ASTRONOMY WORKSHOP 2016

BRICS

ASTRONOMICAL DATA AND COMPUTATION

Ural Federal University

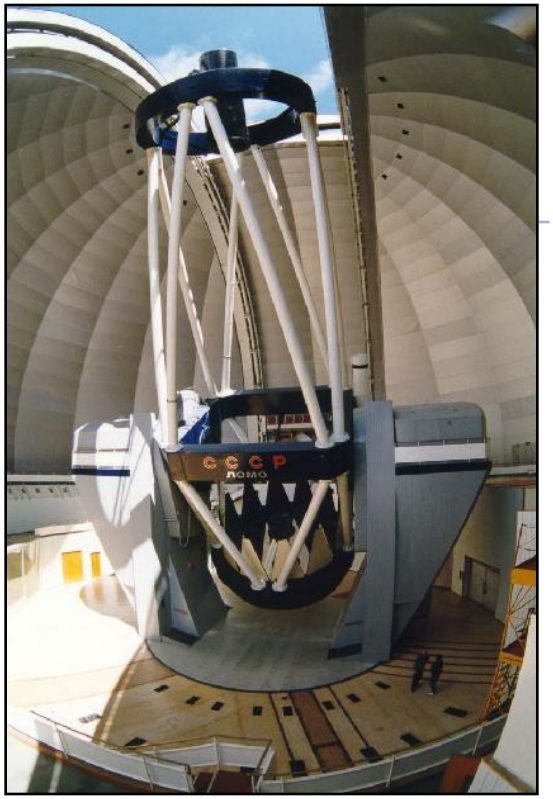
named after the first President of Russia B.N.Yeltsin

Problems of long-term preservation of observation data on the example of SAO RAS archive system

Zhelenkova O.P.[1,2], Vitkovskij V.V.[1,2], Plyaskina T.A.[1], Shergin V.S.[1], Chernenkov V.N.[1]

1-SAO RAS, Nizhnij Arkhyz; 2-ITMO University, Saint-Petersburg

ITMO UNIVERSITY



The Special Astrophysical Observatory of RAS has two the largest Russian telescopes – the 6m optical telescope Big Alt-Azimuthal Telescope (BTA) and radio telescope RATAN-600 with 6 hundred meters diameter antenna.

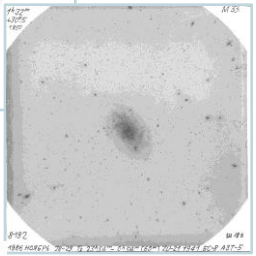
The Observatory became operational in 1966 and this year we are celebrating its 50th anniversary.

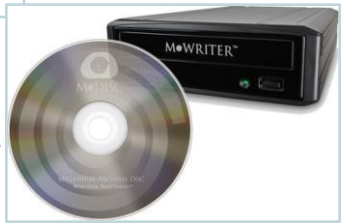
Systematic observations with the radio telescope RAT AN-600 and the 6-m optical telescope BTA started in 1974 – 1975 started and continuously are going on despite the fact that there are already in another state.

Some facts about the telescopes:

- BTA is the world's first large ground-based telescope with alt-azimuth mount, served as the prototype for similar telescopes, which were enacted later.
- For the first time the computer was used to control the telescope.
- Powerful emitters of one of the landing stages of the Apollo space missions observed on RATAN-600 in the first mapping of the Moon.

Information carriers and technologies for long-term keeping





1975

‘80s

‘90s

~2000

2016

?

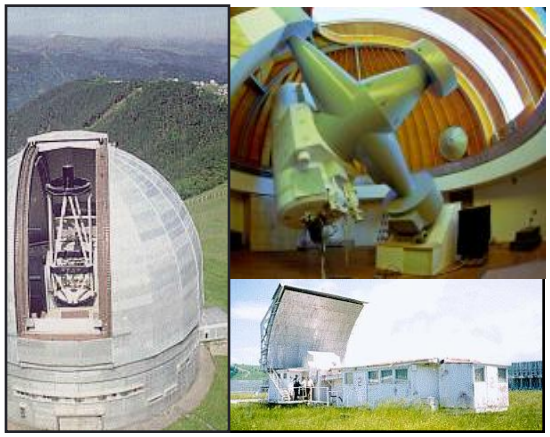
Keeping in boxes

Data bank

Archive system

Search information system
free web access

Migration ….



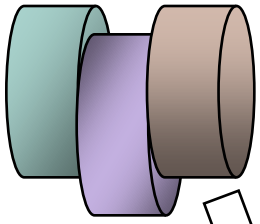
SAO RAS general archive of observational data

