Diversity of domain descriptions in natural science: virtual observatory as a case study[♣]

© Briukhov D.O., Kalinichenko L.A., Zakharov V.N.

Institute of Informatics Problems RAS
brd@ipi.ac.ru, leonidk@synth.ipi.ac.ru, vzakharov@ipiran.ru

Abstract

Main intention of this publication is to attract attention of the DL and e-Science research community to the diversity and complexity of domain descriptions in natural sciences. We have chosen the domain of astronomy where Virtual Observatory development is very intensive around the world. In the domain of astronomy we show the various standards that help to overcome this diversity. These standards are under development by the International Virtual Observatory Alliance (IVOA). The diversity demonstrated is a challenge for the IT people and a warning that should prevent on light weighted promises to create rapidly uniform and integrated definition of a collective memory for the science as a whole (or even for its separate domains).

1 Introduction

The Technical Committee on Digital Libraries of the Institute of Electrical and Electronics Engineers Computer Society (TCDL of IEEE-CS) to define what DL is, uses the more general term "(digital) *collective memory*" to emphasize the convergence of sources of various kinds. Collective memory development faces challenges in several areas, including storage, classification, and indexing; user interfaces; information retrieval; content delivery; presentation, administration; preservation, etc. In contrast to conventional digital library entities, collective memories in different branches of science should be differently structured. More suitable entities would be concept spaces, theories, models, hypotheses, experimental results and measurements, curricula, and educational modules. Scientists have spent centuries to reach well-defined structures, concepts and theories in various branches of science. These definitions cannot be used following the conventional library metaphor, but are more suitable as a guiding principle for information structuring and search in digital libraries. For this reason, the gradual evolution of digital libraries from the currently dominated framework based on the conventional library metaphor to more knowledge-based organization is expected. With time and experience, these frameworks will be upgraded with conceptual definitions (ontologies) of subject domains and curricula along with the conventional metadata so that information resources can be registered in accordance with the proper subject definition and granularity. This trend will also lead to a higher level of coherency of the information collected in a specific subject domain, by contrast with metadata use, where collected materials are more diverse though less relevant to the subject.

e-Science refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections (collective memories), very large scale computing resources and high performance visualisation back to the individual user scientists. Knowledge-based collective memory in a domain of a natural science includes domain terminology and concept definitions, material system description, definitions of various theories and models, observable (measurable) characteristics of real world objects, description of methods and instruments for observation, measurement, observation and experimental data, data analysis results, problem definitions and methods of solution, algorithms and programs, simulations. Integration of such information is driven by scientific and educational needs.

Numerous forms of digital sources representations can be included into collective memories as distributed repositories of knowledge. Until some uniformity can be imposed on the available forms, the collective memory clients will feel themselves in much uncomfortable condition than in conventional libraries. The problem facing researchers and developers in collective memories is fundamental: how to map huge variety of digital sources into their uniform representation and how to support the basic memory function of providing access to the integrated collection of heterogeneous information?

To answer this question just studying an astronomy as an example, we provide an overview of standards that are under development by the International Virtual

Proceedings of the 7th Russian Conference on Digital Libraries RCDL'2005, Yaroslavl, Russia, 2005

