*Bull. Spec. Astrophys. Obs., 2005, 58, 52-62* 

# The SAO RAS observation archive state and prospects of development

**O.P. Zhelenkova, V.V. Vitkovsky, G.A. Mal'kova, T.A. Plyaskina, V.S. Shergin**

Special Astrophysical Observatory of the Russian AS, Nizhnij Arkhyz 369167, Russia

*Received July 15, 2004; accepted July 23, 2004.*

The paper is concerned with the SAO RAS observation archive. By the example of local archives being part of it, an analysis of parameters describing observations is made. The strategy of step-by-step inclusion of the archive into the virtual observatory (VO) environment on the basis of IVOA (International Virtual Observatory Alliance) protocols, interfaces and formats is considered.

**Key words:** astronomical data bases – observation archives

## 1. Introduction

Numerous astronomical projects, telescopes, space missions provide a huge amount of information for the study of the Universe. We can essentially extend the knowledge of physical properties of celestial objects by studying them in different spectral ranges and investigating long observation series for variability search. This needs all available data. In the free web access there are archives containing observation of millions of objects. Such examples are archives of the Hubble and Chandra space telescopes, the Sloan Digital Sky Survey (SDSS), the Two Micron All Sky Survey (2MASS) and the Digital Sky Survey (DSS). The number and volumes of the catalogs, surveys and archives are so large that searching for data in them takes much time and it is impossible to fully copy the information to the user computer. In order to solve the problem of a unified data using, the astronomical community is actively developing a virtual observatory conception. More then one and a half tens of the national projects concerned with the virtual observatory creation are unified in the international alliance IVOA (International Virtual Observatory Alliance, http://www.ivoa.net).

Several working groups are functioning in the IVOA which are engaged in the development of the VO architecture (Williams et al. 2004), technical bases for the publication of astronomical resources, discovery and access to them and also protocols, interfaces, formats and a query language for exchange and work with data. The common astronomical software such as IRAF, IDL, MIDAS gives astronomers the possibility of creation on their basis a set of tools for data processing and analysis. The core of the program systems consists of a set of commands and a command language that enables more complex procedures to be constructed in consideration of specific features of data processing. The methodology of the VO architecture is based on the similar principles. There is a set of simple basic functions realized as web services, and a user can build up from them as from bricks new services for his own purposes. The following functions compose the VO core: publication and discovery of available resources, access to the data (images, catalogs, spectra) by means of simple requests. Even now, providers of large volume data give open web access to them with the aid of the VO protocols. The software is being developed which will allow data providers to publish information in the Internet in the form compliant with the VO standards.

SAO RAS possesses the two largest Russian telescopes — BTA and RATAN-600 designed for fulfillment of investigation of priority astrophysical problems. For more than 20 years a digital observation archive is conducted at the observatory. In the late 1980s a conception of the Astronomical Data Bank (ADB) of the SAO RAS was developed (Kononov et al. 1990), which unified observational data, astronomical catalogs, digital surveys, technical documents and papers into one informational system. The telescopes of the observatory are in shared use, which supposes intensive information exchange and demands organization of operative access to them. The realization of web access to diverse digital collections and the unification of information resources of the observatory are problems of the Integrated Bank of Data (IBD), which expounds the ABD conception in view of the achievements of informational and computer technologies.

To include the IBD into the VO environment, we can suggest a number of sequential steps using the specifications of the services, protocols and standards proposed by the IVOA. Initially this approach is applied to the data of the general observatory archive, and then it can be extended to the other collections (catalogs, surveys, etc.) available and newly included into the IBD.

## 2. On the Virtual Observatory architecture

To turn to the discussion of the VO architecture, let us consider what technological methods are used in the Internet communication environment for data computing. Recall that WWW is a global net by the use of which millions of computers exchange information using standard protocols. The simplest and most widespread methods of exchange is the combination of HTTP and HTML. The HyperText Markup Language is a specialized language intended for the description of hypertext documents and the HyperText Transfer Protocol is a protocol for hypertexts transmission. The work in the net consists in the interaction between two computers with the help of program tools (web browser, web server) in order to transfer information by means of the protocols and to represent the hypertext document in the graphical or text form. For information delivery by HTTP protocol a model "client – server" or the principle "request – response" is used. It means that the information is centrally held (on the server) and accessed via client program at the user request. For this reason, the HTTP protocol is used to send information from the client to the server and vice versa.

Web services are more sophisticated interaction level in the net. There are applications which can be published, discovered and run in the Internet. They exchange messages with nodes (servers) that allows procedures at remote computers to be executed. The messages are created at XML (eXtensible Markup Language (Extensible ... 2004). In comparison with HTML this is more universal language for the creation of hypertext documents which is used for communication between any web applications. It can be illustrated by the services on the basis of the SOAP protocol (Simple Object Access Protocol). This protocol at its lower level uses HTTP as a transport means. With its help the textual information is retrieved in which the data structure, the request, order of use and execution of the service are described by the SOAP rules using XML. The client program fetches the necessary SOAP service and runs it. Since web services run in the net with different program systems and hardware, then the protocols used for data transfer and procedure run must be independent of execution environment. The SOAP possesses these properties. Web nodes supporting this protocol provide web-services which programmatically interact between themselves and operate without human intervention.