Optics: 238 CD/DVD

Radio: 7CD

Total: 245 CD/DVD

Local archives



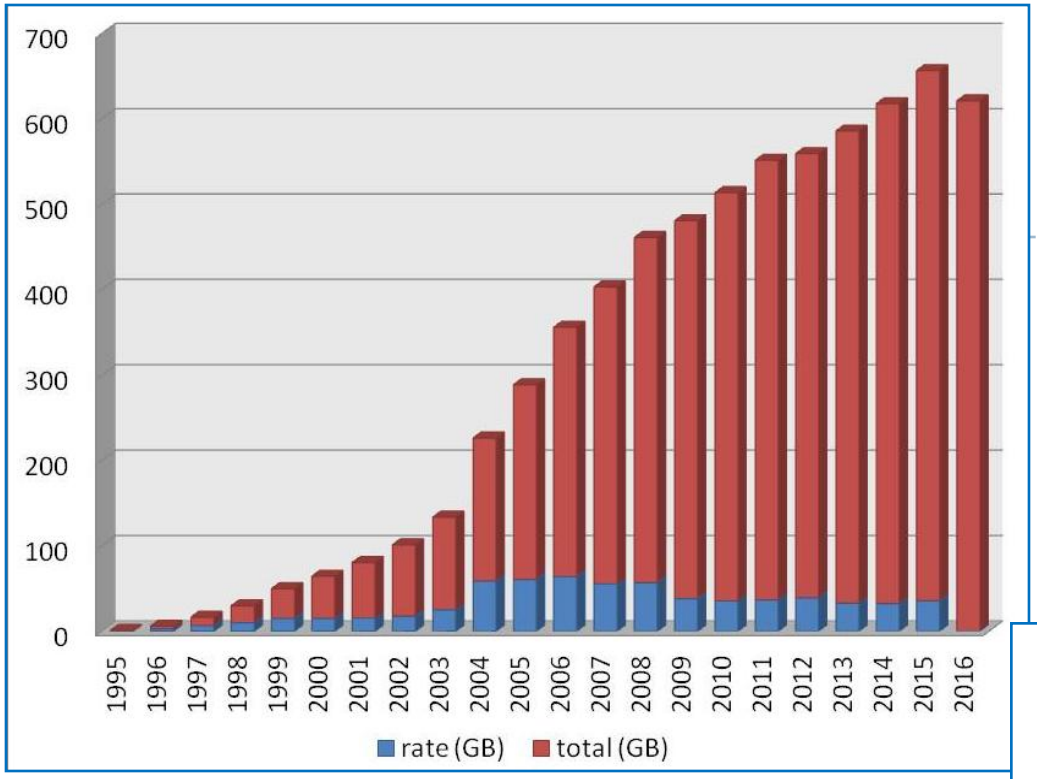
Formats of data: FITS, RFLEX, BINARY

Local archives: 16

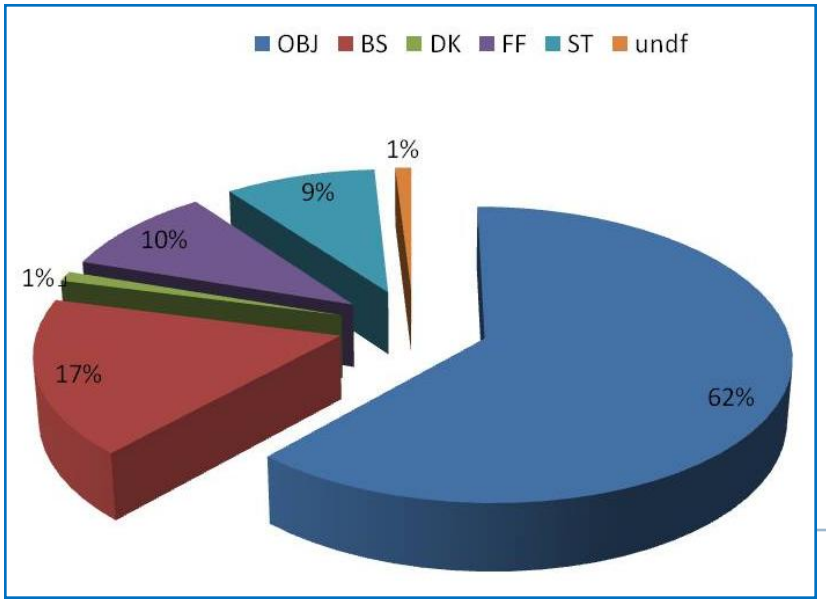
Numbers of files	~540000
Volume	0.5TB
Volume of storage area	1.5 TB
Number of records	>1000000

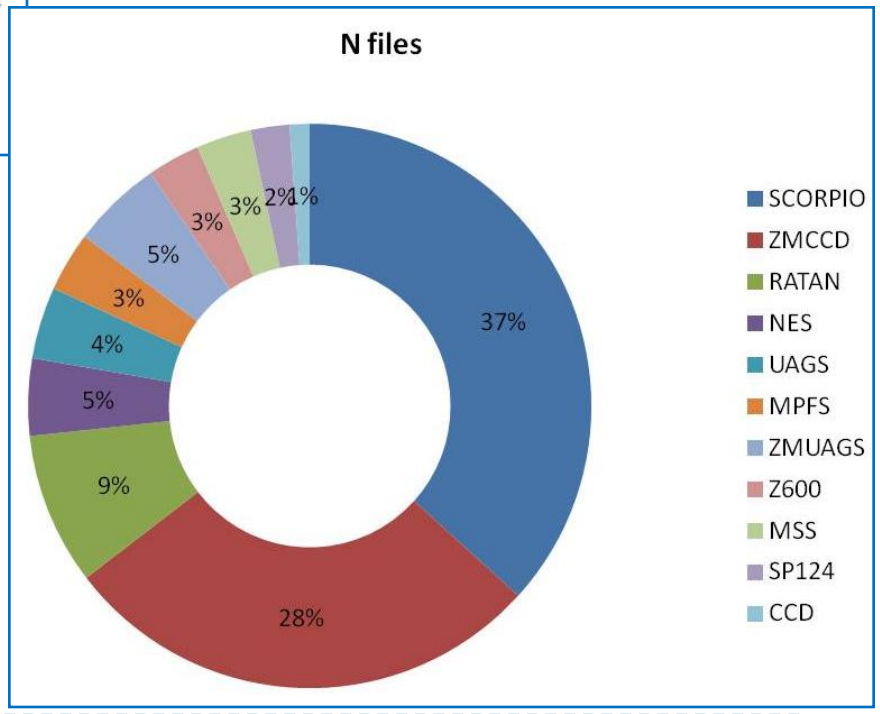
The characteristics of local archives

Archive	$N_{CD/DVD}$	N_{files}	Volume, MB	Range of date
LYNX	8	2106	1806	1996.02 – 2002.05
NES	50	23225	44995	1998.03 – 2014.04
PFES	6	2296	1993	1996.08 – 2001.01
MPFS	27	17998	31030	1996.08 – 2009.10
IFP	2	4503	1090	1997.05 – 2000.03
MOFS	1	1059	642	1997.03 – 2001.08
SCORPIO	65	193184	210650	2000.09 – 2016.01
UAGS	9	21696	4698	1994.11 – 2005.10
CCD	4	6011	2493	1996.02 – 2000.04
SP124	3	11676	1421	1996.02 – 2000.12
MSS	11	16543	14221	1996.05 – 2014.05
ZMCCD	40	146099	103246	1996.12 – 2015.10
ZMUAGS	15	27211	8247	1998.04 – 2016.02
CEGS	3	611	947	1997.03 – 2010.10
Z600	8	15951	4928	1996.01 – 2001.06
RATAN	7	45921	1735	1996.06 – 1999.01

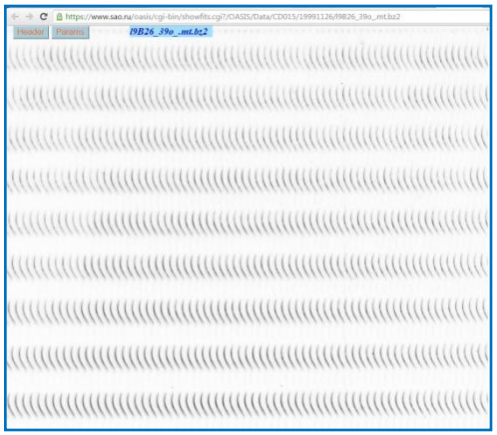


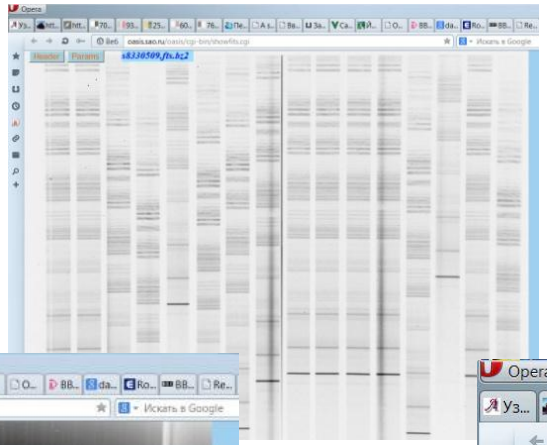
Types of files in the archive:
Observations – 91%
Observation logs - 1%
Auxiliary - 8%

