WD		
N-D cubes; pixelated images, sparse data	NDCube v1.0	2015	WD		SPLAT-VO/time

Other Metadata

Metadata features	Data model name &version	Year	Status	Protocol	Applications
VOEvent: transients observations	VOEvent v2.0		REC	VTP 1.0	Service embedded desc .
Simple time series	Simple time series v1.0	2014	Note		
Simulation Data					
Simulations, data and code description	Simulation v1.0	2012	REC	SimDAL	DEUVO, Meudon PDR code, MilleniumDB, etc.
Micro simulations, Implementations of SimDM	Implementations of the Simulation DM v1.0	2012	Note		
Provenance metadata					
Datasets generation process, Progenitors	Provenance DM v1.0	2016	WD	TBD	

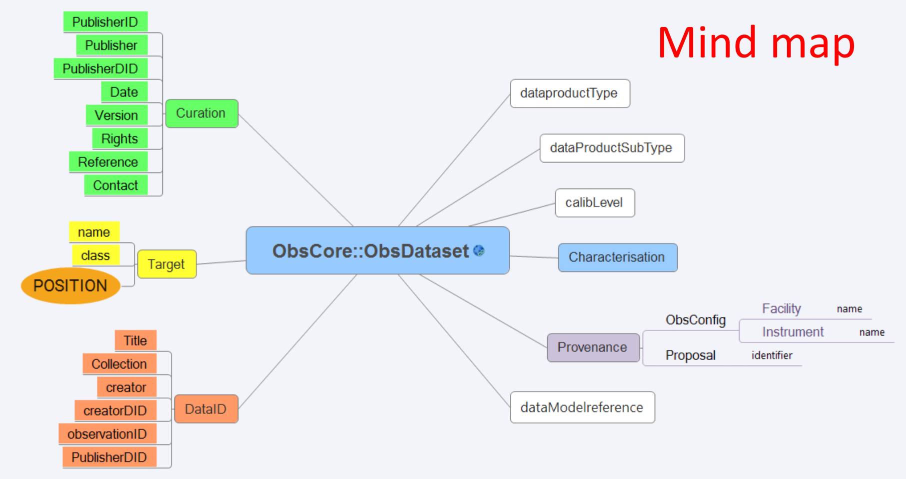

2007
2008
2010
2011
2012
2013
2014
2015
2016
Spectrum 1.0
Spectrum 1.1
ObsCore 1.0
Characterisation 1.13
STC 1.33
Single Spectral Lines 1.0
Photometry 1.0
VOEvent 1.0
VOEvent 2.0
Simulation 1.0
Simple Time Series 1.0
Spectral 2.0
NdimCube 1.0
ObsCore 1.1
DataSet Metadata 1.0
STC 2.0
Provenance 1.0
ivoa.net/Documents
Legend
Re-use Link
In Apps or Protocols
Encyclopedic DM