Computer facilities and data storage systems at enterprises and institutions are rarely fully loaded. From the statistics UNIX systems nearly 90% of time are inactive, the personal computers are inactive to 95%. Without acquisition of additional computers the efficiency of existing facilities can be increased by "learning" the net connected computers to work together. This is exactly the idea of distributed computing. Metacomputing or distributed computing consists in the computer resource use by means of communication environment. Such a level of net interactions between computers is implemented with the aid of actively developing Grid technology (Chetty, Buyaa 2002). The Grid is a method of using distributed processor capacities and storage systems allowing the idle computer resources to get busy. This technology is realized with the aid of a set of standardized services which provide reliable shared work of geographically distributed resources uniting computers, clusters, information storages, nets, scientific program tools etc. It is not of the user's concern where the resources used by him are located. The most important component of the Grid infrastructure is the middleware which controls jobs, ensures secure connections and data staging, replicates data from one geographically remote node to another and syncs their copies. The next generation of the Internet will enable something more: from your computer you can make a net of computers work for yourself. The universal language for such interaction is assigned to the Open Grid Service Architecture (OGSA) (Foster et al. 2002).

After this short review of web technologies we will turn to the VO architecture proposed by IVOA (Williams et al. 2004). The main goals of the VO designers are discovery of new class of objects and pattern recognitions, distributed computing implementation, publication of available information resources and analysis results in the web and also provision of collaborative research work by the same scientific problems. They are realized with portals (web-sites with specialized tuning to their own audience performing the role of the starting point or input when working with information), unified user interfaces and application specific software. As examples we can give the already implemented program packages and systems: VOPlot, Mirage, Topcat and user interfaces and portals: SkyQuery, Aladin, Oasis. The next VO architecture level is interfaces with heterogeneous data sources. Interaction between them is executed with web-services based on HTTP, SOAP and Grid standards of the W3C consortium. These services are chosen by the IVOA as the main VO technologies. HTTP, SOAP and Grid provide different degree of efficiency and functionality in the work with information.

The HTTP services are the simplest and most widespread means of information communication, but they do not ensure the necessary for the VO flexibility and scalability on an information resource request. The SOAP extends capabilities of HTTP which enables unattended resource interchanges to be organized. The Grid is used for controlled and authorized bulk volume data delivery and distributed computing. The VO services are divided into three wide classes: register services for discovery and publication of resources, data services (request and access to data) and computing services (data computing and merging). Let us dwell on each of them in more detail.

The discovery of data is performed through the VO registers (Plante et al. 2004), which contain the resource descriptions and are an important component in effective searching of information in distributed environment. Data and services are considered as resources. Other objects, for example, organizations, projects, software can also be regarded as resources. The register is simply a resource description represented in the form of structured metadata to provide automatic search and request processing. A request to the resource can be performed by the following parameters: type (catalog, archive of images, and education resource), sky region, time, wavelength and so on. The IVOA model contains three types of registers. The searching register is implemented for user applications and contains all descriptions of the resource available in the VO. For this type of register a special process of data harvesting from a great number of registers is executed. Publication registers differ from the search ones by the fact that they do not support searching. They simply display resource descriptions in the Internet and can repeatedly fulfill the search registers by means of sync process. Application-specific search registers belong to the third type. They do not contain all VO information. The registers are specialized in a particular type of resources or scientific themes, for example, associated with the investigation of supernovae. They can perform harvesting over other registers filling up themselves.

Data services provide request and access to information. The most popular request is search for information about a selected sky area (the target and the area size is indicated) and obtaining images, spectra or a list of objects falling within this region. Information is requested by means of the Simple Image Access Protocol (SIAP) (Tody and Plante 2004) which has a particular format of access to resources supporting the service, and obtains the image of the given sky region. This is a simple service and it allows access to one data source. Data services, besides standard graphical formats (gif, jpeg), operate with two astronomical formats. One of them is the Flexible Image Transport System (FITS, Wells et al. 1981), which is an astronomical standard for storage and exchange of data. The second one is the VOTable format (Ochsenbein et al. 2004) used for the output of data obtained at request. The VOTable corresponds to the XML standard (Extensible... 2004) and is dedicated to description and representation of astronomical tables.

To designate one and the same physical quantities and parameters, astronomers use different names. In order to avoid ambiguity in interpreting the values, it is necessary to determine what exactly the different names denote. This is especially important in exchanging information between services. For determination of the semantic type of the quantity it is proposed to use a dictionary developed and controlled by IVOA which is named the Unified Content Descriptor (UCD) (Derriere et al. 2004). The fixed designations of physical quantities and their definitions are consolidated in the dictionary. Subsequently, the IVOA will define specifications of requests at a higher semantic level and UCD will be an important part in this definitions.

The IVOA is going to develop standards for complex services, for instance, for data federation (unified requests to several data sources), image processing and detection of objects, statistics analysis and complex data visualization. For instance, for the requests to a relational database or to several databases simultaneously an OpenSkyQuery protocol (IVOA SkyNode Interface ... 2004) is being developed. In this case the request is written in a language close to the Structured Query Language (SQL) generally used by database management systems (DBMS). The Astronomical Data Query Language (ADQL) (Yusida et al. 2004) is a simplified analog of the SQL. An ADQL operator is translated into XML representation. If there are several astronomical resources which provide the web service with such a protocol then ADQL query can run on them all at once.