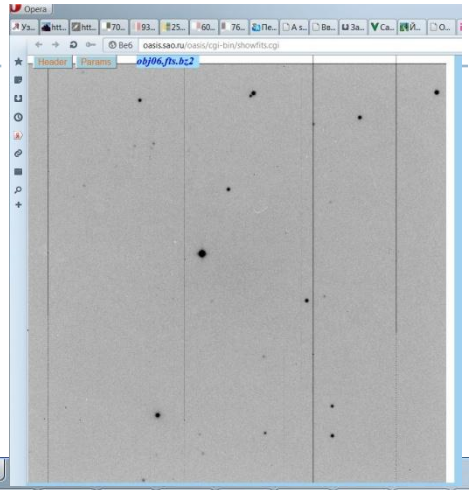


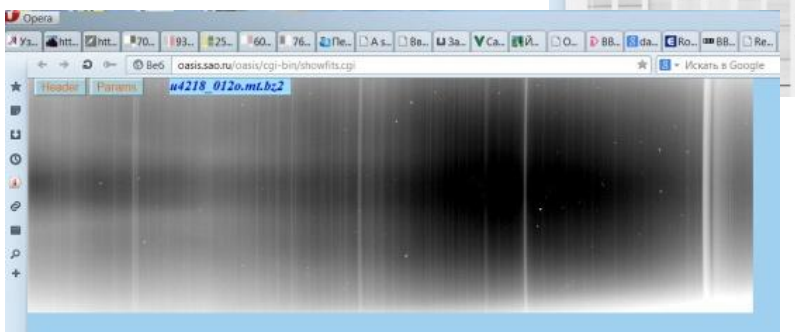


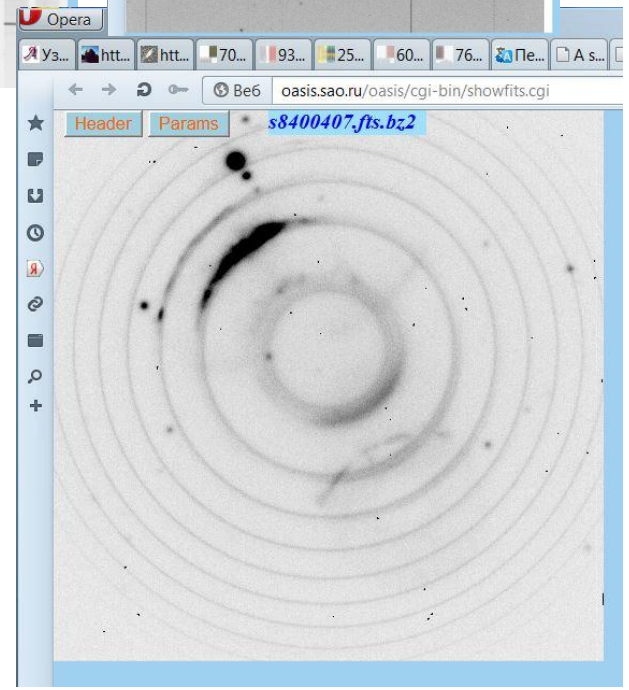
Data with heterogeneous structure. Examples of observations

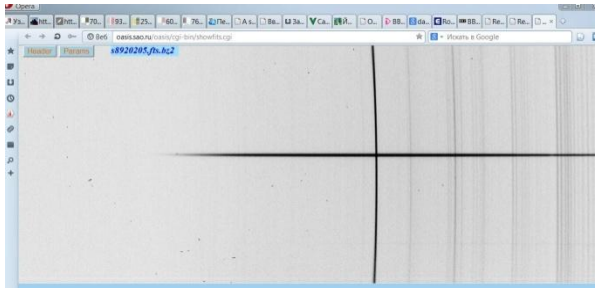












Parameters of an observation file obtained from different telescope systems and requested for the archive system

Observation
- date
- start time
- end time
- frame type

Specification of a file
- identifier
- file location
- inquiry number
- processing level
- file name
- Size
- last modification date

Access level:
- program
- principal investigator
- password
- observer
- origin
- telescope

Detector :
- name
- parameters (readout noise, bitpix

Observation target :
- coordinates
- object name

Data format:
- format type
- acquisation system

Instrument :
- name
- parameters (filter,slit,..)

Weather conditions:
- temperature
- wind
- mirror temperature
- pressure
- seeing

FITS-header for direct images

```
BITPIX = 16              /  No. of bits per pixel
NAXIS = 2                /  No. of axes in matrix
NAXIS1 = 1044            /  No. of pixels in X
NAXIS2 = 1046            /  No. of pixels in Y
BSCALE = 1.00            /  REAL = TAPE*BSCALE + BZERO
BZERO = 32768.0          /
DATAMAX = 9448.0  /  MAX PIXEL VALUE
DATAMIN = 78.0         /  MIN PIXEL VALUE
DATE = '2011-03-31T21:15:58.673'          /  UTC date this file was written
CREATOR = 'CCDServer v2.1'                /  ACQUSITION SYSTEM
DATE-OBS= '2011-03-31T21:15:43.007' /  UTC date of Observation start
TELESCOP= 'BTA 6-meter'                   /  TELESCOPE NAME
INSTRUME= 'SCORPIO '                      /  INSTRUMENT
OBSERVER= 'Makarov, Uklein'               /  OBSERVERS
OBJECT = ' G18.1+2'                       /  NAME OF IMAGE
PROG-ID = 'Groups of dwarf galaxies'      /  observational program
AUTHOR = 'Makarov '                       /  AUTHOR OF PROGRAM
FILE = 'S8560202.FTS'                     /  original name of input file
IMAGETYP= 'obj'       /  object, flat, dark, bias, scan, eta, neon, push
ORIGIN = 'SAO RAS '  /  observatory
START = '01:15:47'  /  measurement start time (local) (hh:mm:ss)
EXPTIME = 10.0        /  actual integration time (sec)
CAMTEMP = 143.251                         /  camera temperature (K)
DETECTOR= 'EEV CCD42-40'                  /  detector
RATE = 160.0          /  readout rate (KPix/sec)
GAIN = 1.946          /  gain, electrons per adu
NODE = 'B'            /  output node (A, B, AB)
BINNING = '2x2'       /  binning
PXSIZE = '27.0 x 27.0'                    /  pixel size (mkm x mkm)
UTC = 76543.200       /  UTC at exposure start (sec) 21:15:43.20
LST = 45424.730       /  Local sidereal time (sec) 12:37:04.73
RA = 165.3988782      /  Right Ascension (degr.) 11:01:35.73
DEC = 30.5988931      /  Declination (degr.) +30:35:56.0
EPOCH = 2000.0        /  EPOCH OF RA AND DEC
Z = 22.9              /  zenith distance
A = 63.0              /  azimuth
PARANGLE= 48.4        /  parallactic angle
ROTANGLE= 127.7       /  field rotation angle
SEEING = ' 1.5 '      /  seeing
FILTER = ' '          /  filter
FOCUS = 40.63         /  focus of telescope (mm)
IMSCALE = '0.355 x 0.355'                 /  image scale ("/Pix x "/Pix)
SLITWID =             /  slit width (")
```

```
MIRRTEMP= 2.4         /  mirror temperature (C)
DOMETEMP= 0.7         /  dome temperature ©
OUTTEMP = 0.6         /  outside temperature (C)
WIND = 5.3            /  wind (m/s)
CLOUDS = 0            /  clouds (%)
PRESSURE= 596.1       /  pressure
MODE = 'Image '       /  mode of instrument
DISPERSE= ' '         /  disperser, dispersion A/px
SPERANGE= ' '         /  spectral coverage
SLITMASK=             /  slit mask
SLITPOS =             /  slit coordinates in 2x2 map
ORDER =               /  order of dispersion TILTPOS
TILTPOS = ' '         /  tilt position
FILTERS = ' '         /  name of both wheels
FILTPOS1= 0           /  position of wheel number 1
FILTPOS2= 0           /  position of wheel number 2
POLAMODE=             /
CAMFOCUS= 5.59        /  focus of reducer (mm)
COLFOCUS=             /  FOCUS OF COLLIMATOR
QGCONST =             /  Queensgate constant
LSCAN =               /  wavelength of IFP scan
CHANNEL =             /  IFP channel
HISTORY
COMMENT   '
MJD-OBS = 55651.8859144 /  Modified Julian Date of observation
AIRMASS = 1.08538  /  airmass for current zenith distance
WCSAXIS = 2           /  Number of WCS axes
CTYPE1 = 'RA---TAN'                       /  RA-Gnomic projection
CUNIT1 = 'deg'        /  RA units - degrees
CRPIX1 = 466.3        /  X reference pixel
CRVAL1 = 165.3988782                      /  RA of reference pixel
CTYPE2 = 'DEC--TAN'                       /  Decl-Gnomic projection
CUNIT2 = 'deg'        /  Decl units - degrees
CRPIX2 = 551.0        /  Y reference pixel
CRVAL2 = 30.5988931                       /  Decl of reference pixel
EQUINOX = 2000.0      /  Equinox of equatorial coordinates
RADECSYS= 'FK5'       /  using FK5 coordinates system
CD1_1 = -0.0000592925                     /  rotation matrix coefficient [1,1]
CD1_2 = 0.0000786492                      /  rotation matrix coefficient [1,2]
CD2_1 = 0.0000786492                      /  rotation matrix coefficient [2,1]
CD2_2 = 0.0000592925                      /  rotation matrix coefficient [2,2]
HISTORY scorpio_wcs: dA=0.01dgr. Scale=0.3546 dX=-54.7 dY=30.0
END
```

SAO RAS general archive of observations

includes:

- scientific data
- calibration

provides:

- free web access
- long-term keeping and preservation

basic principles:

- the semantic item is a file with observation;
- store in the archive all observations carried out at the observatory telescopes;
- archive does not change the format of stored data;
- exclusive copyright to use data containing information about astrophysical objects, within 2 years after the observations belong to the applicants of the observation program.