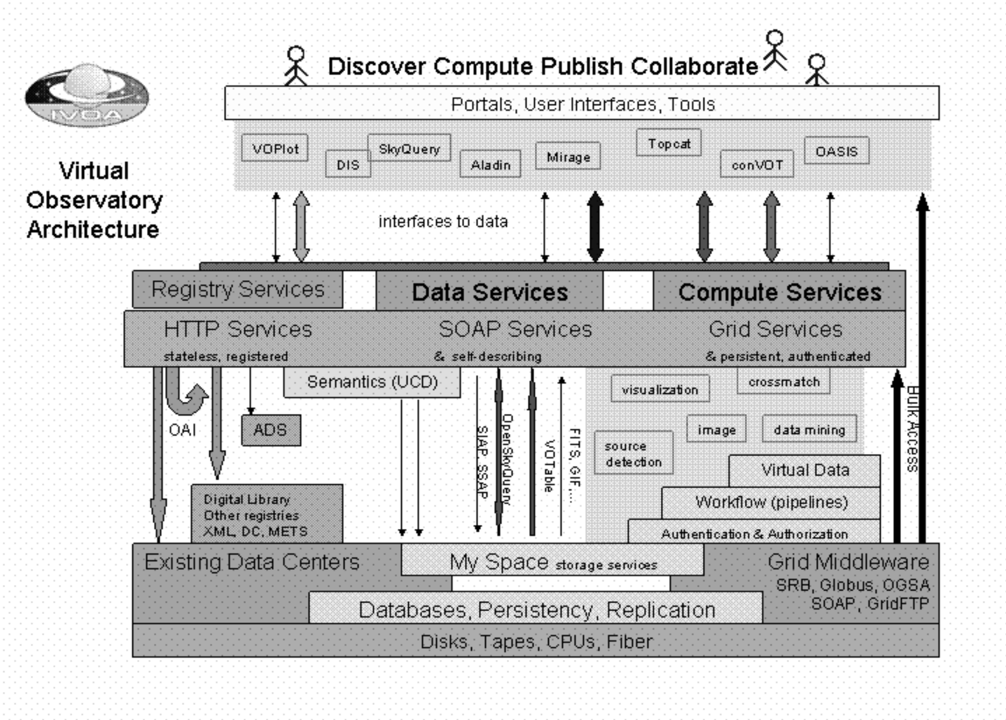

Figure 1. IVOA Architecture

Observatory Alliance (IVOA). The set of standards shows how specific could be domains of natural science in data modeling, data access layering, conceptualization, query languages and interfaces.

With this paper we would like to emphasize how hard is the problem of creating collaborative scientific enterprises in one domain and especially across the domains. Our intention is to show that due to such diversity, very specific facilities are to be designed in order that such semantic collaboratories could start to work.

2 Virtual observatory architecture according to IVOA

This section provides an overview of the IVOA standards [12] in accordance with their state at the end of 2004. The development of architectural decisions and standards is accomplished by 8 IVOA Working Groups: Resource Registry, Data Modeling, Content Description (UCD), Data Access Layer, VOTable, VO Query Language, Grid & Web Services, Standards & Processes.

The architecture of the VO is *service oriented*, meaning that components of the system are defined by the nature of requests and responses to services. Because of this, the description of each service is based on the choice of the protocols for requests and responses, rather than classes and methods. Data is communicated between services in two basic formats: FITS and XML.

Fig. 1 provides a high-level conceptual overview of the IVOA-supported architecture of Virtual Observatories that has emerged over the last few years.The top bar of the figure represents the VO science-oriented objectives: discovery of data and services, reframing and analyzing that data through computation, publishing and dissemination of results, and increasing scientific output through collaboration and federation. The IVOA does not specify or recommend any specific portal or library by which users can access VO data, but some examples of these portals and tools are shown in the grey boxes.

Different vertical arrows represent the different service types and XML formats by which these portals interface to the IVOA-compliant services. In the IVOA architecture, the available services are divided into three broad classes:

- Data Services, for relatively simple services that provide access to data;
- Compute Services, where the emphasis is on computation and federation of data;
- Registry Services, to allow services and other entities to be published and discovered.

These services are implemented at various levels of sophistication, from a stateless, text-based request-response, up to an authenticated, self-describing service that uses high-performance computing to build a structured response from a structured request. In the VO, it is intended that services can be used not just individually, but also concatenated in a distributed *workflow*, where the output of one is the input of another. The registry services facilitate publication and discovery of services.

Each registry has three kinds of interface: publish, query, and harvest. People can publish to a registry by filling in web forms in a web portal, thereby defining services, data collections, projects, organizations, and other entities. The registry may also accept queries in a one or more languages, and thereby discover entities that satisfy the specified criteria. The third interface, harvesting, allows registries to exchange information between themselves, so that a query executes at one registry may discover a resource that was published at another. Registry services expect to label each VO resource through a universal identifier. Resources can contain links to related resources, as well as external links to the literature, especially to the Astronomical Data System. The IVOA registry architecture is compliant with digital library standards for metadata harvesting and metadata schema, with the intention that IVOA-compliant resources can appear as part of every University library.

Data services range from simple to sophisticated, and return tabular, image, or other data. At the simplest level (conesearch), the request is a cone on the sky (direction/angular radius), and the response is a list of "objects" each of which has a position that is within the cone. Similar services (SIAP, SSAP) can return images and spectra associated with sky regions, and these services may also be able to query on other parameters of the objects.