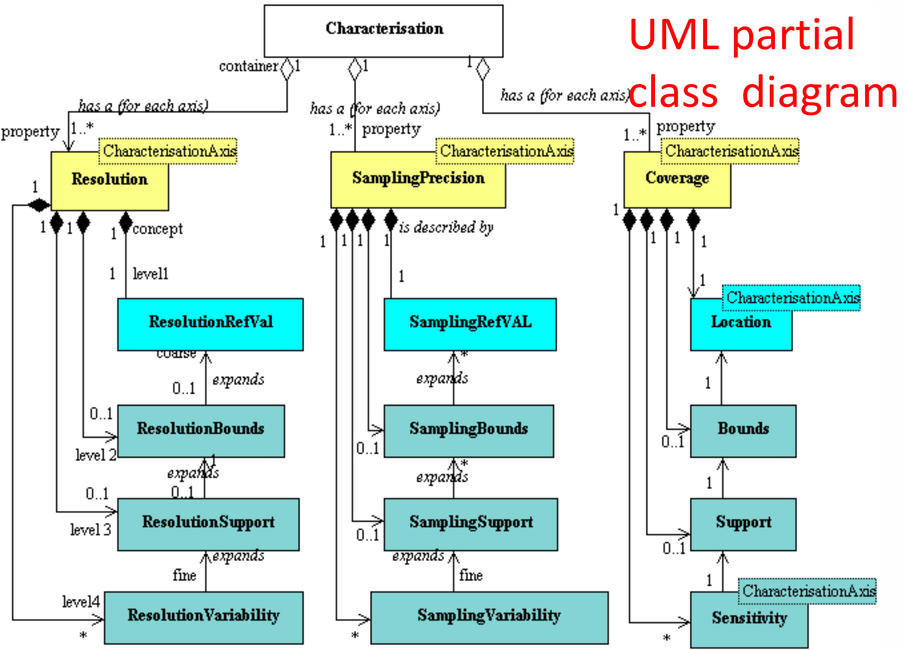
Structural aspect

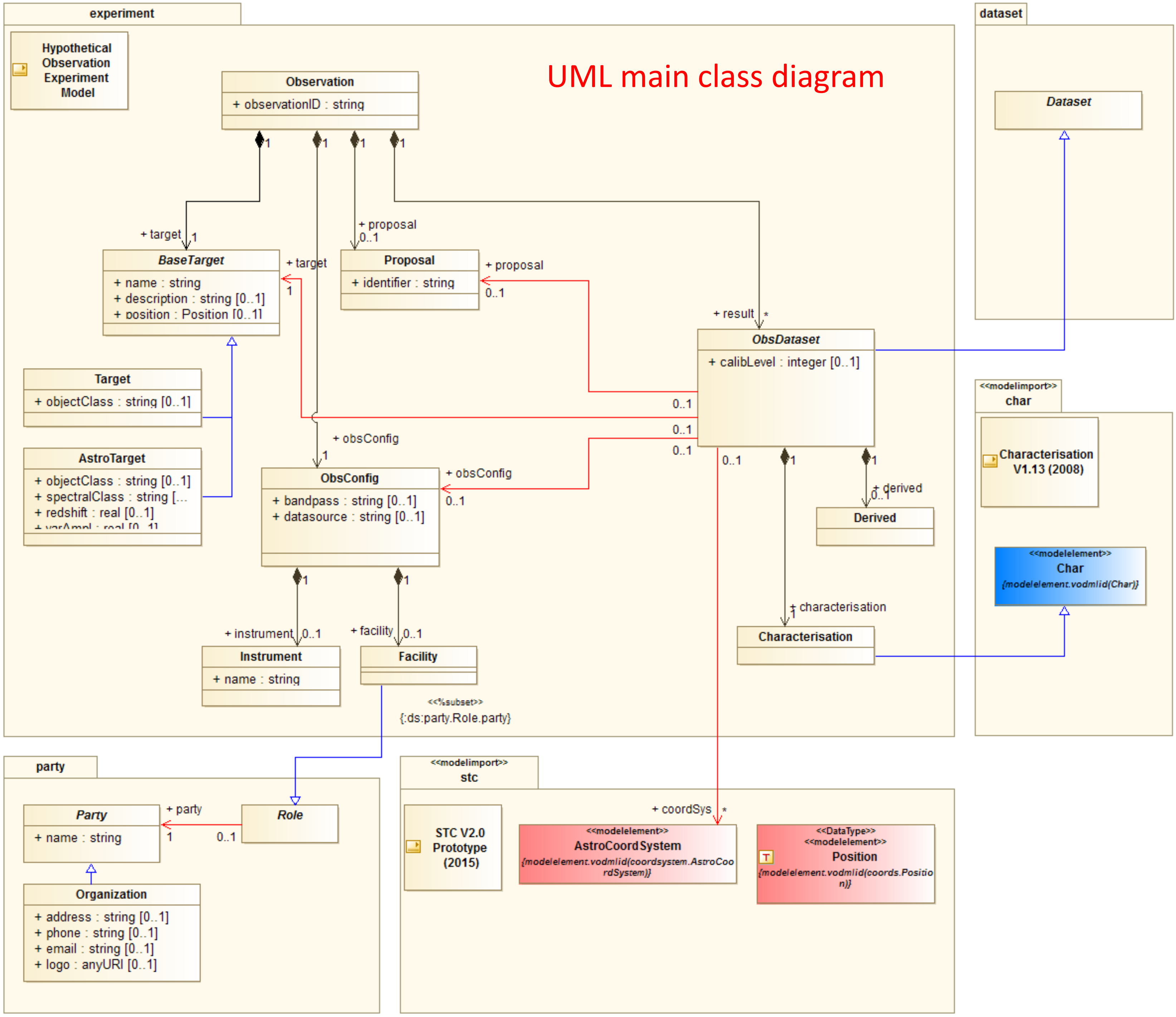
- The DM fields arrangement contains the dependencies between different pieces of metadata
- Can be expressed as
  - UML class diagram
  - Mind maps
  - Tables
  - Lists