The third group of services is related with data computing. We get a sky area image by a digital survey or an observation archive query. For this purpose, it may be necessary make a mosaic from several images, flatten background, rotate and add some images for co-locating, this is to perform a few simple operations. It is desirable that these operations

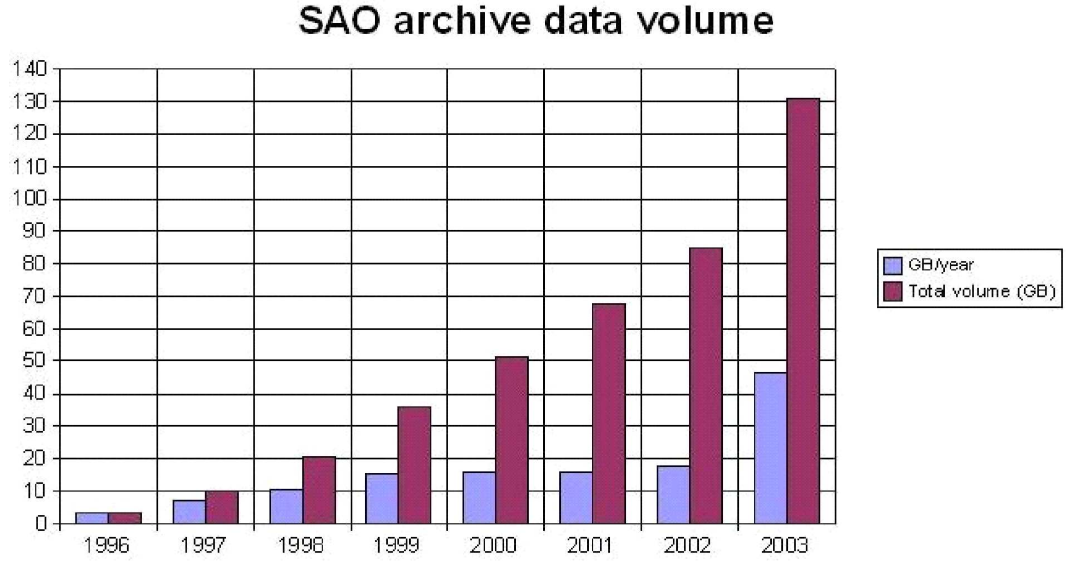


Figure 1: *Volume data change of the SAO RAS observation archive.*

are performed in unattended mode in the process of query answer formation as a consecutive computing sequence. Such sequences are called workflow. In the IVOA development the information processing software is used representing a set of distributed services linked to each other with workflow execution. The flow paradigm for working with distributed data consists in multiple using simple services for construction of complex applications. This components are isolated one from another by means of well specified protocols, namely, rules for job running, the structure for input and output data are defined. An analogous principle can be used for virtual data, that is, for such data that are created dynamically in case of necessity and kept for a certain time in the cache, which allows their reuse in the same requests.

## 3. Current state of the SAO RAS general archive

During its existence the SAO RAS archive has passed a few stages of data representations, types of data storage media, provided services. The stages are closely related with the computer technology progress and also with necessity for international collaboration in astronomical scientific investigations.

Table 1 represents information about the volume of optical and radio data. Fig.1 shows the change of the archive data volume (gigabytes) from 1996 to 2003 and also the annual rate of data increase.

The stages of the observation archive existence may be considered as a chain: ***local archive* → *base archive* → *general archive* → *IVOA compliant archive***, which makes it possible to trace the process of unification of heterogeneous in file formats and data structures and also organize access to observation files.

A **local archive** is a digital collection of data obtained by one or a few similar observation methods. Since 1988 FITS-format for observation data description and storage has been used at the observatory (Vitkovskij et al. 1988). The FITS file consists of symbolic and binary data blocks. The blocks with symbolic information contains a set of pairs "keyword – its value" which describe the file parameters. The observations proper are binary data. The FITS keywords can be considered as metadata when describing the binary contents of the file. The FITS keyword sets were specified for the observatory observation data, and also structures of a binary part of a file were fixed and reflected in its header. At the observatory telescopes diverse instruments are used for an investigation of celestial objects. The consequence of these is the existence of observation methods distinguished by various parameters. Each method is linked with a certain computer-instrumental complex — an acquisition system. The digital data formats differ for different acquisition systems and have own keyword sets for observation parameters.

The **basic archive** is a collection of the local archives. Since 1994 the observational data obtained with the observatory instruments have been kept on optical disks. The physical unit of storage in the archive is an optical disk. The observer decides what refers to observations and should be kept in the archive. The administrator forms a disk image, can

Table 1: *The SAO RAS observational data archive (CD-disks, 15.12.04)*

| Archive | Number of disks | Average daily flow of data | Data volume | Number of files |
|---|---|---|---|---|
| Optics | 125+125 copies | 85.2MB | 165.4GB | 136589 |
| Radio | 7 | 4.3MB | 3.6GB | 45921 |
| Users' archives | 30 copies | | | |

add, when necessary, auxiliary information and writes archive data on an optical disk. In the end, on the disks are written: files with observations, service files used for the instrumental error corrections and calibrations, observation logs, accompanying information prepared by the observer, and also additional information text reference files used for disk identification, software related with the disk content, check sums. At the present time, the basic archive contains 16 local archives which differ by data formats, keywords and structures of data. In Tables 2 and 3 are presented principal characteristics of the local archives.