The OpenSkyQuery protocol drives a data service that allows querying of a relational database or a federation of databases. In this case, the request is written in a specific XML abstraction of SQL that is part of ADQL (Astronomical Data Query Language).

The IVOA architecture will also support queries written at a more semantic level, including queries to the registry and through data services. To achieve this, the IVOA is developing a structured vocabulary (simple ontology) called UCD (Unified Content Descriptor) to define the *semantic type* of a quantity.

The IVOA expects to develop standards for more sophisticated services, for example for federating and mining catalogs, image processing and source detection, spectral analysis, and visualization of complex datasets. These services will be implemented in terms of industry-standard mechanisms, working in collaboration with the grid community.

Members of the IVOA are collaborating with a number of IT groups that are developing workflow software, meaning a linked set of distributed services with a dataflow paradigm. The objective is to reuse component services to build complex applications, where the services are insulated from each other through well-defined protocols, and therefore easier to maintain and debug. IVOA members also expect to use such workflows in the context of *virtual data*, meaning a data product that is dynamically generated only when it is needed, and yet a cache of precomputed data can be used when relevant.

In the Fig. 1, the lowest layer is the actual hardware, but above that are the existing data centers, that implement and/or deploy IVOA standard services. Grid middleware is used for high-performance computing, data transfer, authentication, and service environments. Other software components include relational databases, services to replicate frequently used collections, and data grids to manage distributed collections.

A vital part of the IVOA architecture is *MySpace* so that users can store data within the VO. MySpace stores files and DB tables between operations on services; it avoids the need to recover results to the desktop for storage or to keep them inside the service that generated them. Using MySpace estblishes access rights and privacy over intermediate results and allows users to manage their storage remotely.

The IVOA architecture uses services at different levels: HTTP GET/POST services, SOAP services, Grid services. In the IVOA architecture, a VO-compliant web service is defined as one that can also supply a VOResource description of the service, including curation, description, sky region, IVOA identifier, and other information.

2.1 Data Modeling

2.1.1 A unified domain model for astronomy, for use in the Virtual Observatory

The document “A unified domain model for astronomy” [3] is the IVOA attempt to define a conceptual model, created as the result of the domain knowledge extracted from the modelers (some of whom are astronomers) and their direct coworkers as well as from literature and other external references. Authors consider that the model can be used in various ways:

- It can be used as the basis for a meta-data repository that archives can use to describe their data products in a common model;
- It can be used as a model describing the entities (classes and attributes) that can be used in a common query language for these astronomical archives and for the relations between that can be followed from these entities in navigation to related ones;
- It can be mapped to an XML schema, to a Java or C# class library, to a relational database schema, allowing reference implementations for these particular bindings;
- It can simply serve as a formal, common language in “whiteboard discussions” about the structure of particular data products.

Though for RVO the approach looks attractive (domain model might be used as a mediator schema), it is doubtful that global data model for the whole domain of astronomy could succeed. Each class of astronomical problems will introduce its own concepts, data structures, behaviors convenient for the respective problems (see section 7 on the classes of astrophysical problems). Each new instrument and changing in observational technology will lead to new kinds of data that could not be foreseen in advance. Therefore, it seems that data modeling approach should provide much more flexibility to survive.

It is said in the document that the way of using the common domain model is equivalent to an *ontology*. At the same time the main difference between conceptual model and ontology consists in the following. Conceptual model can be used as a global schema over existing heterogeneous data sources and services. It means that existing sources/services can be registered at the conceptual model, mapped to it so that querying through the domain definition of the registered sources could be possible. Ontology is used as a reference definition of the domain concepts and relationships between them. Such definitions of concepts can be used for annotation of elements of various data models in the domain to provide them with the adequate semantics (cf. UCDs as a step towards simple ontology).

What is defined now in the document "A unified domain model for astronomy" might be more suitable to consider as an attempt to provide a draft definition of an ontology for the domain as a description of sets of concepts and relationships between them.

Alongside with a unified domain model, specific data models [3] are being defined for various kinds of astronomical data, such as Spectra, Quantity, Observations, Transforms, Catalogs, Inteferometry, Simulations, Passband, Error/Accuracy. Some of these models are overviewed in the subsequent subsections.

2.1.2 Data model for quantity

A VO data model to describe the semantic content of sets of astronomical data values and their most closely associated metadata has been defined. The model may be used by aggregation or extension in higher level models describing astronomical datasets. Any value must be associated with a physical concept which can be tagged as a UCD, and with a physical unit. A set of interfaces to an object called Quantity and to some related objects is intended to be defined.