Basic requirements for the organization of data in the archive

- Archive disc has a disc label and contains directories with data
- One directory on the disk contains one observation date
- The directory name corresponds to the observation date
- One observation is stored in a single file
- There are no restrictions on file format

STANDARD QUERIES

- Observation date
- Coordinates
- Source name
- Observation program
- Principal investigator
- Filter
- Type of file
- Mode of observation

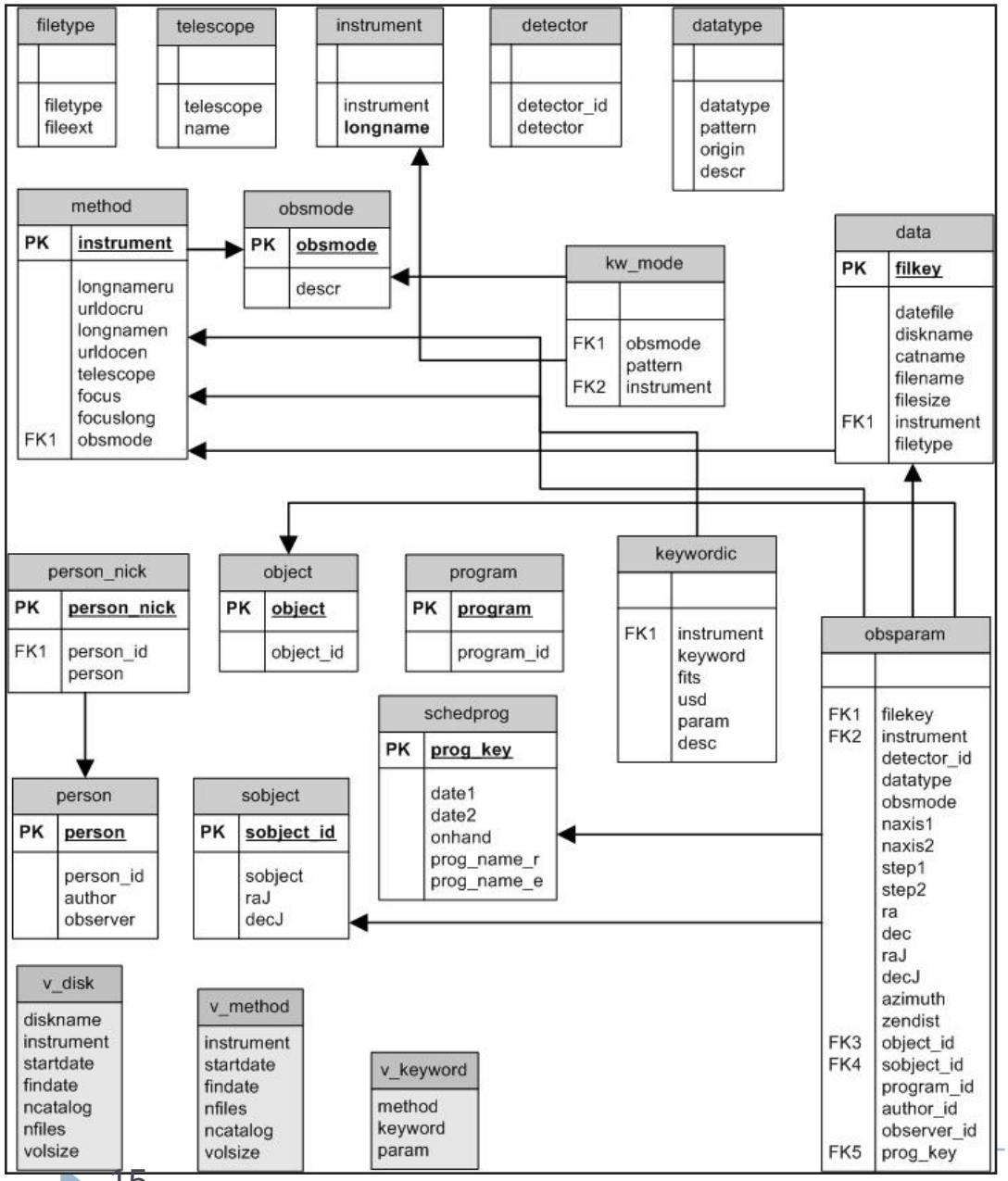
Standard requests, except for selection by observation date, due to lack of the necessary parameters in the file header are implemented to the part of observation files:

- no coordinate values - 17%
- no object name - 4%
- not defined program name - 30%
- not defined applicant - 29%
- unidentified observers - 29%

OASIS - Observation Archive Search Information System

SIS of the archival system realized in 3-tiers architecture ИПС «client-application server-database server».
The database (OS Fedora 12, DBMS PostgreSQL8.4) includes 17 tables and views.
The web interface (Perl, CGI, DBD/DBI) supports the set of standard queries and и http-copying of requested data.

OASIS table schema

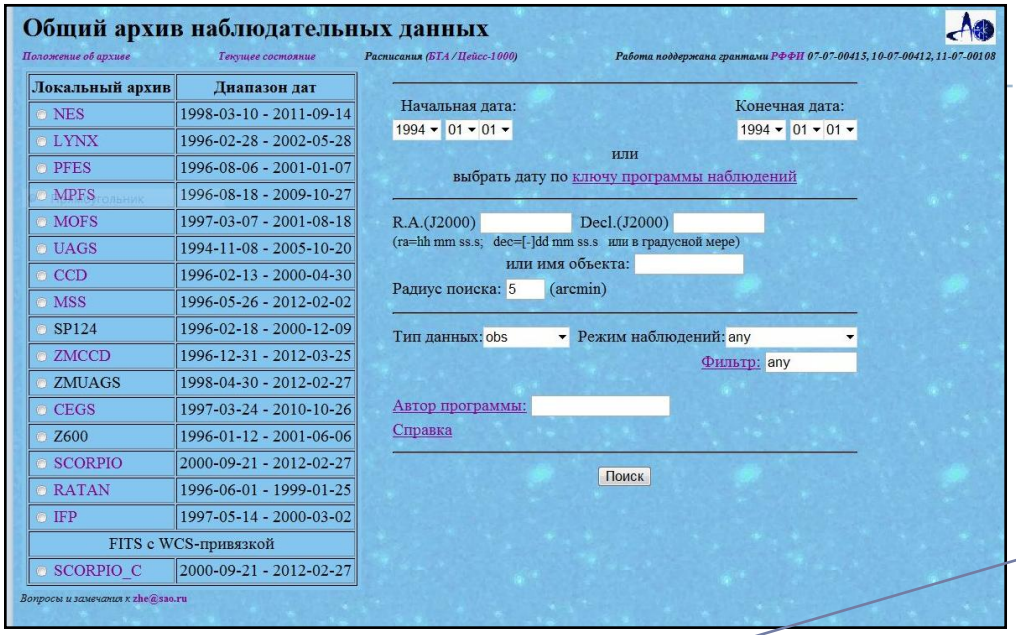


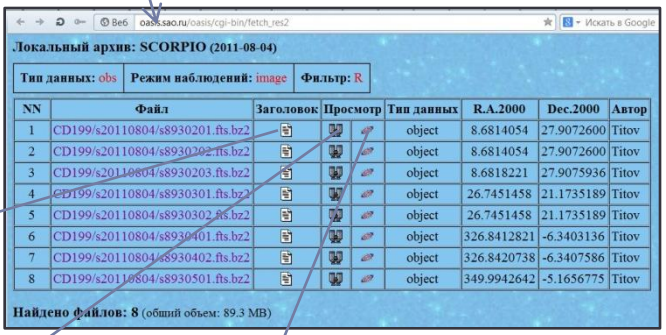
Each file is described with about 60 parameters.