From the experience of work with such a collection, rules have been defined which are supposed to be followed in the work with the archive. Now it is necessary to take into account the possibility of addition of new digital collections, old observation data and observations obtained at other observatories. These are the rules:

- the archive is transparent for the user. It does not change formats and parameters of stored data; the formats of the output data retrieved at requests are the same as the input data formats; additionally a possibility is available to retrieve the data in the FITS format;
- the archive keeps raw data written on optical disks; access is provided to everything that is written on disks;
- the logical unit of storage in the archive is an observation;
- access to the data is permitted on the basis of the SAO RAS observation archive regulation (http://www.sao.ru/Doc-k8/Science/arch_regul.html).

The **general archive** is the basic archive with web access to data. To realize such access to local archives, different types of observation files written on CD disks are analyzed. The main problem to be overcome in the archive federation is the difference of formats, parameters and data structures. For this purpose, an information search system (ISS) based on a database management system (DBMS) is used (Vitkovskij et al. 2000).

The basic source of information about observation parameters is the file header. For the description of observation in different acquisition systems use from 0 to 75 parameters in optics and to 289 in the radio range. The values of these parameters are received from the control systems of the telescope and from the acquisition system. Part of the parameters comes to the file header automatically, part is entered interactively by the observer. Since the acquisition systems are being improved, then the data formats and parameters change. This is why, each archive includes more than one version of format.

The sources for getting information about observation parameters are full names of files obtained when executing a system command ls, keywords (descriptors) from the file header, night logs, observation schedule. The information can be repeatedly duplicated, but sometimes, only the principal investigator (PI) can give or specify it. For some observations only the full name of file is given. From this name, which contains the path from the root catalog of the optical disk, one can determine the observation date, the compression method, the image type, the filter, the format. The system command ls gives information about the file size, time of file recording (copying on the disk).

Several popular types of requests for the archive data can be specified from interviews of users. Hereafter we will mention these types as standard requests. Here belong requests the date of observation, instruments, types of files, coordinates of the observed field/object, name of the astronomical object, observation program, PI's and observers. To make standard requests, the file header must have necessary parameters. For this purposes, the local archives were checked for the adequacy of keywords in the header, the necessity for the involvement of other sources and what exactly, rules (relations between keywords) by which the search for data corresponding to the requests is possible.

Due to errors of observational file headers, insufficient number of keywords, absence of necessary parameters, it is impossible to organize standard requests to all archive files without exception, but for one type of request – by the observation date. Other types of requests can be executed only to a part of local archives.

Now turn to the procedure of CD disk archiving.

Table 2: *Local archives included into the general observation archive of SAO RAS*

| **Local archive** | **Telescope** | **Number of CD** | **Date range** |
|---|---|---|---|
| LYNX | 6m | 7 | 1996.02 - 2001.12 |
| NES | 6m | 23 | 2001.07 - 2004.05 |
| PFES | 6m | 1 | 1999.12 - 2001.01 |
| MPFS | 6m | 17 | 1996.08 - 2004.02 |
| IFP | 6m | 2 | 1997.05 - 2000.03 |
| MOFS | 6m | 1 | 1997.03 - 2001.08 |
| SCORPIO | 6m | 18 | 2000.09 - 2004.04 |
| UAGS | 6m | 8 | 1994.11-2003.09 |
| CCD | 6m | 4 | 1996.02 - 2000.04 |
| SP124 | 6m | 2 | 1996.02 - 1999.05 |
| MSS | 6m | 1 | 1996.05 - 2001.01 |
| ZMCCD | z1000 | 18 | 1996.12 - 2004.04 |
| ZUAGS | z1000 | 3 | 1998.04-2000.09 |
| CEGS | z1000 | 1 | 1997.03 - 2000.12 |
| Z600 | z600 | 8 | 1996.01 - 2001.06 |
| RATAN | RATAN600 | 7 | 1996.06 - 1999.01 |

Table 3: *Characteristics of the local archive*

| **Local archive** | **Volume (Mb)** | **Number of files** | **Mean size of file (kb)** | **Number of nights** | **Daily average flow of data (Mb)** |
|---|---|---|---|---|---|
| LYNX | 8790 | 5397 | 1700 | 291 | 30 |
| NES | 28334 | 4529 | 6400 | 137 | 207 |
| PFES | 1248 | 771 | 1700 | 36 | 35 |
| MPFS | 20830 | 10819 | 2000 | 186 | 112 |
| IFP | 2180 | 4502 | 500 | 28 | 78 |
| MOFS | 1286 | 1059 | 1200 | 26 | 49 |
| SCORPIO | 21730 | 21274 | 1000 | 189 | 115 |
| UAGS | 8678 | 19809 | 450 | 397 | 22 |
| CCD | 4988 | 6012 | 800 | 127 | 39 |
| SP124 | 2338 | 8162 | 300 | 173 | 14 |
| MSS | 1284 | 2986 | 400 | 153 | 8 |
| ZMCCD | 22958 | 21737 | 1100 | 345 | 67 |
| ZUAGS | 364 | 3855 | 100 | 107 | 3 |
| CEGS | 462 | 220 | 2200 | 18 | 26 |
| Z600 | 10134 | 15968 | 650 | 415 | 24 |
| RATAN | 3554 | 45921 | 80 | 861 | 4 |

First, the CD disk is copied on the hard disk of the archive server and examined by the administrator for peculiarities. It is desirable that data on the disk will be presented in a certain standard manner. For this, it is necessary that observation obtained during one night should be in the catalogs whose names contain the date. One observation is considered to be a logical unit of storage in the ISS, that is why, several observations archived in one file are unpacked. The operations which are performed with the disk to bring it to the standard form are written in the command file.