The following concepts are involved into the quantity model: accuracy, quality, array axes, coordinates, frames, coordinate systems, units, transformations. XML serialization for the proposed values is also defined.

2.1.3 IVOA Observation data model

A comprehensive data model named ’Observation’ for observational data is currently being defined (Data Model for Observation, IVOA WG internal draft). An Observation can be a spectrum, an image, a time series, or a higher dimensional combination of those. This model attempts to identify the different aspects that fully describe either a single observation of the sky, or a dataset derived from a number of observations. It therefore represents a description of all the metadata that may be required by both data discovery and retrieval services and data analysis applications. Metadata in this document means any data associated with the observation except for the astronomical measurements themselves.

The Observation DM can be used in different ways depending on the context. In frame of the DAL (IVOA Data Access Layer), the DM will provide standard tags to formulate a query to a VO-compliant data provider (the Coverage part of the model described below will play a frequent role here) and a standard to describe the results of such a query (like the metadata tree used in IDHA). In the context of data processing and analysis, the DM will provide a standard way to describe the accuracy, the resolution and the sampling applied to any observation. This lets tools handle observations from different archives in a systematic way. The description of the instrument configuration used to collect the data is useful in a variety of analysis and query contexts.

2.1.4 Simple Spectral Data Model

This is a data model describing the structure of spectrophotometric datasets with spectral and temporal coordinates and associated metadata. This data model may be used to represent SED (spectral energy distributions), spectra, and time series data. Spectra are stored in many different ways within the astronomical community. The IVOA model presents an abstraction for spectral data. It is required to represent a single 1-dimensional spectrum, time series photometry, spectral energy distributions which consist of multiple spectra and photometry points.

Spectral data model is based on such concepts as Spectrum and Time Series, Spectral coordinate, Flux (Spectral Intensity) Object, BackgroundModel Object, Time coordinate, Position coordinate, Accuracy Fields. Associated Metadata Fields include Coverage Fields, Frame fields, Derived Data Fields, Curation model, Data Identification model.

The Spectrum model involves objects addressed by the proposed VO Observation and Quantity data models. A single Spectrum maps to the Observation model, which will include the Curation and Coverage objects. The Flux and the spectral coordinate entries together with their associated errors and quality will be special cases of the Quantity model, as will the simpler individual parameters. FITS serialization, VOTable Serialization and Direct XML serialization are defined for the spectral model.

2.1.5 Simulation Data Model

A data model for simulation data (named ’Simulation’) is being developed within the framework outlined by the Observation model. The three main sub-categories – Simulation Data, Characterization and Provenance are still applicable. However, for simulation data it is the Provenance object, rather than Characterization that contains the real descriptive content of the model.

This object remains essentially the same as in the Observation model – a subclass of the Quantity object, used to contain the main data output of the simulation. However, for simulated data there is potentially a much wider range of quantities to be stored. In Observation at least one quantity in the data must be an observable; this is not the case in Simulation. The metadata structure – the set of UCD’s used to describe each quantity must be enlarged to incorporate data clearly labeled as being ’theoretically derived’.

The Provenance object contains most of the information describing the simulation. This is because, unlike during an observation, most of the effort in acquiring the data is not through measurement but through the execution of numerical routines, thus creating the data set. The Provenance object is defined as ’the description of how the dataset was created’ which for a simulation is possible to describe entirely.

Provenance can be broken down into the Theory, Computation and Parameters. Theory describes the underlying fundamental physics upon which the simulation is based. Computation describes the technique used to evaluate the physics described in Theory through the execution of numeric routines. Parameters not only define the physical context of the simulation, but also the resolution and detail. If the algorithms are analogous to a mathematical function, the parameters are the values of the input variables.

2.2 Unified Content Descriptors (UCD)

The Unified Content Descriptor (UCD) [10,11] is a formal vocabulary for astronomical data that is controlled by IVOA. The major goal of UCD is to ensure interoperability between heterogeneous datasets. The use of a controlled vocabulary will hopefully allow an homogeneous, non-ambiguous description of concepts that will be shared between people and computers in the IVO. A UCD is a string which contains textual tokens that are called *words*, which are separated by semicolons. A word may be composed of several *atoms*, separated by period characters. The order of these atoms induces a hierarchy. The UCD system is an attempt to describe simply the most commonly used quantities that astronomers want to exchange. It gives *standard names to properties of instances of concepts.*

UCD will be used in practice for *exchanging* information using a controlled vocabulary. They are used in the VOTable standard to attach a standard description to table column names. What is needed for interoperation with other systems is a "*translation layer*" that is able to associate UCD to the parameters that are used internally, so that the output of the service contains a standard description that can be interpreted by other VO services.

