Astronomy and Computing 11 (2015) 74–80

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Data modeling for the virtual observatory

Mireille Louys *

Centre de Données astronomiques de Strasbourg, Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550,
11 rue de l' Université, F-67000 Strasbourg, France
ICube, Université de Strasbourg, CNRS UMR 7357, 300 bd Sébastien Brant - F-67412 Illkirch Cedex, France

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form
6 March 2015
Accepted 6 March 2015
Available online 18 March 2015

Keywords:
Data modeling
Astronomical databases
Virtual observatory tools

ABSTRACT

The data modeling effort has played a key role in the Virtual Observatory project, and contributed to the effort to build a common reference framework to describe the necessary information attached to astronomical data: the metadata. Such metadata describe the observing parameters and characterize and qualify the observed measurements. These pieces of information are produced and stored in project archives. Standardizing a homogeneous representation of metadata allows uniform discovery and use of the data in the Virtual Observatory infrastructure. This paper describes the context of data modeling in the VO architecture and shows how data models support requirements on the data access layer and for applications development. How the modeling process has been undertaken is explained with a short overview of the different data models. We also discuss in some detail the lessons learned in this modeling and standardization effort.

1. Introduction

This paper highlights the collaborative work undertaken in the Virtual Observatory (VO) project for modeling the observational metadata published by various astronomical data centers and used by scientists for their research programs. We present an overview of how the data modeling effort in the Data Model Working Group (DM WG), gathered and structured knowledge about observations and their metadata descriptions in a set of articulated data models.

During the VO development a number of other modeling efforts have been undertaken. The VOEvent group defined a description of sky events, with an adhoc protocol and data model (Seaman et al., 2011). Simulation codes and simulation data are modeled in a specific top-down approach (Lemson et al., 2014) led by the Theory interest group and endorsed by the DM WG. In this paper we concentrate on observational data, and explain the DM WG effort and its interactions with archive data providers, the Data Access Layer Working Group, and the applications developers. This work has involved strong interaction with other working groups efforts and developments. Section 2 outlines the approach to observational data. Section 3 describes the IVOA landscape and Section 4 provides details on the data modeling process. The current data models are described in Section 5 with discussion of the lessons learnt in Section 6. The *Glossary* section at the end of the paper defines the acronyms and development tools currently used in the VO initiative.

2. A dedicated approach for observational data

The VO data modeling effort is intended to organize and offer a description of the observational datasets in a logical and comprehensive way. It encodes common reference knowledge about metadata associated with observations that helps users navigate on-line distributed data collections, and allows publishers to efficiently describe their resources. It is also driven by science cases and the necessity to sort out, compare and confront data files from different observation programs. Interoperability has become a common concern for most disciplines in observational sciences; this has led to the emergence of similar projects such as Helio-VO[1] in heliophysics, or VAMDC[2] for atomic and molecular physics.

In the context of distributed astronomical science products, data are generally public after some proprietary period. The scientific value of data lasts for a long time. There is intrinsic value in data from different epochs for understanding time-dependent

* Correspondence to: Centre de Données astronomiques de Strasbourg, Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l' Université, F-67000 Strasbourg, France.
E-mail address: mireille.louys@unistra.fr.

[1] http://www.helio-vo.eu/.
[2] http://www.vamdc.eu/.

http://dx.doi.org/10.1016/j.ascom.2015.03.004
2213-1337/© 2015 Elsevier B.V. All rights reserved.

phenomena. The diversity of observing programs provides rich, heterogeneous data in many forms: light curves, spectra, spectral energy distributions, sky images, and velocity or spectral datacubes.

Astronomy data are stored for long term preservation so that they can be re-used in various studies later on. It is important to trace the data precision and statistical signature of each dataset individually, in order to characterize the content. Multiwavelength studies need precise descriptions of the instrumental parameters and of the statistical behavior of the measurements in order to reliably compare, match, superimpose, or combine observations in various regimes.

Moreover astronomy archives use a wide variety of database systems and architectures, with each organization using its own design for table definition, column names, etc. Therefore in order to allow scientists to seamlessly access a large collection of various archives, without having to learn each particular interface for accessing data of interest, there was a need for a common description frame for all astronomical metadata. This has been a strong incentive for the development of data modeling in the Virtual Observatory project and to build up protocols and applications in a consistent manner.

The approach taken for constructing data models was to gather various use-cases for data discovery, data retrieval, and data analysis, by interviewing astronomers for examples of usage and relying on the existing know-how of large data centers. From the collected ideas, we have tried to propose reconciling schemes for astronomical metadata description.