Suitable the most for interpretation by humans (graphs) or by machine (tables, lists, XML documents)

Graphs


Mind map
ObsCore::ObsDataset
Curation
PublisherID
Publisher
PublisherDID
Date
Version
Rights
Reference
Contact
Target
name
class
POSITION
DataID
Title
Collection
creator
creatorDID
observationID
PublisherDID
dataproductType
dataProductSubType
calibLevel
Characterisation
Provenance
ObsConfig
Facility
Instrument
Proposal
name
identifier
dataModelreference


UML partial class diagram
Characterisation
container
has a (for each axis)
property
CharacterisationAxis
Resolution
SamplingPrecision
Coverage
concept
is described by
level1
level 2
level 3
level4
ResolutionRefVal
SamplingRefVAL
Location
ResolutionBounds
SamplingBounds
Bounds
ResolutionSupport
SamplingSupport
Support
ResolutionVariability
SamplingVariability
Sensitivity
expands
coarse
fine


UML main class diagram
experiment
Hypothetical Observation Experiment Model
Observation
+ observationID : string
BaseTarget
+ name : string
+ description : string [0..1]
+ position : Position [0..1]
Proposal
+ identifier : string
Target
+ objectClass : string [0..1]
AstroTarget
+ objectClass : string [0..1]
+ spectralClass : string [...
+ redshift : real [0..1]
ObsConfig
+ bandpass : string [0..1]
+ datasource : string [0..1]
ObsDataset
+ calibLevel : integer [0..1]
Derived
Characterisation
Instrument
+ name : string
Facility
<<%subset>>
{:ds:party.Role.party}
+ target
+ proposal
+ result
+ obsConfig
+ instrument
+ facility
+ derived
+ characterisation
dataset
Dataset
<<modelimport>>
char
Characterisation V1.13 (2008)
<<modelelement>>
Char
{modelelement.vodmlid(Char)}
party
Party
+ name : string
+ party
Role
Organization
+ address : string [0..1]
+ phone : string [0..1]
+ email : string [0..1]
+ logo : anyURI [0..1]
<<modelimport>>
stc
STC V2.0 Prototype (2015)
+ coordSys
<<modelelement>>
AstroCoordSystem
{modelelement.vodmlid(coordsystem.AstroCoordSystem)}
<<DataType>>
<<modelelement>>
Position
{modelelement.vodmlid(coords.Position)}

Data Model Table description (ObsTAP ex.)