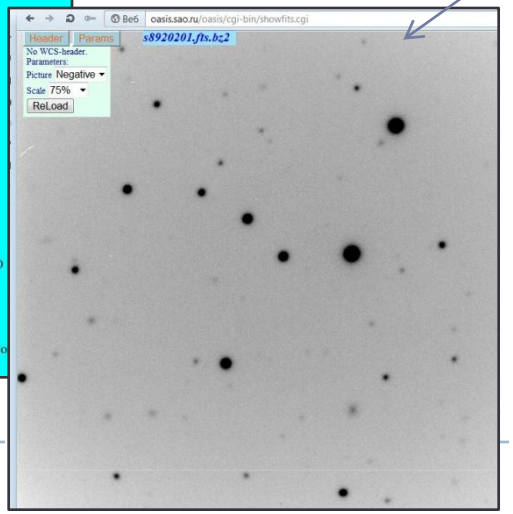
They are used for the dynamic formation of the web interface, for mapping parameters FITS-header and attribute of SIS tables, identify files, specify a file type, etc.

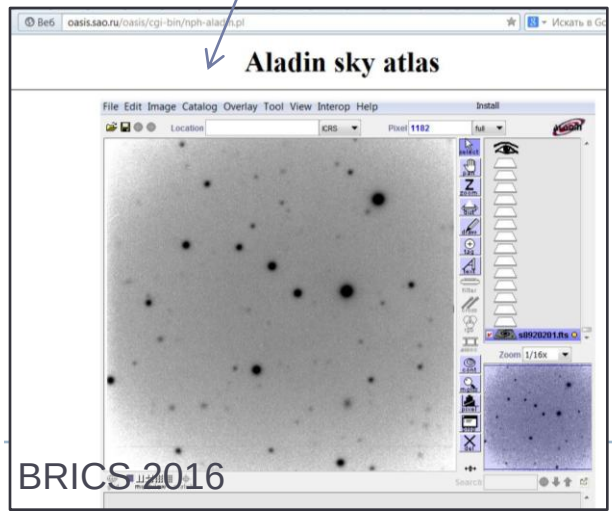
Some tables are vocabularies, others contain parameters describing each observation file.

OASIS web interface (I)

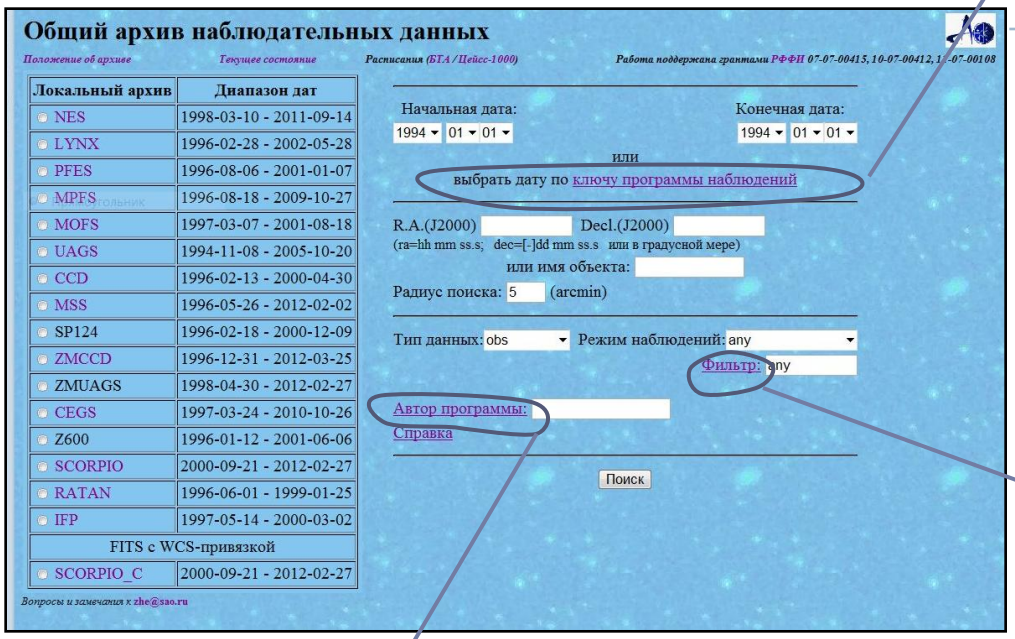


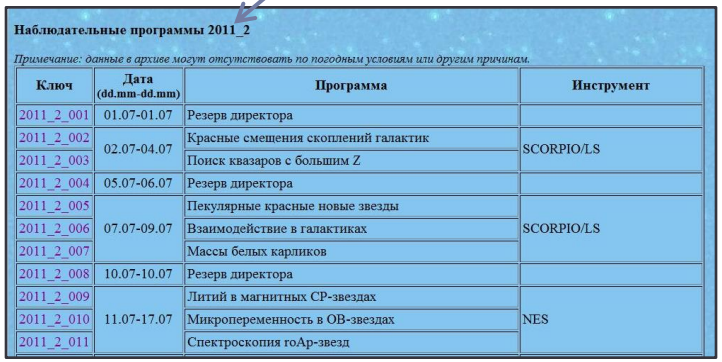


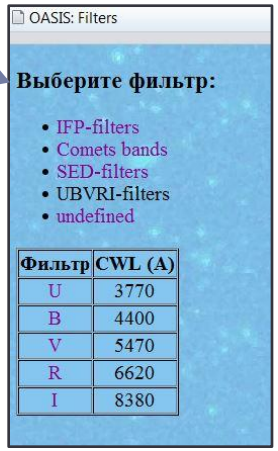


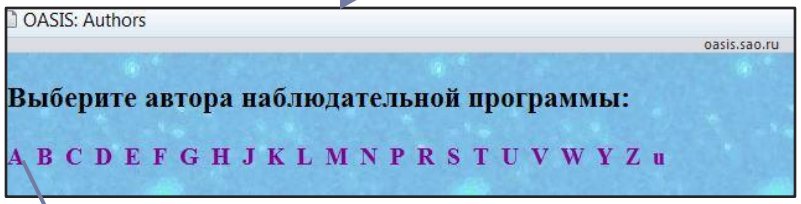


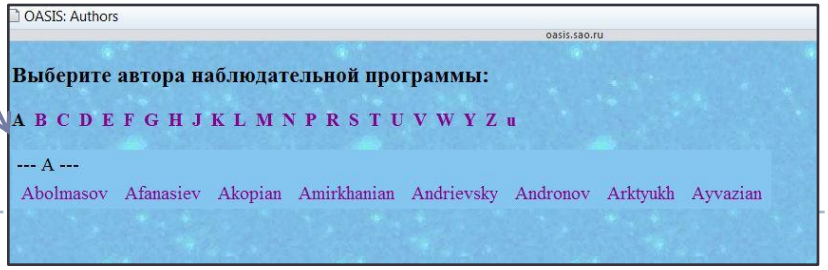
OASIS web interface (II)