3. The data model landscape of the IVOA

Data modeling has been a central activity for the VO development as shown in the IVOA architecture document (Arviset and Gaudet, 2012). The interactions of the data modeling task principally lie in the definition of search parameters and representation of returned results in access protocols, so namely with the Data Access Layer Working Group (DAL WG). The class definitions elaborated by the DM WG can also feed the design of VO-aware applications by the Applications WG. The choice of a serialization format, to transport the modeled metadata, also involves the VOTable and the Semantics Working Groups. The ***ivoa.net*** (IVOA, 2014) document repository retains copies of the various data models available today. These have been designed for particular kinds of data products, first for space–time coordinates, then spectral datasets, 2D sky images, spectral lines, and data cubes. Most of the metadata used in astronomical protocols and applications are now derived from a stable set of data models as show in Table 1. These models are implemented and used by access protocols and client applications to effectively transport, visualize, transform, and interpret science observations.

4. Data modeling process

4.1. Metadata all around

Data modeling is focused on the metadata that describe the measurement values within an observation file: instrument name, file identifier, data provider, date of observation, date of publication, rights, position on the sky, field of view, instrument configuration, measurement quality, etc.

Historically this information was generally expressed in the FITS header keywords, with only a small set of standardized keyword labels for numerical data formats and WCS (World Coordinate Systems) location information. Most of the keywords used to express observation conditions, processing configuration, etc. do not obey a standardized vocabulary across the astronomical data centers.

Therefore a common, homogeneous, and structured representation of all these metadata was highly desirable in order to facilitate interoperability. Diverse use cases from protocol design in DAL WG, and from dataset handling in applications (Applications WG) helped to clarify usage of metadata and to sort out categories and roles for different pieces of metadata.

The concept of object oriented description appeared to be an adequate mechanism to represent various categories of metadata and group them logically. For instance, classes for Dataset, Curation, and Identification had been designed early in the Resource Metadata standard and organized in a tree-like structure as an XML schema[3] in VOResource and VODataService (Plante et al., 2010). These concepts, first stated in the Registry WG, are valid throughout the VO and are reused as key building blocks in other data models.

Coordinates (positional, temporal, spectral) are central in astronomy and have been modeled in detail, together with the various Coordinate Systems in the Space–Time Coordinates (STC)specification (Rots, 2011).

UML has been used since 2003 to express relationships between different concepts using class diagrams, and specifically to define each class and its attributes in detail. A text description is necessary to explain the properties of each class or attribute. It is provided along with each UML class diagram in the IVOA standard documents.

4.2. First steps to resolve metadata heterogeneity

Up to the 2000s, most astronomical data providers and reference archives used to store and distribute their observational datasets and the metadata attached to it following their own logic and policy. Each archive had its own interface. In order to homogenize the descriptions of source catalogs, the VOTable format has been proposed right at the beginning of the VO experience, to encode data and metadata in a tabular format. This very common data structure allows storage of all kinds of lists or collections with rows to store individuals, like sources, datasets, events, etc. and with columns representing properties measured or assessed for such individuals. The VOTable specification (Ochsenbein et al., 2011) defines basic XML elements such as FIELD, PARAM, and GROUP with attributes that qualify them. Among these qualifiers, two of them have led to the first steps of vocabulary standardization in the VO project: 'ucd' for the semantic content and 'unit' for expressing the units used for a column value.

The Unified Column Descriptor (UCD) (Preite Martinez et al., 2011) specifies a controlled vocabulary for the classification of the physical quantities exposed as a value in a column of a table. The collection of UCD terms covers most of the kind of measurements recorded in catalogs and observations in general. In the Data modeling effort, UCD words are used to add semantic value to some classes' attributes, like for instance to disentangle various kinds of flux measurements.

Another specification on which VO data modeling is based on, is the syntax definition of units strings for all measurement or metadata exposed in the VO system. Attributes describing such units in a VO Data Model should conform to the VO Units specification (Derriere et al., 2014) which gives the rules to compose a unit expression.

4.3. Metadata's scope and data model coverage

The DM WG has produced IVOA data models that are as comprehensive as possible with respect to their use cases, and

[3] See schemata at http://www.ivoa.net/xml/index.html.

Table 1
Current IVOA Data models with their scope and year of endorsement by the IVOA. All these standards specifications are available on the IVOA repository at http://www.ivoa.net/documents/.

Metadata	Data model name, version	Year of publication
Space–Time coordinate	STC v 1.33	2007
Physical axis description and properties	Characterization v1.13	2008
Spectral Line Transitions	Simple Spectral Line	2010
1D Spectrum, Light Curves	IVOA Spectrum v1.1	2011
Observational dataset (All data products)	ObsCore v1.0	2011
Photometric calibration	Phot v1.0	2012
SED, Photometric Points, Time series, Multi-segment 1D spectrum	Spectral v2.0	2014
N-D dataset, cubes complex observations sparse data	Image	2014 in review

thus cover the requirements for simple usage, but also more. This makes the models not only rich and very detailed, but also easily applicable by defining a set of mandatory data model items for the main usage (see Characterization DM, Spectrum DM, Obscore DM in Table 1).