nameattr	unit	ucd	SEMANTICS	utype	STRUCTURE	dataType	description
obs_publisher_did		meta.ref.url;meta.curation		obscore:Curation.PublisherDID		VARCHAR	publisher dataset identifier
obs_collection		meta.id		obscore:DataID.Collection		VARCHAR	short name for the data colection
facility_name		meta.id;instr.tel		obscore:Provenance.ObsConfig.Facility.name		VARCHAR	telescope name
instrument_name		meta.id;instr		obscore:Provenance.ObsConfig.Instrument.name		VARCHAR	instrument name
obs_id		meta.id		obscore:DataID.observationID		VARCHAR	internal dataset identifier
dataproduct_type		meta.id		obscore:ObsDataset.dataProductType		VARCHAR	type of product
calib_level		meta.code;obs.calib		obscore:ObsDataset.calibLevel		INTEGER	calibration level (0,1,2,3)
obs_release_date		time.release		obscore:Curation.releaseDate		TIMESTAMP	timestamp of date the data becomes publicly available
target_name		meta.id;src		obscore:Target.Name		VARCHAR	name of intended target
s_ra	deg	pos.eq.ra		obscore:Char.SpatialAxis.Coverage.Location.Coord.Position2D.Value2.C1		DOUBLE	RA of central coordinates
s_dec	deg	pos.eq.dec		obscore:Char.SpatialAxis.Coverage.Location.Coord.Position2D.Value2.C2		DOUBLE	DEC of central coordinates