The administrator uploads in the archive system the data on the new disk, supports the database server, controls conditions of optical and hard disks with observations. The administrator needs information about each optical disk: disk size, number of files and catalogs, full file names, checksums. Archive users have access to observation data. Besides, the administrator has more access to data of other types (software, reference and service information).

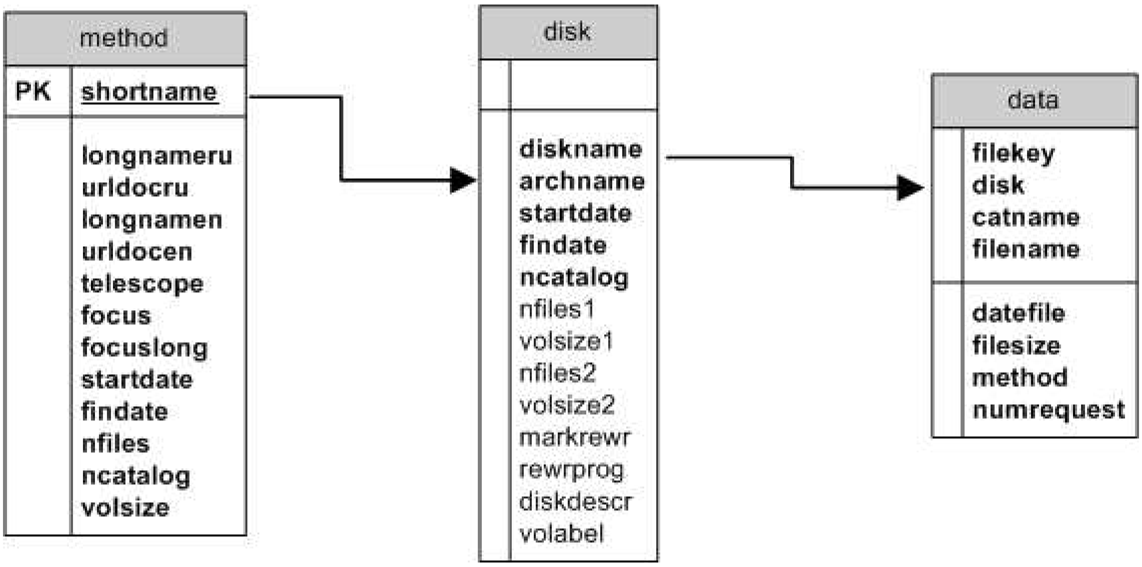


Figure 2: *The scheme of the information system tables for the date request.*

A dedicated server on which the information search system based on the Oracle DBMS and file warehouse are located is used for the observation archive. The structure of tables for the date request was designed. This type of request can be performed for all local archives. The loading of tables is accomplished by Perl scripts. These scripts analyze flat text files got with the aid of the command 'ls' applied to optical disks. The user interface for on-line access to the observations is done on the basis of CGI (Common Gateway Interface), DBD (DataBase Driver) and DBI (DataBase Interface) specifications. The tables are uploaded on an on-going basis. The scheme of the tables is presented in Fig. 2.

The services given to the users by the general archive are as follows: web-interface, http-retrieval of the selected data, quick-look FITS files with observations (grey scale picture in the jpeg-format is generated), getting data on the observation set from the archive schedules (via the name of the observation program). The information search system functions are supposed to be extended by adding new types of requests to the general archive. The developed system unified heterogeneous collections, which are the local archives, and made it possible to refer to them through the common web-interface.

## 4. The SAO RAS observation archive as the VO-compliant resource

The astronomical community extensively uses new computer science technologies for the creation of an environment which would unite all available data into a common warehouse. Since the moment the projects connected with the VO appeared, specifications of protocols, formats and types of requests used for creation of software allowing operations with the astronomical data in the Internet have been developed and specified. These standards are already used in a number of astronomical data centers for access to observation archives, digital surveys, catalogs. It is noted in the quarterly NVO report that several thousands of astronomical resources are available in the VO registers now (Building ... 2004). In three – four years the publication of data in the Internet with the use of IVOA standards will become an ordinary and necessary procedure. It is important that the SAO RAS archive would be included in the informational structure of the virtual observatory. It is necessary for this to choose from the specifications those ways and methods of inclusion which are suitable for our collection.