The IVOA STC specification focuses on the definition and description of Coordinates on all physical axes used in astronomy. It includes many types of coordinates systems and is very useful to express positions, observation footprints, and sky regions.

The nature of the physical measurement of an observation is treated in detail in the Characterization DM (Louys et al., 2011b) following a coarse to fine progressive layered structure. It represents how data values span along all physical axes (spatial, spectral, temporal, flux, …), and defines the following properties for each axis: the *coverage* of an observation, the *resolution* information, the *sampling* and the *accuracy*. The Characterization DM represents these properties as classes in the description of each physical axis. Each property can be described using a common layered pattern: reference values, bounds, support, and variability, from the coarsest down to the finest level of description. It can fit many different situations by adjusting the desired level of detail on a combination of axes. These metadata are the core of a general data model, ObsCore DM (Louys et al., 2011a), that has been defined in order to homogenize the description of observational datasets across many VO-compliant archives. Indeed, when a user asks for observational data he/she will want to use constraints that focus on **what** data product is available, **where** on the sky it is observed, **when** it was observed, and **how** (which wavelength, frequency, filter, detector, polarimetric type, etc.). As an example a search in a database could use such criteria:

```
- product type = spectrum
- wavelength includes 6500 Angstroms
- spectral res > 15 Angstroms
- spatial res > 2 arcsec
- exptime > 3600 s
- data quality = any
```

ObsCore DM defines the necessary data model items that allow a translation into an ADQL, the dedicated query language developed in the VO for astronomical data bases. The corresponding query would be like:

```
SELECT * FROM ivoa.ObsCore
WHERE dataproduct_type = 'spectrum'
  AND t_exptime > 3600
  AND s_resolution< 5.5e-4
  AND 6500e-10 between em_min and em_max
```

These parameters already exist in data centers with their particular column name. Instead of modifying all column names for existing archives, IVOA data models annotate existing metadata with a homogeneous language and wrap them in a VOTable document. Data centers can then map their own metadata column names to the data model's definitions.

ObsCore DM exposes general metadata for all types of data products. It is restricted in terms of details and encompasses 25 mandatory metadata keywords. On the contrary Spectrum DM and Image DM (Cresitello-Dittmar et al., 2014a) are focused on particular data products: Spectrum DM v1.1 (McDowell et al., 2012) covers simple or multi-segment spectra and in the more recent update (Cresitello-Dittmar et al., 2014b) spectral energy distributions and light curves, while Image DM covers 2D sky images, data cubes, and more generally all N-dimensional datasets. Phot DM (Salgado et al., 2014) deals with photometric metadata such as calibration references and filter attributes. It shares the description of the coordinate system together with the Spectrum DM.

In terms of strategy, building up a complete and extensive Observation data model as a whole and in one go was not realistic. Data models needed to be iteratively elaborated, implemented, and tested for different types of data products and various actors contributing to the VO project. Therefore these different data model 'building-blocks' have been designed, with special care to stabilize valuable concepts for the domain of astronomy, to promote class reuse and avoid overlapping definitions.

In order to fit the evolution of the needs, any data model requires update, as appropriately mentioned in Dowler (2012). It is up to the IVOA data modeling group to maintain and warrant the consistency of this evolution path for the IVOA data models.

4.4. Validation of data models

Each DM in the VO is published in a standard document available in the IVOA standard documents repository[4] and validated by at least two reference implementations. In practice simple DAL protocols have been considered as partial implementations of the corresponding model, like for instance SSA for Spectrum DM, ObsTAP for ObsCore DM, or SIA version 2 for Image DM.

The complete validation of a data model, i.e., testing the pertinence and appropriate definition of all pieces of metadata in one single reference application, is generally not possible. The VO data models have been designed following a comprehensive view with respect to their corresponding use cases and so cover more than simple usage. This implies that several scenarios must be tested to assess the full sustainability of our model design.

Data model classes principally represent the metadata structure with attributes names, data types, text description, units, and UCD when applicable. They are valid and effective when they are proven to work for a large set of archives or data collections. This means that reference implementations can attach methods to each class data structure in order to map the data archived in a database to the objects exposed in the transport layer, and then propagated in the data serialization documents.

[4] http://www.ivoa.net/documents.

Table 2
Example of astronomical VO-aware applications using IVOA data models.