s_fov	deg	phys.angSize;instr.fov		obscore:Char.SpatialAxis.Coverage.Bounds.Extent.diameter		DOUBLE	size of the region covered (~diameter of minimum bounding circle)
s_region	deg	phys.outline;obs.field		obscore:Char.SpatialAxis.Coverage.Support.Area		REGION	region bounded by observation
s_resolution	arcsec	pos.angResolution		obscore:Char.SpatialAxis.Resolution.refval.value		DOUBLE	typical spatial resolution
s_xel1		meta.number		obscore:Char.SpatialAxis.numBins1		BIGINT	dimensions (number of pixels) along one spatial axis
s_xel2		meta.number		obscore:Char.SpatialAxis.numBins2		BIGINT	dimensions (number of pixels) along the other spatial axis
t_min	d	time.start;obs.exposure		obscore:Char.TimeAxis.Coverage.Bounds.Limits.StartTime		DOUBLE	start time of observation (MJD)
t_max	d	time.end;obs.exposure		obscore:Char.TimeAxis.Coverage.Bounds.Limits.StopTime		DOUBLE	end time of observation (MJD)

VO-DML data model description

- Derived from UML modeling
- Expressed following the VO-DML meta-model
- Describes all data model classes, attributes and relations in a dedicated XML document
- On going specification in proposed recommendation phase

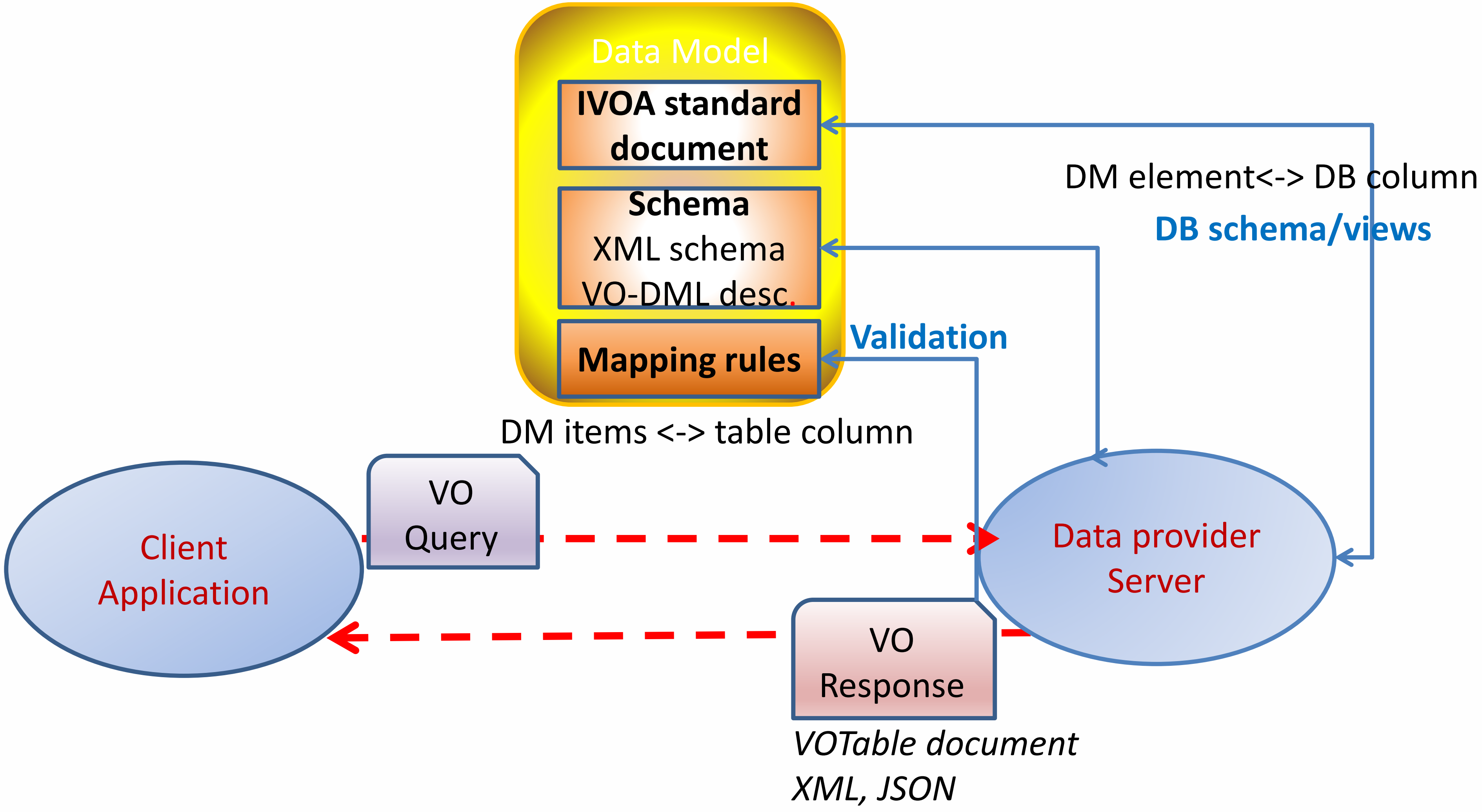
International
Virtual
Observatory
Alliance