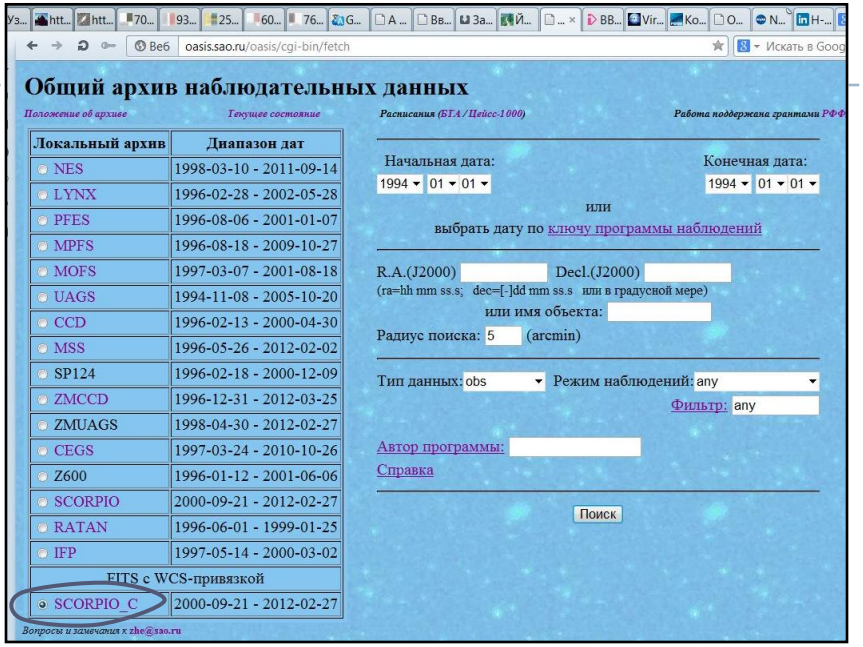


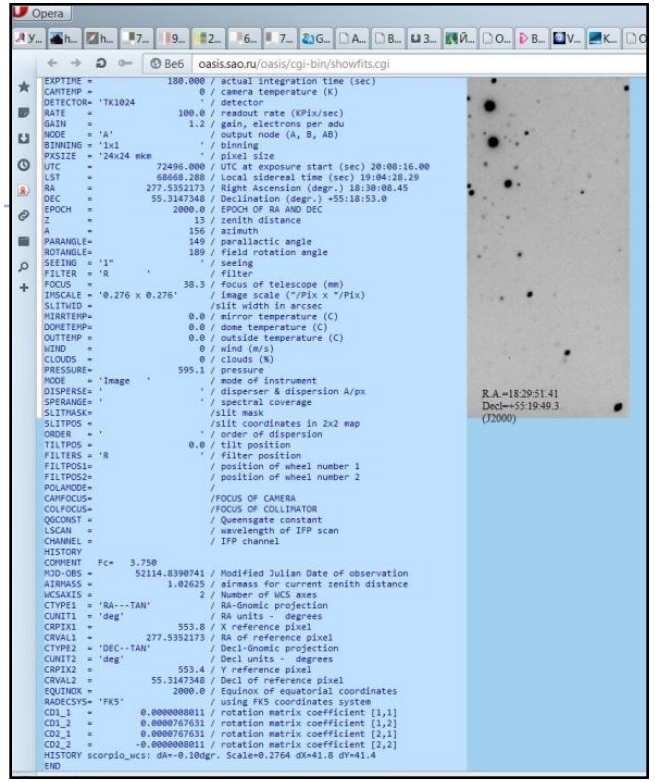


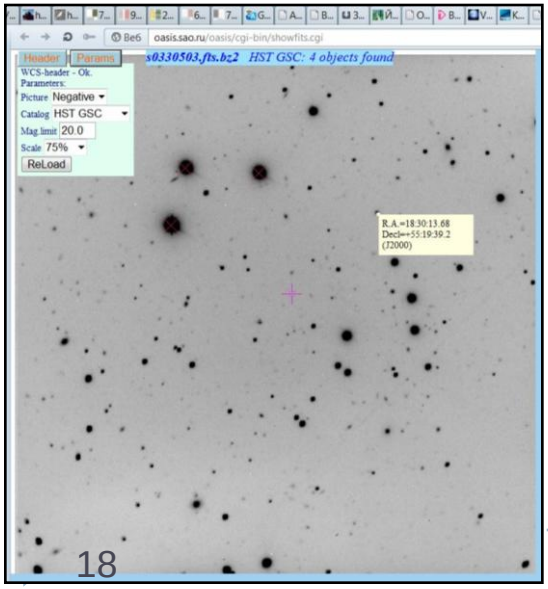


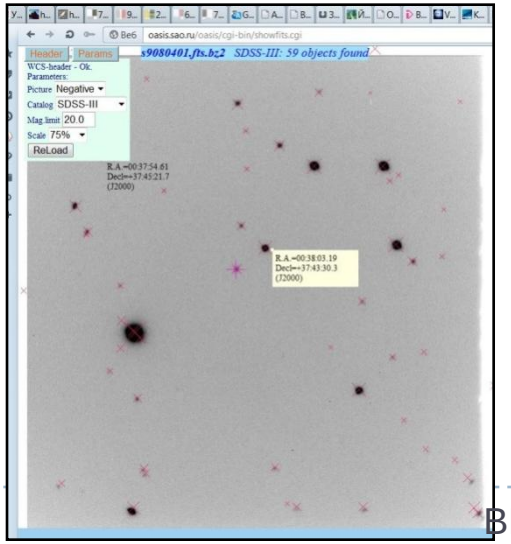


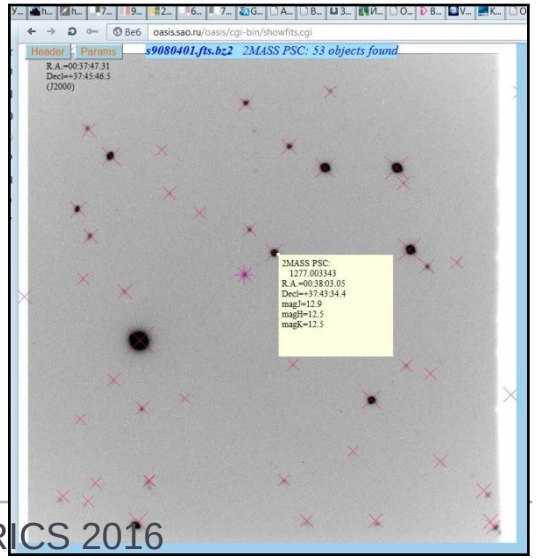
OASIS web interface (III)