Application	Url	Data model
IRIS	http://cxc.cfa.harvard.edu/iris/v2.0.1/index.html	Spectrum 1.1
SPLAT-VO	http://www.g-vo.org/pmwiki/About/SPLAT	Spectrum 1.1 ObsCore 1.0
SAADA	http://saada.unistra.fr	ObsCore 1.0
SED-Builder	https://sdc.cab.inta-csic.es/vosed/index.jsp	PhotDM 1.0 Spectrum 1.1
Aladin	http://aladin.u-strasbg.fr/	Char 1.13, ObsCore 1.0

4.5. *Mandatory and optional data model items*

Depending on the use case, different subsets of metadata may be chosen to suit the purpose and be declared mandatory or optional. This allows homogeneity for baseline usage in the data access protocols, and extensibility when building up advanced interfaces or supporting dedicated data types such as those in high-energy physics.

With the data models now in place we have a representation framework that covers many types of observational metadata. Some classes are reused from one model to another, enriched or slightly modified, but each reuse context is fully documented in each VO data model. Instead of providing fine building blocks that can be arranged in various ways, VO data models highlight usage and know-how in the astronomy domain and warrant some logic and consistency in the arrangement of classes. The concepts defined comply with domain expertise.

5. Data models in use

As mentioned in Section 3 different participants developing the VO layer can take advantage of the IVOA data models:

- Application developers can find in the IVOA standards documents the meaning and usage of data model items and conform their variables or classes to them.
- Client developers can automatically refer to the XML schema representation of each model and then parse and validate the serialized data.
- Server application developers can translate or map their internal database column labels to an IVOA homogeneous interface.

Data models also serve as an interpretation language for applications to analyze the result of a query. Applications, as shown in Table 2, can build up their classes from data model serializations to organize their code.

Fig. 1 shows the various checking mechanisms taking place in a client–server interaction scenario. In the figure, the data model standard document (for instance Characterization DM or Spectral DM), lists in details the classes definitions, their properties, attributes and meaning, data types, recommended units, UCD tags, and usage. The client programmer can then design his/her own classes, by re-use and extension of the DM classes, and build up a query to ask for datasets in a VO query syntax (either ADQL, with TAP) or parameter query (with SIA/SSA protocols). On the other side, the server is made data model-aware by mapping its table columns (from a data base) to attributes of the data model class. This server will then deliver a VO Response file as a VOTable, with fields defined using the data model attributes. The spatial resolution information for some images discovered in an image archive, for instance, could be expressed as

<FIELD name=‘spatres’ datatype=‘double’
ucd=‘pos.angResolution’
unit=‘arcsec’
utype=‘char:Char.SpatialAxis.Resolution.Refval.Cresolution’>

This field exposes the spatial resolution stored in the column `spatres` of this database. The client could get another VOTable query response from a different archive, and still recognize the values for the spatial resolution, provided the other archive uses the appropriate data model markup. Here the mapping is done for a single Utype string on a single value. The data model standard document, XML schema, and Utype list make up the shared information between a client and a server application and help to realize the Virtual Observatory interface layer.

More generally, the VO response should contain a data model markup in order to map a GROUP of FIELD values to their corresponding object in a data model, if necessary. This strategy is currently developed in the VO-DML initiative.

Data models can help when designing a VO interface for a server to expose the content of an existing archive: they not only define the fields names, but they also specify the physical meaning and the computation and interpretation rules carried along with those fields. In other words, data models provide an interpretation framework for existing database content without knowing the implementation details of this particular database. The Table Access Protocol illustrates this situation very clearly.

The Table Access Protocol (TAP) allows the exposure of any kind of table content in the VO and supports high flexibility. Using references to a data model provides insights and clear explanations of the metadata exposed with a TAP service (Dowler et al., 2010, 2011).

TAP defines a specific schema called *ivoa* which can hold tabular views on VO data models. For instance, the table *ivoa.obscore* is a tabular representation of the observation data mapped against the ObsCore data model. The point here is that the columns of that table do not belong to the database design. They are defined by the ObsCore data model. That data discovery queries (as shown above) may be posed without probing the service metadata since semantics, units, columns names, etc., are common to all TAP services embedding ObsCore DM. This fosters a wide range of data discovery across a large number of data archives.

For instance the Xcatdb (Motch et al., 2007) builds up an ObsCore view of the XMM catalog data. In this way the service can serve any spectra, 2D-image, or time-series described with the ObsCore metadata attached. The native data remains accessible through other protocols (Web or VO).

As another example, the GAVO-TSAP service (Theoretical Spectral Access Protocol) also exploits TAP services to expose theoretical spectra. The spectral files are served with TAP with a subset of Spectral DM classes, adjusted for theoretical content.

At a larger scale, the ObsTAP strategy (TAP + ObsCore DM) is currently under consideration for building enhanced bibliographic services which will deliver, together with a scientific publication, the science data files used by the authors for the described work.