VO-DML: a consistent modeling language for IVOA data models

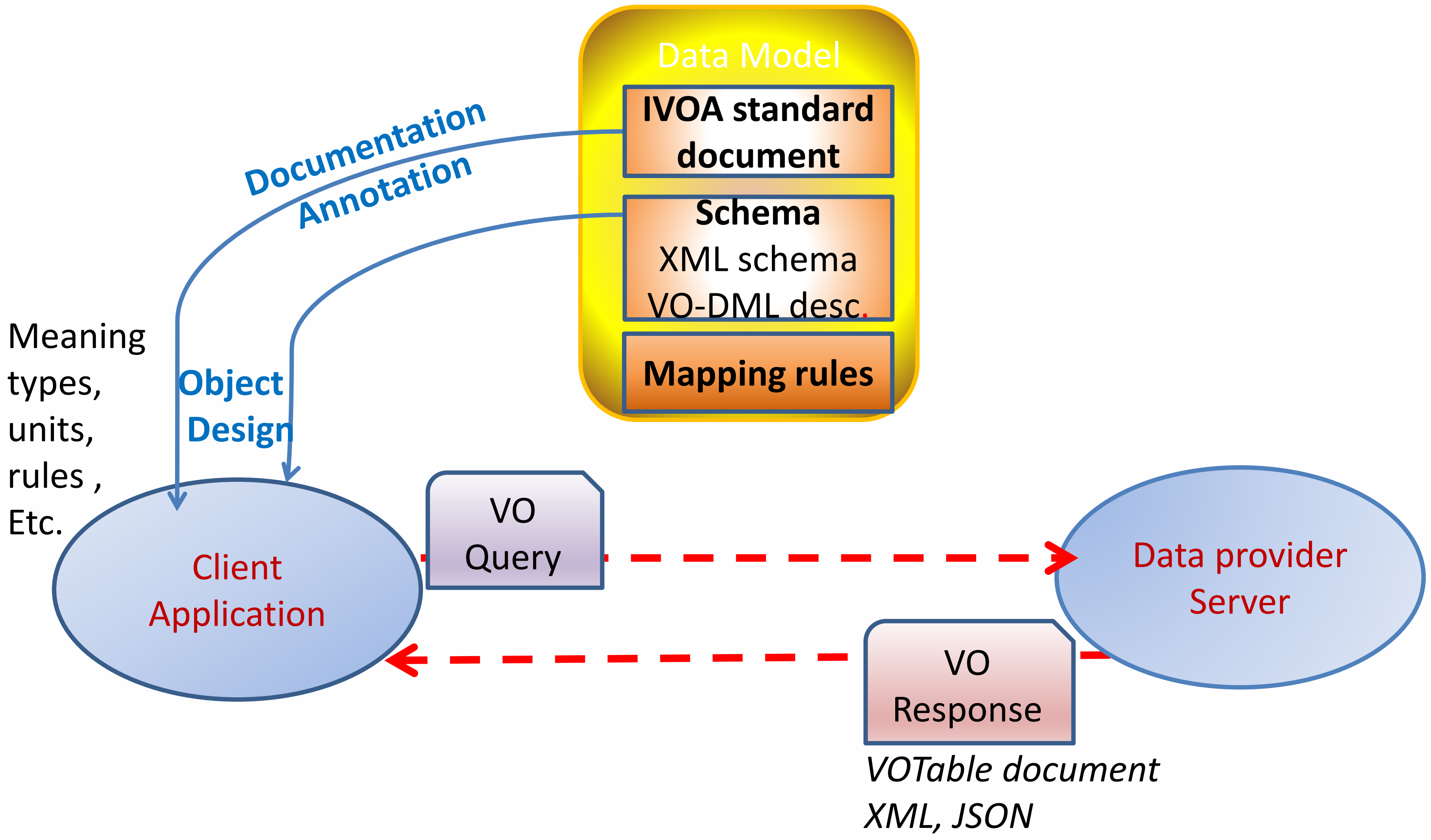
Version 1.0-20160923

Proposed Recommendation 2016 September 23

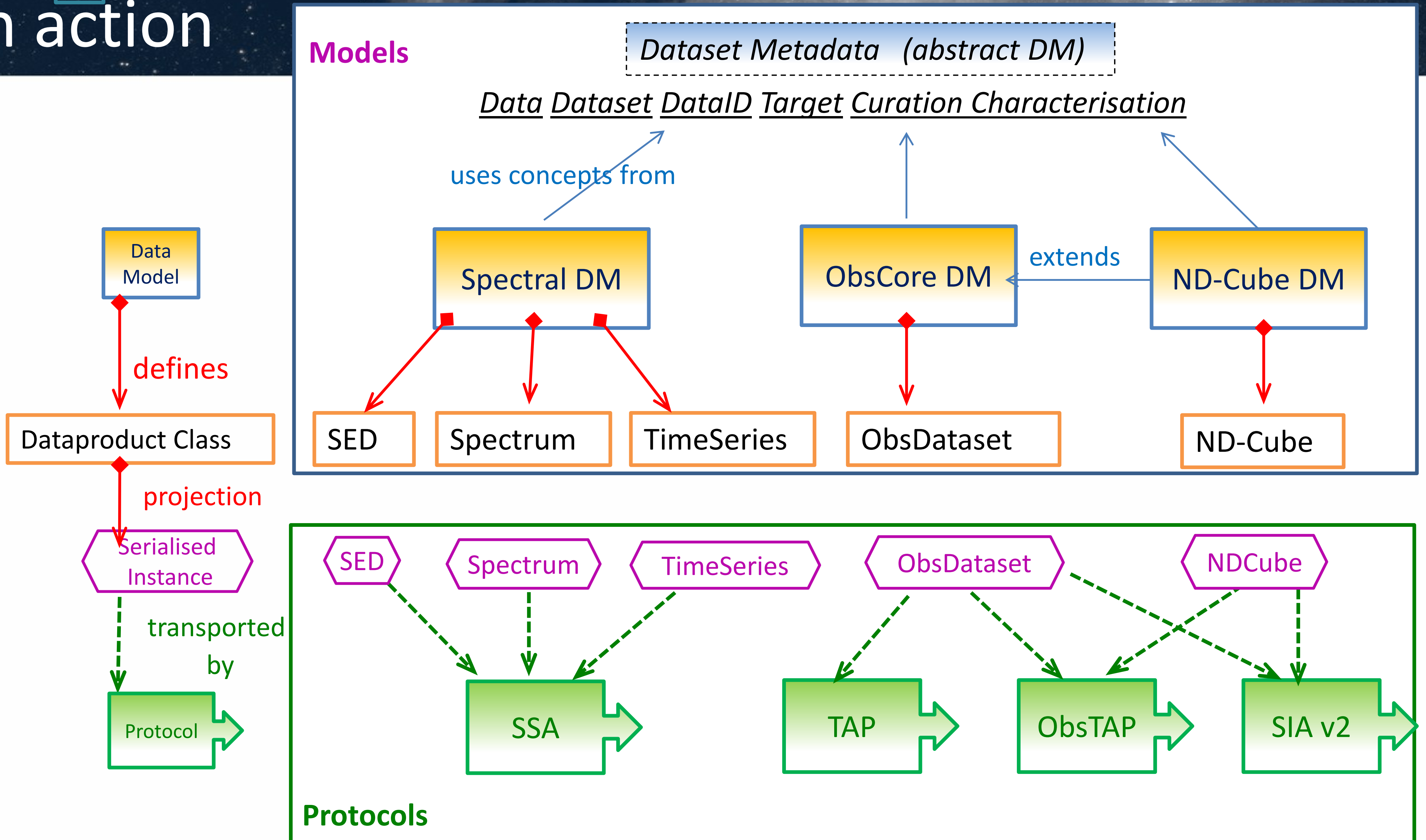
Interaction aspect /server side



Interaction aspect /client side



Data models in action



Lessons learnt (development strategy)

- Incremental development
- Gathering use-case has proven to be effective
- Discussions with data providers can be cumbersome but are essential
- Testing in real context with applications, protocols and data model update sets up a positive development feedback
- It was challenging to adjust the granularity level :
  - Details specific to particular data archives are not covered, but the common interoperable description layer works

Lessons learnt (technical aspects)

- Easy to prototype, but long and painful to get everything consistent and persistent in a large reference picture.
- Important to find the appropriate scope for data modeling:
  - Common features → mandatory data model fields
  - Specific features → optional features
  - We target interoperability and not exhaustivity
- Data model compliance checks improved when we can validate serialisations (owing to the Operations WG)

Conclusion