It will be recalled that the informational structure or, in other words, environment, infrastructure of the virtual observatory holds computers and data warehouses connected by WWW and includes a few interpenetrating tiers of software based on the common and specialized specifications and standards for network data transmission, database interaction as well as user interfaces, visualization and analysis of data. For inclusion of astronomical data into the structure a few tasks need to be performed. First, it is necessary to obtain and present the information in the digital form then put a dataset in the computer, which can be done by different methods: from a simple — recording files on a hard disk, to a complicated, but more flexible one in functionality and reliability — creation of a database. After the creation of web-interface to data the community should be informed about it, so that the web-services requesting data could find the resource in unattended mode. For this purpose, the digital collection is put into the VO register to which services operating with user interfaces apply.

In the NVO, the process of registration of astronomical resources is being tested. Two registers are available in which astronomical resources are published, in the California Institute of Technology (Caltech) and the National Center for Supercomputer Applications (NCSA) (Plante et al. 2004). On each of them data source can be registered by filling the web-form. Input resource descriptions are stored as XML documents. The search registers use for gathering information a special process (harvesting) based on the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) (The Open Archives ... 2002).

If the owner of data has several static, rarely changing collections, the simplest way to control the description of his own resources is to use the Caltech or NCSA sites with already existing registers. When the provider has several dozen resources of data, which change and better control of the descriptions is needed, then a register of his own is created on his site. This method of registration is more suitable for the observatory archive since it allows one to work with the included resources at ones own discretion. It is also suitable for the publication of digital collections, included in the IBD, which differ in the structure of data from the local archives, namely, for catalogs, surveys.

So far there has been no finished program solutions for the creation of ones own local register on the site of any IVOA member, but they will appear soon. For this reason, for the local archives included in the SAO RAS general archive, it is supposed, first, to perform registration in one of the above-mentioned sites and then transfer their descriptions to the local register of SAO RAS.

For the integration of information resources, it is important to find methods of addition of new collections to the environment, so that it can be possible to retrieve the requested data through the unified interface. When the collection is entered into the register, information appears that the resource is available, but one cannot work with the data themselves yet. To search for and retrieve the data from this collection, it is necessary to realize web service. Such service is created on the basis of the corresponding specification. Specifications of the following protocols have been developed: ConeSearch — request and retrieval of data from the catalog for the objects falling within the given sky region, SIA — obtaining an image of the target sky area, for spectra — Simple Spectrum Access (SSA) (Dolensky, Tody 2004). The register contains the information on what type of services are provided for the resource, so that the web-services, which execute search for all published collections, can work with data by descriptions in unattended mode. For instance, on the Caltech and NCSA sites, where the VO registers are located, there is a list of digital collections, from which one can thus retrieve the requested images. That is why, apart from the registration of the observatory archive, yet it is necessary to realize analogous web services for it. It will permit the archive to become a part of the integrated astronomical resources.

Images of sky regions and spectra of celestial objects are stored in the local archives. If one applies the VO technologies, then for observation data to be selected from the general archive, web service are required, which permit retrieval images of target sky regions or spectra of objects falling there from the local archives. For the creation of these services, the protocols SIA and SSA are used. A protocol is a set of rules for the client and server with data interchange. Let us dwell at length on the SIA protocol. As requested by a client, the SIA service must transmit the image of the sky region of the defined size. In an ideal case it gives the astronomer a certain region of the virtual sky, which in reality may consist of several digital images covering this region, and the user should not worry about joining the boundaries of individual frames and their calibration. In obtaining data, the following operations must be performed: request for the image, preparation of the image for transmission and the transmission itself. This protocol is subdivided in four more categories by the type of images and operations with them. Briefly about each type of service.

***Image Cutout*** service is intended for cutting out regions from large digital images, usually from digital sky surveys. The user received back the required region which may be composed from the mosaic of frames. No additional operations is performed with obtained image, for instance, rescaling, pixel resizing, etc. Initial data do not change.

***Image Mosaicing*** service resembles the previous one, but additionally performs defined by user data processing: image rescale, pixel resample, reduction to another type of projection. There is a possibility to generate pixels of the image from data obtained in other spectral regions and perform superposition of the images. The initial data change in this case, and the service operates slower than Image Cutout.

***Atlas Image Archive*** provides access to already processed images of different type of surveys. No extractions of portions from images is carried out. The user obtains the image back from the survey according to the conditions of the requests.

***Pointed Image Archive*** service provides access to image collections of small sky regions the most frequent aim in which is a certain particular object. These type of service is used for observation archives.

From the type of the service mentioned above the Pointed Image Archive suits for the SAO RAS archive. Further, when considering the specifications in more detail, we dwell on the parameters and rules for this type of the service and also on the sequence of actions which it uses.

The first action is the request of the image. The size, scale, type of projection can be used as additional parameters for specification of the request. The service must return the URL-links to the images which are the most suitable to the conditions of the request. Input of data is executed as GET request of the HTTP protocol. It looks as follows:

**http:// <server-address> / <path-to-service-program>?[extra-GET-arguments>&[...]]**

The service must transmit two parameters, which define the coordinates of the center and the size of the rectangular sky region. The coordinates of the center are given in degrees in the ICRS (International Celestial Reference System) coordinate system, which corresponds to the catalog FK5 for the epoch 2000.0. The size of the area are also assigned in degrees. The parameter INTERSECT, which defines how the selected images must coincide with the requested sky region, can be passed in the request. This parameter is omitted for the SAO RAS archive since it is supposed by default that INTERSECT=OVERLAY, that is the image satisfies the request condition if it partially overlaps the demanded area. In the request line for the second type of service additional parameters, which determine the desired size, scale orientation and projection of an image are used. Other SIA types must not give error message if they are present in the request line. The service must support the parameter FORMAT which marks the formats of the obtained image. They may be the following: FITS, html, jpeg, png.