Fig. 2 illustrates the collaborative aspects between the Data Model Working Group, whose scope is shown in the top red rectangle, and the Data Access Layer Working Group, represented in the lower green rectangle. DAL protocols handle serialization documents that are elaborated according to data model concepts, with metadata keywords belonging to the data model representation and values bound to a particular dataset.

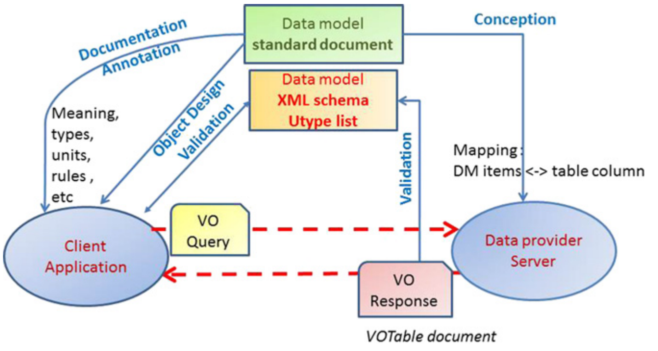

Fig. 1. Client and server applications exchange serialization documents, based on objects described in the data model standards documents. The client code may use the data model classes to build up queries following the DM keywords, get a VO response from the server, and validate the metadata content against the XML schema. Then it uses the values for computation or visualization in some classes derived from the data model specification (for instance built up from the XML schema). The data model documentation on the various data model items can also be exposed in a user-friendly interface (e.g. as tooltips). On the server side, the data model standard document and Utype lists are used to map the columns of server data base with Utypes tags, in order to attach a data model attribute to each piece of metadata serialized in the VO response.

The mechanism for binding one single piece of metadata, expressed as one column name in a VOTable, to a particular data model item is by using a Utype string that points to the data model description of the metadata keyword.

There have been different views about how to define this referencing mode:

- by using a Utype path to point to the appropriate class attribute in the DM hierarchy, as described in a serialization document
- by packaging the involved class structure into groups of FIELDREF in the VOTable document.

For homogenizing the representation of metadata across archives and taking into account the expertise at data centers, it was decided to standardize the data access layer and the representation of datasets instead of modifying each archive architecture and imposing constraints on data base schemata to all.

The interoperability effectively takes place when data centers produce outputs following DAL response formats and the circulating metadata documents use the data model terms to tag the exposed metadata.

At the user's end, also, applications recognize the DAL serialization keywords and use them for manipulating the data, e.g. TapHandle (Michel et al., 2014), cutout services, etc.

6. Discussion and lessons learnt

6.1. Adopt the appropriate granularity

Selection of and references to the most used VO data model attributes in an unequivocal way have worked well for data discovery so far. The ObsCore DM take-up has been a success, not only due to appropriate granularity and complexity of the data model, but also to the simultaneous take-up of the TAP protocol. Today the ObsCore data model can be used for exposing any data collection, within an ObsTAP service, and very soon with SIAV2 protocol.

6.2. Different perspectives

We agreed about the VO strategy to first focus on data discovery and data retrieval and secondly on data analysis use cases. Supporting all data analysis use cases is a challenging task as applications and analysis software evolve more quickly than archive infrastructure. While all physical dimensions of datasets are fully covered by existing VO data models, the Provenance information, which deals with the way data have been obtained and reduced, is not yet modeled. This is an ongoing task in the DM WG, but not yet completely delimited or scaled. Coping with the variety of dependencies between data products elaboration steps and instrumental contexts is a challenge.

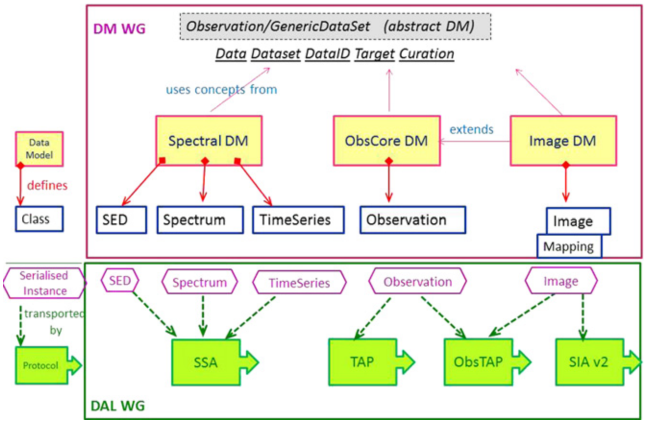

Fig. 2. Data models define major classes that represent common metadata to deal with the different kinds of data products. They are based on abstract concepts like Dataset, DataID, Curation, which are necessary for data management in general. Each data model specifies the properties of a major class, Observation/Dataset (ObsCore, ImageDM) with attached attributes and subsidiary classes. Protocols use instantiated objects, (class structure + attribute values) stored in VOTable serialization documents (in purple). They pass this information to applications which parse the VOTable document, check the data model attributes, and launch their processes accordingly. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.3. Data model extensions

From the object-oriented paradigm adopted in the VO data models, one can build up ad hoc models by deriving and extending classes of the current VO data models. This introduces flexibility for application developers who can design richer class libraries and extend their methods. However, it is difficult to maintain a large collection of derived classes in general in object oriented programming. Failure to coordinate various derived models in the VO framework could put interoperability at risk. Up to now, we do not encourage data model extensions, unless a clear new use-case is exposed and worked out, which then could lead to an update of the related data model, in the IVOA frame.