- VO data models offer a rich set of data models describing most of observation and simulation data products
- A good learning curve for VO developers thanks to the motivation of the actors
- Data models build up the underlying semantic and structural layer that binds together the VO framework
- Join us at IVOA meetings or contact us on dm@ivoa.net

Thanks for your attention

Thanks to all people involved

A collaborative effort thanks to

Jesus Salgado, Carlos Rodrigo, Pedro Osuna, Mark Allen, Mireille Louys, Jonathan McDowell, Deborah Baines, Jesus Maiz Apellaniz, Evanthia Hatziminaoglou, Sebastien Derriere, Gerard Lemson, Laurent Bourges, Miguel Cervino, Claudio Gheller, Norman Gray, Franck LePetit, Benjamin Ooghe, Rick Wagner, Herve Wozniak, Igor Chilingarian, Marie-Lise Dubernet, Evelyne Roueff, Matteo Guainazzi, Enrique Solano, Joe Mazzarella, Raffaele D'Abrusco, Tamas Budavari, Markus Dolensky, Inga Kamp, Kelly McCusker, Doug Tody, Pavlos Protopapas, Arnold Rots, Randy Thompson, Frank Valdes, Petr Skoda, Bruno Rino, Jim Cant, Omar Laurino, Patrick Dowler, Daniel Durand, Laurent Michel, François Bonnarel, Mark Cresitello-Dittmar, Markus Demleitner, Tom Donaldson, Matthew Graham,Jose Enrique Ruiz, Kristin Riebe, Michèle Sanguillon, Mathieu Servillat, Markus Nullmeier, Catherine Boisson, Claudia Lavalley