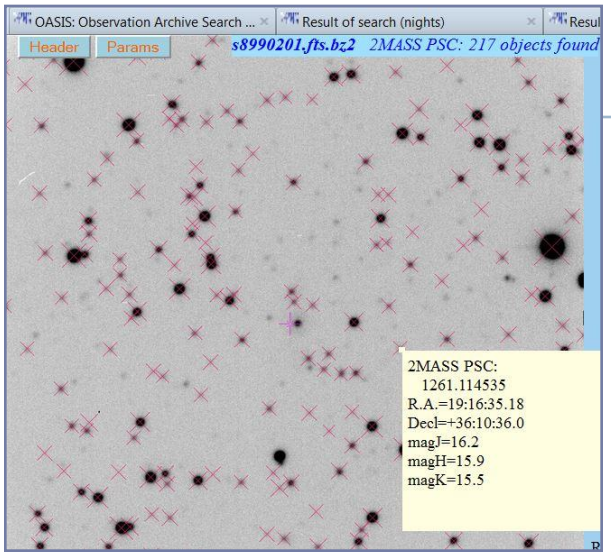


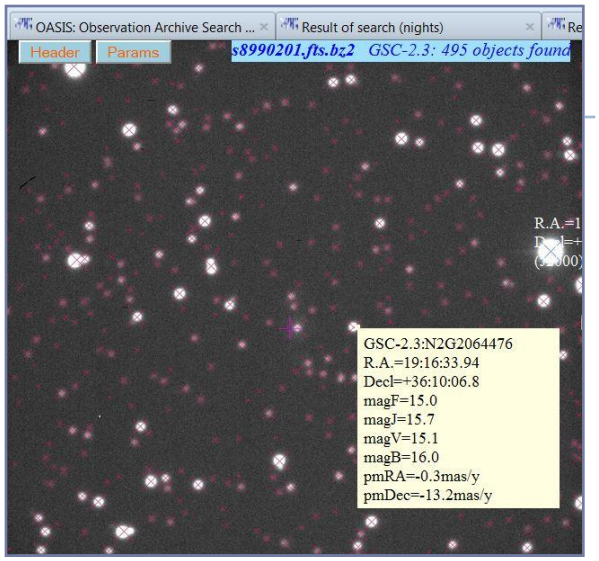


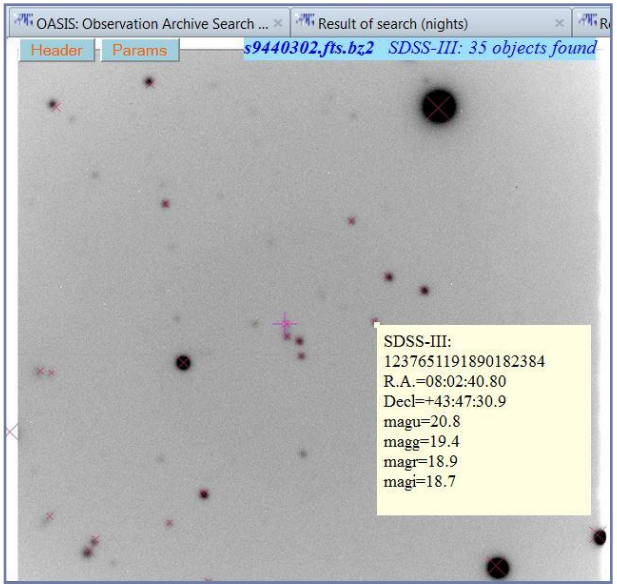


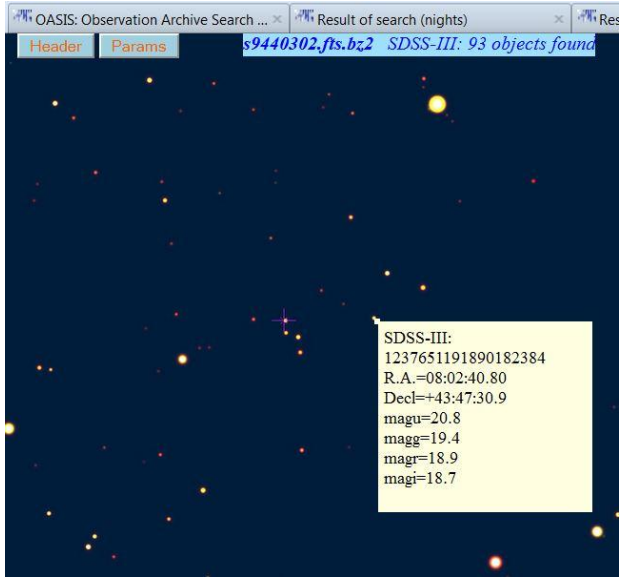


OASIS web interface (IV) – on-fly visualization

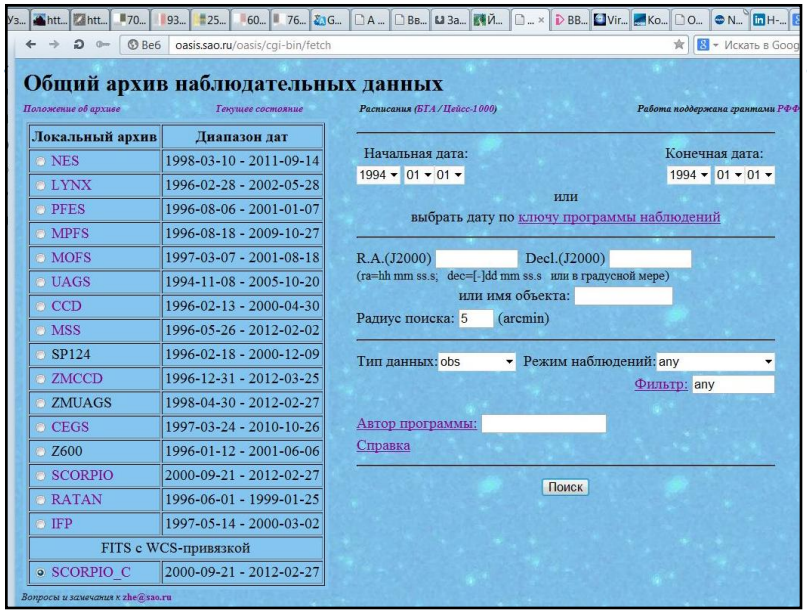


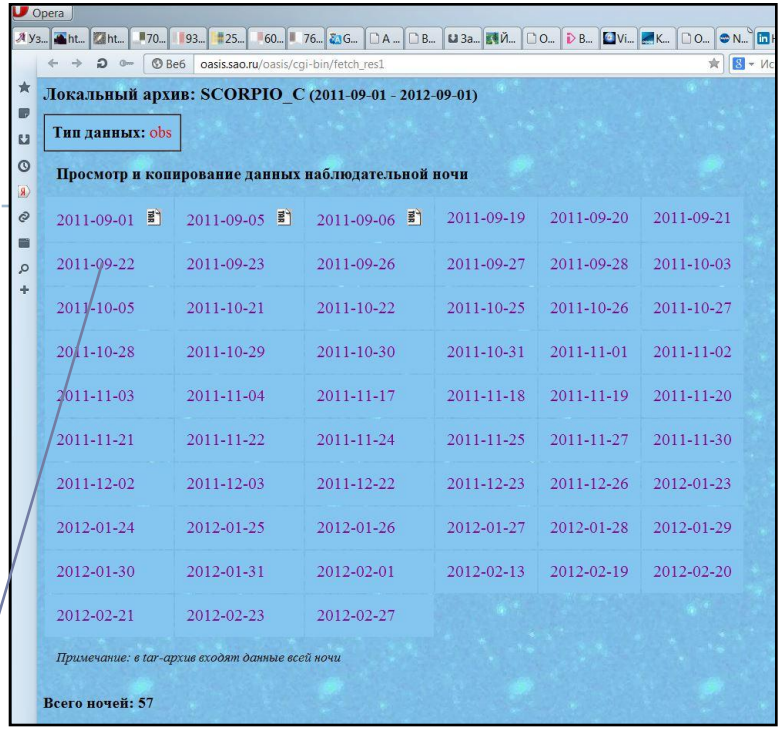


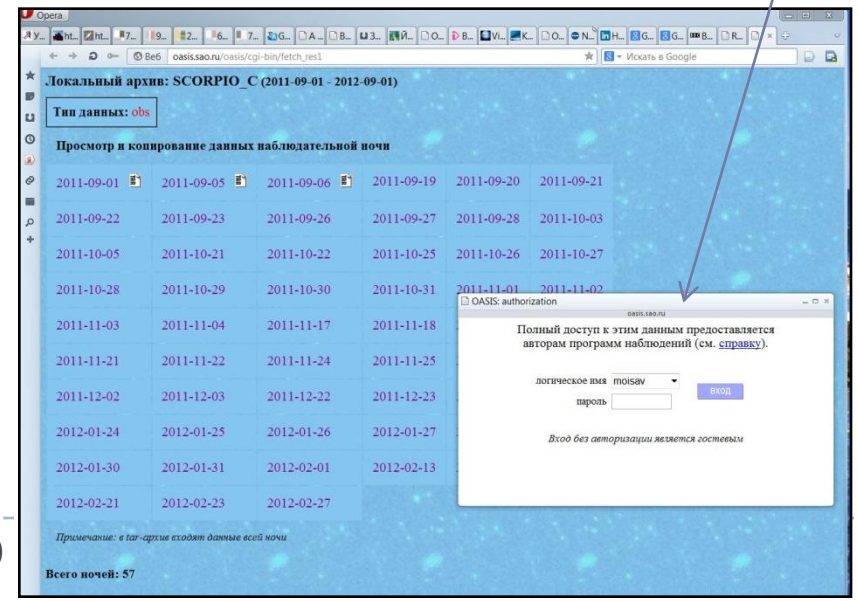




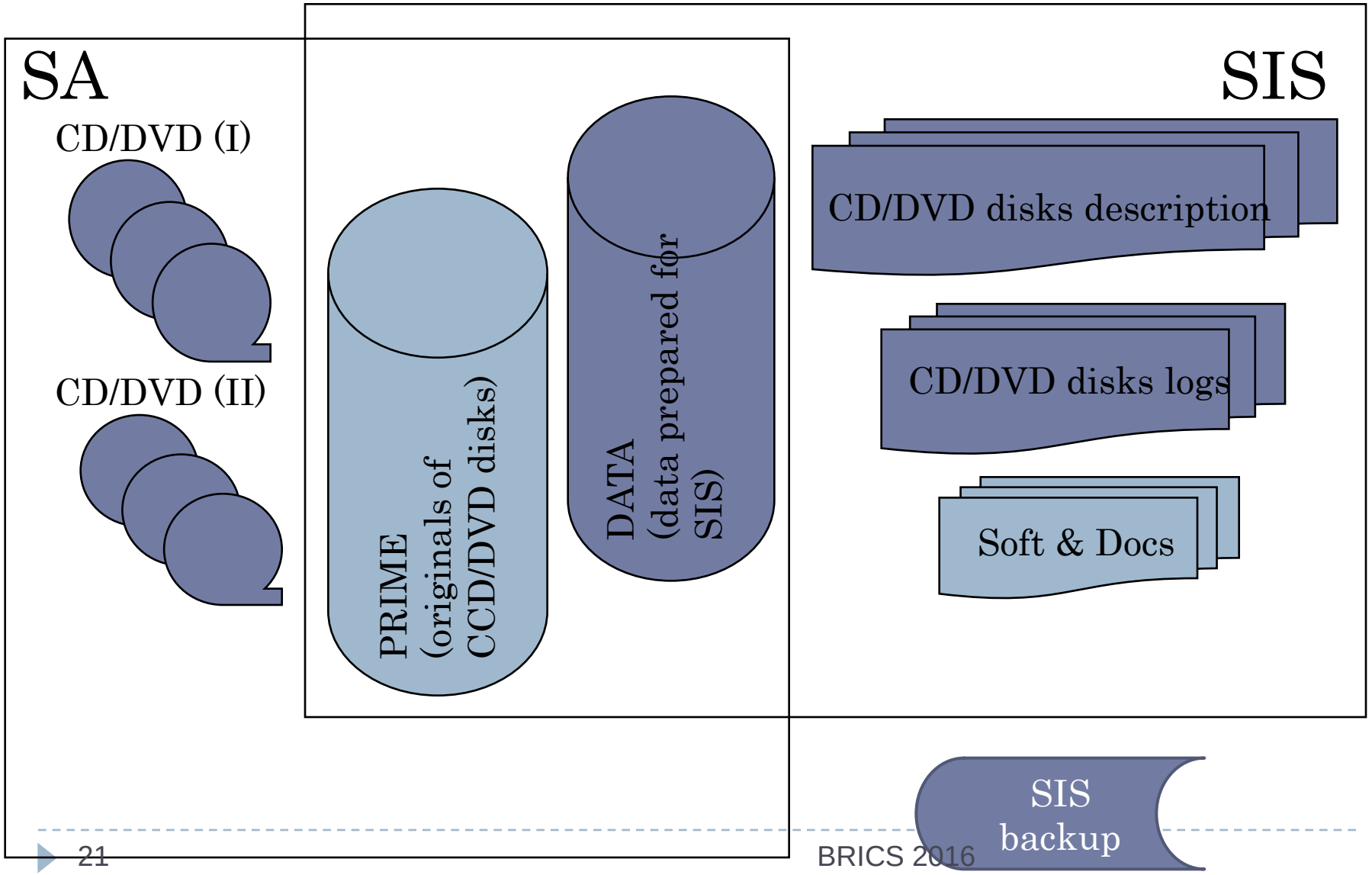