6.4. Serialization strategies

Circulating the metadata information as described in VO data models requires a communication framework to allow data providers to wrap their metadata in VO-compliant serialization formats. Applications can unwrap and interpret these metadata, provided they are 'data model-aware'. DAL protocols concern the s description of the communication process for the transport layer including: access mode, required metadata, output format, parameters, etc. but do not fully describe the interpretation context for the data.

VOTable is the format of choice to transfer lists of data sets in the Virtual Observatory, and especially the metadata attached to such datasets. Quite naturally the VOTable format defines a Utype tag for each VOTable element like FIELD, PARAM, and GROUP, which are used as extra information to locate and interpret the values stored via VOTable elements within the context of a data model. This mechanism binds a value to an attribute in the data model representation, generally an XML Schema. This is a bottom-up, fine grain labeling system that works effectively for small amounts of metadata, and scales nicely for large data collections. The Utype string matches a path in the UML class diagram of a data model. Originally these data model tags were loosely defined in the VOTable specification. This has been discussed extensively, but not yet finalized in a specification like the other VO recommendations. Each IVOA data model is published together with a list of Utypes derived from its class architecture, and can be used as a reference list. A better way of exposing Utypes definition and circulating examples of usage is planned to serve future users and developers.

The Utype string is to be distinguished from a UCD string and serves a different purpose. A UCD tag offers a classification of a quantity, according to general physical knowledge of the measurements that astronomers study: temperature, position, proper motion, atomic and molecular lines, energy bands, for instance. UCDs build up a controlled vocabulary to support the astrophysical science. On the contrary, the scope of a Utype is constrained by the point of view taken in the data model context; it is bound to a role attached to its attribute partner with respect to the class structure in the referenced data model.

Up to now, the Utype string mapping has been based on a one to one relationship between a simple quantity and a label in the data model. It fits the current needs of tabular serialization as applied in many data centers and experienced in the Saada project (Michel et al., 2006). It allows annotation of partial views of a DM, to describe only some attributes of a model object and therefore offers a simple and easy mechanism for datasets annotation on the data provider's side.

However, another way to serialize data model instances in VOTable is currently being examined. VO-DML seeks to bring more of the object-encapsulated structure into the serialization document instance. It is meant to automatically derive the data model tags from a UML class diagram, and to explicitly describe class aggregation and relations. This is a work in progress and is currently being discussed in the DM WG.

6.5. *Testing the pertinence of the modeling effort*

The major part of metadata described in the current IVOA data models has been effectively used and tested in real scale, namely with the development of DAL protocols and applications. Still, because they have been built on a large set of science driven use cases, some situations anticipated at the modeling phase may not have been tested extensively due to priority constraints, new data collections, etc. This means that IVOA data models may evolve with the requirements of new science use cases and new development frameworks emerging in information technology.

7. Conclusion

The data modeling effort has played a key role in the development of the VO initiative. It has been driven by astronomical community use cases. It has gathered information and structured knowledge about metadata that is now formalized as a set of VO-recommended standard data models. This constitutes a reference frame for further development in the astronomical community. This happened due to a good cooperation inside and between working groups and as such was a good experience. Identifying common objectives across national projects, in compatibility with the astronomical community's interests and with a goal of reaching consensus, has proven to be quite effective.

Acknowledgments

Many collaborators in the VO working groups and especially in the Data Model Working Group have contributed to this work. I want to thank Jonathan McDowell (CfA) who did a lot to launch the Data Modeling effort in the VO. Many fruitful discussions happened with people involved in data centers, including Anita Richards (University of Manchester), Alberto Micol (ST-ECF, ESO), Daniel Durand (CADC), Doug Tody (NRAO), François Bonnarel (CDS), Laurent Michel (SSC-XMM, Saada), Arnold Rots (CFA), Jesus Salgado, and Pedro Osuna (ESAC) who brought their expertise and insights to make cooperative progress happen between data modeling and data access strategies. The widespread adoption of the ObsCore data model has been greatly and efficiently supported by Pat Dowler and CADC colleagues and by all DAL developers implementing the ObsTAP specification. Thanks to the VO-DML initiative, introduced by G. Lemson (MPI) and O. Laurino (CfA), a framework is proposed to consistently wrap and unwrap the data model representation across the data access protocols using a fully formalized mapping of object instances to VOTable serializations. Special thanks are also owed to Mark Cresitello-Dittmar (CfA) for providing a comprehensive and advanced model for N-dimensional datasets based on the building blocks elaborated throughout this VO project. In addition, much feedback from the community from users, developers, and data publisher have been very helpful. This long term data modeling effort has been funded by VO-involved national projects and their respective funding agencies. This work has been supported by the French CNRS agency and European projects, namely EURO-VO projects as AVO, VO-Tech, AIDA, VO-ICE, and CoSADIE.