2.3 Metadata Registries for VO

2.3.1 Resource Metadata for the Virtual Observatory

A *registry* is a query service for which the response is a structured description of resources. Resource metadata constitute a "yellow pages" of astronomical information. Metadata about resources and services in VO are standardized. *Resource metadata* [6] are generic, high-level, and independent of any specific service.

Resource metadata are typically not queryable parameters in the underlying services, but rather they encompass information that now is simply "known" to users, or must be discovered through other means. *Service metadata* are an extension of the general resource metadata describing *how* to access the resource. Resource metadata are collected through resource registration services. The most general resource metadata is similar in concept to the Dublin Core metadata definitions.

IVOA document describes the *concepts* needed in the resource metadata. These concepts may be instantiated in a variety of standard forms, e.g. XML, UCD tags, or FITS keywords, and with a variety of mechanisms, such as Topic Maps, OWL, or RDBMSs.

2.3.2 IVOA Metadata Registry Interface

IVOA has developed the standard interfaces [5] that enable interoperable registries. These interfaces are based in large part on a Web Service definition in the form of a WSDL document. Through these interfaces, registry builders have a common way of sharing resource descriptions with users, applications, and other registries. Client applications can be built according to this specification and be able to discover and retrieve descriptions from any compliant registry.

A searchable registry is one that allows users and client applications to search for resource records using selection criteria against the metadata contained in the records. A searchable registry gathers its descriptions from across the network through a process called *harvesting*. A publishing registry is one that simply exposes its resource descriptions to the VO environment in a way that allows those descriptions to be harvested. A full registry is one that attempts to contain records of all resources known to the VO. A local registry, on the other hand, contains only a subset of known resources.

The IVOA Registry Interface consists of three query operations:

- Search searches the Registry in order to obtain the VO resources.
- KeywordSearch is a helper query based on a set of key words.
- GetRegistries is another helper query to obtain Registry VO resources.

and six harvesting operations, which support resource harvesting in accordance with the OAI-PMH definition.

2.4 VOTable Format Definition

The VOTable format [13] is an XML standard for the interchange of data represented as a set of tables. A table is an unordered set of rows, each of a uniform format, as specified in 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 has built-in features for big-data and Grid computing. It allows metadata and data to be stored separately, with the remote data linked. Due to that it is possible to send metadata-rich pointers to data tables in place of the tables themselves. The overall VOTable document structure is described and controlled by its XML Schema referenced at its top. Data Access Layer

2.4.1 DAL Architecture

The task of the IVOA DAL working group is to define and formulate standards for uniform access to VO data that may have heterogeneous representations by different data providers. Architecturally DAL (http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDAL) consists of a family of data access services that provide access to both data and computation – access to VO resources. Client data analysis software will use these services to access data via the VO framework; data providers will implement these services to publish data to the VO. Principal data types within the scope of the DAL and mapping of data types to access services (e.g., image, table, spectrum, time series, etc.) are to be defined. Each service supports a specific data model and is implemented through the respective data access protocol. Access protocols form a middleware between the VO resources and client data analysis programs. Thus distributed multiwavelength data access and analysis is planned to be developed.

Current DAL services include:

- Cone search (access to astronomical catalogs; simple query based on position, search radius; returns a VOTable containing results);
- Simple Image Access (SIA) (uniform access to image archives; atlas and pointed image archives; image cutouts, image mosaics; image is returned as a FITS file or graphics file);
- Simple Spectral Access (SSA) (access to 1D spectra and SEDs; spectra is returned as ASCII, VOTable, FITS).

2.4.2 Simple Image Access Protocol Specification

This specification [7] defines a protocol for retrieving image data from a variety of astronomical image repositories through a uniform interface. A query defining a rectangular region on the sky is used to query for candidate images. The service returns a list of candidate images formatted as a VOTable. For each candidate image an access reference URL may be used to retrieve the image. Images may be returned in a variety of formats including FITS and various graphics formats. Referenced images are often computed on the fly, e.g., as cutouts from larger images. Data collections are often distributed, and the client may query multiple image services simultaneously, e.g., to gather data from multiple wavelength regimes or surveys to analyze a single region on the sky.