The request response must be in the form of a table in the VOTable format. All the images satisfying the request condition are transmitted in it. It is defined in the specification what elements of the VOTable are obligatory. To describe each image in the table row is used with a set of parameters in the form of UCD descriptors. Information identifying the image, its coordinates and size, spectral range, actions performed with the image is transmitted. The coordinates of image are represented in a simplified version FITS WCS (World Coordinate System) (Greisen, Calabretta 2002). The service must give error messages.

The next step is connected with preparation of image. The server takes out data from the store and places them into temporary memory (cache) located on another computer, or produces dynamical generation of an image (scaling, rotation, resampling and so on). This process is essential in bulk data retrieving. When transmitting data from the SAO RAS archive, this step may not be considered. It will be performed by standard means of the operation system. The last stage is the transmission of data. In this version of the specification the URL link to each image is transmitted to the client, from which he can get data using the mechanisms of the Internet.

These rules should be followed in realizing the web service for the observatory archive.

To accomplish the SIA request, the following parameters of the observation should be included into the structure of the ISS tables: the observation date, type of file, telescope, device, detector, acquisition system, coordinates of object, frame size. Fig. 3 presents the scheme of tables for getting a sky region from the archive. The coordinate request is applicable to part of the collections, because the data have the set of coordinate parameters not in every local archive. For the local archives, there are variants in the format and name of the parameters which represent one and the same physical quality. For instance, the observation date in different digital collections can be obtained from the values of the following keywords: DATE, DATE-OBS, "Date of observation", OBS-DATE. This situation with parameters complicates the procedure of uploading the tables. The list of images, satisfying the request, outputs in a table in the columns of which the parameters of each file are written. These parameters are UCD-descriptors. For this reason, for the local archives it is necessary to establish correspondence between the observation parameters and the elements of the UCD-dictionary. This is accomplished through the local dictionary where the inner names of observation parameters and the corresponding UCD-elements are matched.

To include a new digital collection into the general archive of the observatory, the following procedures should be performed. The collection is published in the observatory register. The collection is described by the IVOA rules of registration for information resource. Then the correspondence between new parameters, which are used in the given collection for description of observations, and keywords of local dictionary is found, and when necessary, additions are made. The web-service to the added collection is created on the base of chosen protocol specification.