References

Arviset, C., Gaudet, S., IVOA Technical Coordination Group, 2012. The IVOA Architecture. In: European Planetary Science Congress, p. 626, September.

Cresitello-Dittmar, M., Tody, D., Bonnarel, F., Louys, M., Rots, A., Ruiz, J.E., Salgado, J., 2014a. IVOA image data model version 1.0. IVOA Proposed Recommendation. URL: http://www.ivoa.net/Documents/ImageDM/.

Cresitello-Dittmar, M., et al. 2014b. IVOA spectral data model version 2.0. IVOA Proposed Recommendation. URL: http://www.ivoa.net/documents/SpectrumDM/.

Derriere, S., Gray, N., Louys, M., Demleitner, M., Ochsenbein, F., 2014. Units in the VO. IVOA Recommendation. URL: http://www.ivoa.net/Documents/VOUnits/index.html.

Dowler, P., 2012. CAOM-2.0: the inevitable evolution of a data model. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. p. 339.

Dowler, P., Rixon, G., Tody, D., 2010. Table access protocol version 1.0. IVOA Recommendation. URL: http://www.ivoa.net/Documents/TAP.

Dowler, P., Rixon, G., Tody, D., 2011. IVOA Recommendation: Table Access Protocol Version 1.0. ArXiv e-prints arXiv:1110.0497.

IVOA 2014. IVOA standard documents. Repository. URL: http://www.ivoa.net/documents.

Lemson, G., Bourges, L., Cervino, M., Gheller, C., Gray, N., LePetit, F., Louys, M., Ooghe, B., Wagner, R., Wozniak, H., 2014. IVOA Recommendation: Simulation Data Model. ArXiv e-prints arXiv:1402.4744.

Louys, M., Bonnarel, F., Schade, D., Dowler, P., Micol, A., Durand, D., Tody, D., Michel, L., Salgado, J., Chilingarian, I., Rino, B., Santander-Vela, J.D., Skoda, P., 2011a. IVOA Recommendation: Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0. ArXiv e-prints arXiv:1111.1758.

Louys, M., Richards, A., Bonnarel, F., Micol, A., Chilingarian, I., McDowell, J., the IVOA Data Model Working Group, 2011b. IVOA Recommendation: Data Model for Astronomical Dataset Characterisation. ArXiv e-prints arXiv:1111.2281.

McDowell, J., Tody, D., Budavari, T., Dolensky, M., Kamp, I., McCusker, K., Protopapas, P., Rots, A., Thompson, R., Valdes, F., Skoda, P., Rino, B., Derriere, S., Salgado, J., Laurino, O., IVOA Data Access Layer, t., Data Model Working Groups, 2012. IVOA Recommendation: Spectrum Data Model 1.1. ArXiv e-prints arXiv:1204.3055.

Michel, L., Louys, M., Bonnarel, F., 2014. Browsing TAP services with TAPhandle and datalink. In: Manset, N., Forshay, P. (Eds.), Astronomical Society of the Pacific Conference Series. p. 15.

Michel, L., Nguyen, H.N., Motch, C., 2006. How to publish local data into the VO with Saada. In: Gabriel, C., Arviset, C., Ponz, D., Enrique, S. (Eds.), Astronomical Data Analysis Software and Systems XV. p. 25.

Motch, C., Michel, L., Pineau, F.X., 2007. The XCATDB: a complex database based on saada. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), Astronomical Data Analysis Software and Systems XVI. p. 699.

Ochsenbein, F., Williams, R., Davenhall, C., Durand, D., Fernique, P., Giaretta, D., Hanisch, R., McGlynn, T., Szalay, A., Taylor, M.B., Wicenec, A., 2011. IVOA Recommendation: VOTable Format Definition Version 1.2. ArXiv e-prints arXiv:1110.0524.

Plante, R., Stébé, A., Benson, K., Dowler, P., Graham, M., Greene, G., Harrison, P., Lemson, G., Linde, T., Rixon, G., 2010. VODataService: a VOResource schema extension for describing collections and services version 1.1. IVOA Recommendation. URL: http://www.ivoa.net/Documents/VODataService/.