This specification is based primarily on two documents. The first document, "Simple Image Retrieval: Interface Concepts and Issues", describes a longer term view of how simple image access can fit into a more general framework for image access in the VO. The URL-based implementation is intended to be consistent with the concepts discussed in this document. A prototype SOAP/WSDL based Web Services implementation is also planned. The second document, the "Simple Cone Search specification" provides a means to query catalogs via HTTP with a uniform interface. The Simple Image Access interface (SIA) follows a similar to Cone Search approach.

The image data model assumed is minimal at this point. An image should be a calibrated object frame imaging some region of the sky. Only two dimensional images are fully supported within the interface at this time. Images can be returned as either FITS files or as graphics images. Ultimately, VO data models will provide a means to describe more complex data objects within the VO than be directly addressed by the SIM prototype.

2.4.3 Simple Spectral Access Specification

The goal of the Simple Spectral Access (SSA) specification [8] is to define a uniform interface to spectral data including spectral energy distributions (SEDs), 1D spectra, and time series data. In contrast to 2D images, spectra are stored in a wide variety of formats and there is no widely used standard in astronomy for representing spectral data.

The data model for spectral energy distributions defines a set of spectra or time series, some of which may have only one or few data points (photometry) and each of which may have different contextual metadata like aperture, position, etc. A *SED* object has a number of global attributes indicating the number of SED segments and *curation* information. Each *segment* has a *frame*, *coverage*, *curation* and *data identifier* object. The *frame* object is a simplified instance of the space-time coordinate system object The *coverage* object holds info about the observed region on the sky, the time range and spectral range. The *time coordinate* contains elapsed times relative to a reference time. The *spectral coordinate* can be expressed as a wavelength, frequency or energy plus velocity.

The purpose of a spectrum query is to determine the availability and characterization of data satisfying the constraints. The result is encoded as a VOTable. Queries can be restricted to certain types of data using the keywords *findSED*, *findSpectrum*, *findTimeSeries*. Technically based on SOAP/HTTP, an SQL query is generated. The format of the data returned in the retrieval mode could be a VOTable, FITS, native XML, a graphic file or some *foreign* format used by a data provider.

2.5 IVOA Query Language

2.5.1 IVOA SkyNode Interface

The SkyNode Interface describes the minimum required interface to participate in the IVOA as a queryable VONode as well as requirements to be a Full OpenSkyNode, part of the OpenSkyQuery Portal. OpenSkyQuery opens up the SkyQuery [9] protocol to enable other databases and servers to become “Full SkyNodes”. It should be noted that the SkyNode Interface is also related to Data Access Layer WG of the IVOA.

The Astronomical Data Query Language (ADQL) is considered as an XML document format for transported queries to IVOA SkyNodes. Different SkyNodes may not support all features of the Language. Hence ADQL would be passed from the SkyQuery Portal to the SkyNodes or it may come directly from a client or the VOQL portal. All nodes and the portals should be accessible via SOAP services. Additionally for the Open SkyQuery Portal some form of string based query like the current SkyQL would be accepted. A parser would easily convert this to ADQL, i.e. SkyQL would have the same semantics as ADQL but the syntax would be an SQL like string rather than XML.

Basic SkyNode is the minimum IVOA SkyNode Interface – this is useful in itself as it allows one to send queries to a system using ADQL. This is also just one step up from cone search. A matrix has been used as any feature on their own may be useful, i.e. a node which can do XMATCH (spatial matching) is already useful even if it may not participate in the portal because it lacks other features.

2.5.2 Astronomical Data Query Language (ADQL)

ADQL [2] is based on a subset of SQL plus region with, as a minimum support, for circle (Cone Search). ADQL is designed to be the request format of the OpenSkyQuery protocol. The OpenSkyQuery protocol drives a data service that allows querying of a relational database or a federation of databases. In this case, the request is written in a specific XML representation of ADQL. ADQL has two forms:

- *ADQL/x*: An XML document conforming to the XSD;
- *ADQL/s*: A String form based on SQL92 (the BNF exactly defines the form of SQL92) and conforming to the ADQL grammar. Some non standard extensions are added to support distributed astronomical queries.