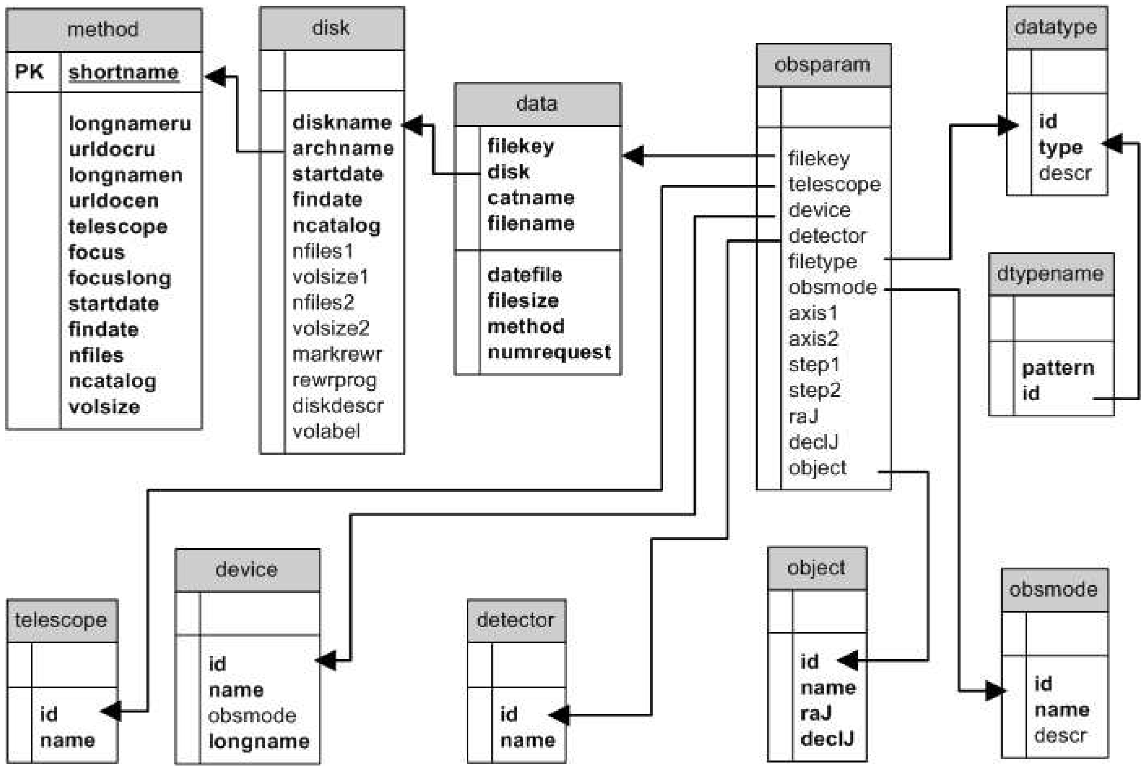


Figure 3: *Scheme of the table for coordinate request.*

## 5. Conclusions

The SAO RAS general archive includes 16 digital collections. Raw observation data are stored in it on the optical disks. The total volume of the archive is about 170GB (compressed data) with an average daily data rate of 90MB (at the date 15.12.2004). The information about the current state of the archive is available at the observatory site (http://www.sao.ru/oasis/archive/docs/boacs.html). The volumes of the local archives of the following instruments of the 6 m telescope: NES, SCORPIO, MPFS, and the 1 m telescope ZMCCD increase the most extensively. In the informatics department, step-by-step work is being done on the unification of heterogeneous digital collections for carrying out requests to them. For this purposes, the structure of the ISS tables has been developed. Storage, upload and requests to them are provided with the aid of DBMS. Programs for loading the tables and web-interface of data access have been developed. At the first stage, request to the local archives for the observation date with previewing of the data and FITS file headers was realized. The development of the scheme and table loading, software for coordinate request is in progress.

The astronomical community actively uses the Internet and new information technologies for retrieval information from available astronomical resources. The IVOA coordinates and governs the work of the community in integration of heterogeneous data, development of standards and specifications. All this work is united by a common notion “virtual observatory”. Some time later, the procedure of publication of any astronomical resource (catalog, survey, archive, web service, site, observatory, etc.) in the VO register will be a standard action as well as the use of special services to work with it.

At the present time, it is necessary to consider and test the IVOA specifications for the federation of digital collections of the observatory. In particular from the available specifications for the general observation archive the following protocols — SIA, SSA, VOTable format, UCD are selected. The next step in the process of integration of the local archives is the use of these standards for the development and realization of web service access to the data.

**Acknowledgements.** This work is supported by the RFBR grant No.07-03-90032.

## References

Chetty M., Buyya R., 2002, CiSE, **4**, 61, http://www.gridbus.org/papers/WeavingGrid.pdf

Derriere S., Gray N., Mann R., Matrinez A.P., McDowell J., McGlynn T., Ochsenbein F., Osina P., Rixon G., Williams R., 2004, UCD (Unified Content Descriptor) – moving to UCD1+, IVOA Workig Draft 2004-04-26, http://www.ivoa.net/Documents/ UCD/WD-UCD-20040426.html

Dolensky M., Tody D., 2004, The Simple Spectrum Access Protocol, SPIE, 5493-47

Foster I., Kesselman C., Nick J.M., Tuecke S., 2002, The Physiology of the Grid. An Open Grid Services Architecture for Distributed System Integration, Draft 2.9, 31pp

Greisen E.W. and Calabretta M.R., 2002, A&A, **395**, 1061

Kononov V.K., Monosov M.L., Vitkovskij V.V, Lipovetskij V.A., 1990, Soobsch. SAO, **65**, 32

Ochsenbein F., Williams R., Davehall C., Durand D., Fernique P., Giaretta D., Hanish R., McGlynn T., Szalay A., Tailor M.B., Wicenec A., 2004, VOTable Format Definition, ver. 1.094, IVOA Proposed Recommendation 2004-06-04, http://cdsweb.-strasbg.fr/doc/VOTable/v1.09

Plante R., Green G., Hanish , McGlynn T., O'Mullane W., Williams R., Williamson R., 2004, ADASS XIII, ASP Conf. Series, **30**, 585

Tody D., Plante R., 2004, Simple Image Access Specification ver.1.0, IVOA WG Working Draft, http://www.ivoa.net/Documents/WD/SIA/sia -20040524.html

Wells, D. C., Greisen, E. W., and Harten, R. H., 1981, A&AS, **44**, 363

Vitkovskij V., Zhelenkova O., Kalinina N., Shergin V., Chernenkov V., Kiyashko T., Pischic L. and Kondratenko V., 2000, Baltic Astronomy, **9**, 578

Vitkovskij V.V., Zhelenkova O.P., Ryadchenko V.P., Shergin V.S., 1988, Soobsch. SAO, **59**, 60

Williams R., Hanisch B., Linde T., McDowell J., Moore R., Ochsenbein F., Ohishi M., Rixon G., Szalay A., Tody D., 2004, Virtual Observatory Architecture Overview, IVOA Note 2004-06-14, http://www.ivoa.net/twiki/bin/view/IVOA/ IvoaArchitecture

Ysuda N., Mizumoto Y., Ohishi M., O'Mullane W., Budavari T., Haridas V., Li N., Malik T., Szalay A., 2004, ADASS XIII, ASP Conf. Ser., **30**, 293

## Specifications and protocols

Bulding the Framework for the National Virtual Observatory. NSF Cooperative Agreement AST0122449, Quarterly Report, April-June 2004, 44 pp.

Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation 04 February 2004, http://www.w3.org/TR/2004/REC-xml-20040204

IVOA SkyNode Interface, ver.0.7.4, IVOA Working Draft 2004-05-17, http://www.ivoa.net/internal/IVOA/IvoaVOQL/ SkyNodeInterface-0.7.4.pdf

The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14, http://www.openarchives.org/OAI/2.0/ openarchivesprotocol.htm