Preite Martinez, A., Derriere, S., Delmotte, N., Gray, N., Mann, R., McDowell, J., Mc Glynn, T., Ochsenbein, F., Osuna, P., Rixon, G., Williams, R., 2011. IVOA Recommendation: The UCD1+ Controlled Vocabulary Version 1.23. ArXiv e-prints arXiv:1110.0518.

Rots, A.H., 2011. IVOA Recommendation: Space–Time Coordinate Metadata for the Virtual Observatory Version 1.33. ArXiv e-prints arXiv:1110.0504.

Salgado, J., Rodrigo, C., Osuna, P., Allen, M., Louys, M., McDowell, J., Baines, D., Maiz Apellaniz, J., Hatziminaoglou, E., Derriere, S., Lemson, G., 2014. IVOA Recommendation: IVOA Photometry Data Model. ArXiv e-prints arXiv:1402.4752.

Seaman, R., Williams, R., Allan, A., Barthelmy, S., Bloom, J., Brewer, J., Denny, R., Fitzpatrick, M., Graham, M., Gray, N., Hessman, F., Marka, S., Rots, A., Vestrand, T., Wozniak, P., 2011. IVOA Recommendation: Sky Event Reporting Metadata Version 2.0. URL: http://cdsads.u-strasbg.fr/abs/2011arXiv1110.0523S.

Glossary

ADQL*:* IVOA Astronomical Data Query Language. ADQL is derived from the Structured Query Language (SQL) language dedicated to support generic and astronomy specific search operations at archives centers.
http://www.ivoa.net/documents/latest/ADQL.html.

ObsTAP*:* ObsTAP is a dedicated table definition for the TAP protocol to query and discover datasets described by the Observation Data Model Core components. Its definition is in the same document as ObsCore.
http://www.ivoa.net/documents/ObsCore/.

Provenance*:* For an observation, there is a chain of actions taken to transform the raw signal in a telescope into science ready data. This implies both an instrumental and a processing side, that are respectively, instrumental settings and observing conditions, and the processing steps, parameters configuration, and calibration details. This has been a research topic for other domains and conceptually formalized by the W3C consortium as shown in
http://www.w3.org/TR/prov-overview/.

SIA*:* Simple Image Access; this protocol has capabilities for the discovery, description, access, and retrieval of multi-dimensional image datasets, including 2-D images as well as datacubes of three or more dimensions.
http://www.ivoa.net/documents/SIA/.

SSA*:* Simple Spectral Access; this protocol defines a uniform interface to remotely discover and access one dimensional spectra.
http://www.ivoa.net/documents/SSA/.

UCD*:* The Unified Content Descriptor (UCD) is a formal vocabulary for astronomical data that is controlled by the IVOA. The vocabulary is restricted in order to avoid proliferation of terms and synonyms, and controlled in order to avoid ambiguities as far as possible. It is intended to be flexible, so that it is understandable to both humans and computers. UCDs describe astronomical quantities with string labels; these are built by combining words from the controlled vocabulary. http://www.ivoa.net/documents/REC/UCD/UCD-20050812.html.

UML*:* The Unified Modeling Language (UML) is a formal and graphical language used for the design and description of applications and information systems following the principles of Object Oriented Programming. It is used in the IVOA data modeling effort to set up class diagrams which show details on classes attributes and logical relationships between classes.
http://www.uml.org.

Utype*:* A Utype is a label used as an identifier for a concept defined within an IVOA data model. Utypes are semantically equivalent to a URI or XPath in XML. Although simple in principle, parsability or non-parsability of these strings in the VO applications led to intensive discussions. http://wiki.ivoa.net/internal/IVOA/Utypes/WD-Utypes-0.7-20120523.pdf.

VO-DML*:* A serialization framework to map object-oriented data models defined by the IVOA into the tabular structure of a VOTable. See the discussion page at
http://wiki.ivoa.net/bin/view/IVOA/VODML.

VOTable*:* The VOTable format is an XML standard for the interchange of data represented as a set of tables. In this context, a table is an unordered set of rows, each of a uniform structure, as specified in the table description (the table metadata). Each row in a table is a sequence of table cells, and each of these contains either a primitive data type or an array of such primitives. VOTable is derived from the Astrores format, itself modeled on the FITS Table format. This is used all around the VO framework to describe lists of data and/or metadata. http://www.ivoa.net/documents/VOTable/20130920/REC-VOTable-1.3-20130920.html.

XML*:* eXtensible Markup Language. A structured language that allows to describe any structured document. It defines a set of markup tags, and their structure in an XML Schema. This allows metadata files circulating in the VO infrastructure to be checked or be transformed using XML tools such as XSLT.
http://www.w3schools.com/xml/.