The XML expression of ADQL (ADQL/x) is recommended in the Virtual Observatories for communications between portals and data servers. The string version of ADQL (ADQL/s) is more suitable for human to understand the queries.

2.5.3 VO Query Language

The Virtual Observatory Query Language (VOQL) [4] is an ambitious language at a higher level than ADQL. A VOQL portal would take VOQL programs. This would need all the work of the SkyQuery portal and more to make it function. There are 3 layers of VOQL:

- VOQL1 WebServices : ADQL and VOTABLE to exchange information between machines;
- VOQL2 Federation : SQL-like query language and federation system, i.e. combination of SkyQuery , JVOQL and VO standards;
- VOQL3 SkyXQuery: future XML-based query language.

The highest level of VOQL is a semantics-based language that allows astronomers to build queries in the language of astronomy rather than the language of databases. Efforts with an ontology of units allows queries expressed in one unit to engage resources expressed in another unit. Similarly astronomical coordinates can be fungible, so that a query in equatorial coordinates can return a resource expressed in galactic coordinates – but in the correct part of the sky. A similar approach allows federation of spectral data that uses different spectral coordinates.

This level of semantics, describing the structure of astronomical datasets, interacts with the astronomical semantics provided by the UCD schema to quantify use of astronomical knowledge. For example, a data model to define spectra may specify that a spectrum has an array of data representing an observable quantity and an array of values representing the spectral coordinate. The UCDs associated with an instance of this data model will specify whether that particular spectrum has an observable of flux or surface brightness, and a spectral coordinate of frequency or wavelength. A data model may also represent a higher level resource such as a compute service, in which the input parameters required by a particular class of service such as source detection programs are defined. Again, the values of some data model metadata may be UCDs which describe what kind of parameters are to be returned by the source detection.

3 Conclusion

Main intention of this publication is to attract attention of the DL and e-Science research community to the diversity and complexity of domain descriptions in natural sciences. Scientists have spent centuries to reach well-defined structures, concepts and theories in various science domains. Specificity of material systems, complexity of theories, variability of instruments and methods of research in each domain explain the diversity required. We have chosen here the domain of astronomy where Virtual Observatory development is very intensive around the world. The IVOA standards shown constitute only a very thin layer of the overall standards that will be required to define the knowledge-based collective memory in astronomy. It is supposed that such eventual system of standards will acquire a form of modular organization where each module corresponds to a subdomain related to a specific class of astrophysical problems (very roughly, such as cosmology, galaxy formation and development, star formation and evolution, sun and planet systems, etc.). The diversity demonstrated is a challenge for the IT people and a warning that should prevent us on light weighted promises to create rapidly uniform and integrated definition of a collective memory for the science as a whole (or even for its separate domains). Serious research, new methods and technologies will be required to overcome the existing obstacles. Some of them are investigated during the creation of the Russian Virtual Observatory [1], the lowest part of which is based on the existing IVOA standards.

References

[1] Briukhov D.O., Kalinichenko L.A., Zakharov V.N., Panchuk V.E., Vitkovsky V.V., Zhelenkova O.P. Information Infrastructure of the Russian Virtual Observatory (RVO). M.:IPI RAN, 2005. – 150 p.

[2] IVOA Astronomical Data Query Language. Version 0.7.4, IVOA Working Draft, 2004-05-17

[3] IVOA Data Model ,

http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDataModel
[4] IVOA Query Language, http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL
[5] IVOA Resource Registry, http://www.ivoa.net/twiki/bin/view/IVOA/IvoaResReg
[6] Resource Metadata for the Virtual Observatory. Version 1.01, IVOA Recommendation 2004 April 26
[7] Simple Image Access Specification. Version 1.00, IVOA Working Draft 24 May 2004
[8] Simple Spectral Access for ISO data. Version 1.00, IVOA Note 21 May 2004
[9] SkyQuery, http://www.skyquery.net
[10] UCD (Unified Content Descriptor). Version 1.9.9, IVOA Working Draft, 2003-10-15
[11] UCD (Unified Content Descriptor) - moving to UCD1+. Version 1.03, IVOA Working Draft 21 May 2004
[12] Virtual Observatory Architecture Overview. Version 1.0, IVOA Note 15 June 2004
[13] VOTable Definition. Version 1.1, IVOA Proposed Recommendation 08 June 2004

♣ This work was partially supported by the RFBR grant 05-07-90413