OASIS web interface (IV)







The archival system includes a storage area (SA) and seach information system (SIS)



Storage area of the archival system

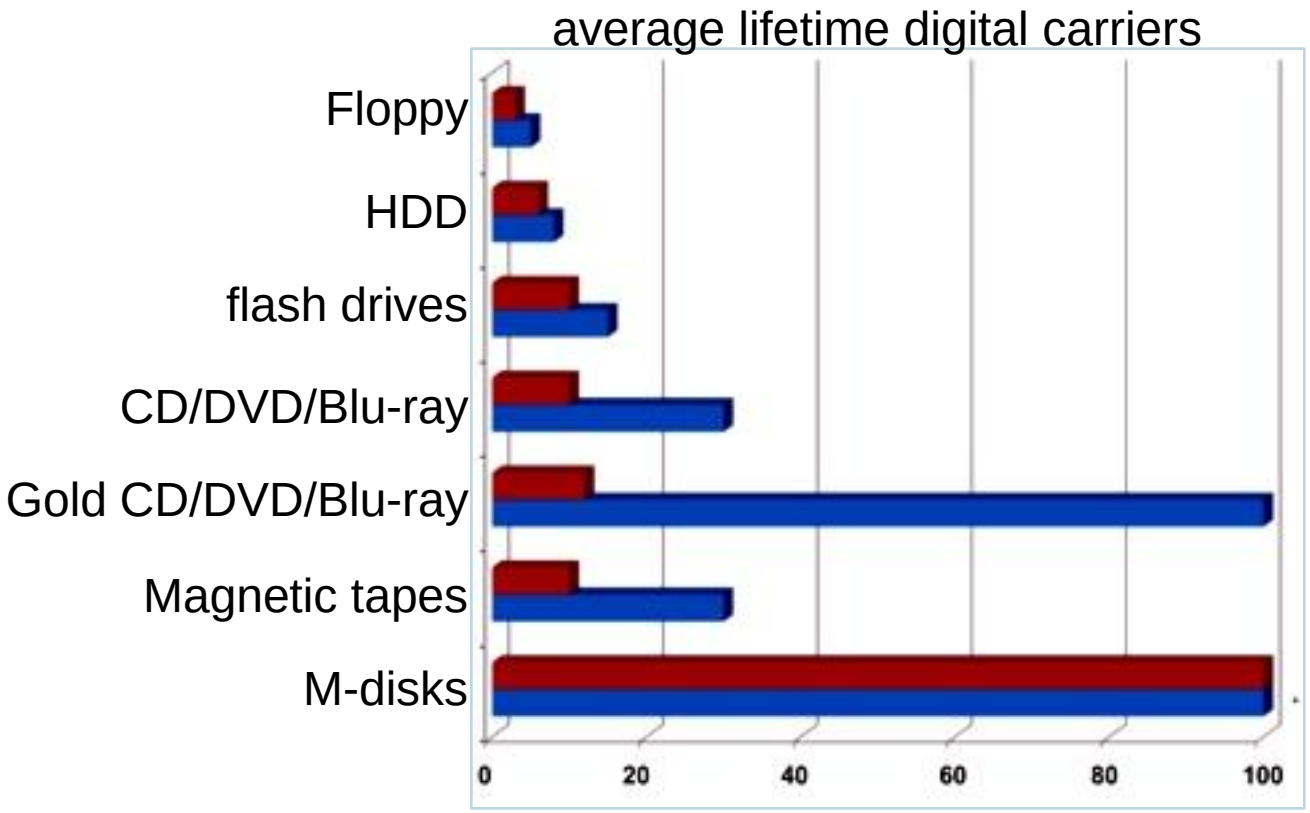
- 2 copies of the archived data on CD/DVD: MASTER (I) и COPY (II);
- on a dedicated server for the working version of SIS: PRIME and DATA (~1.5TB);
- on the file server for the test case of SIS: PRIME and DATA (~1.5TB);
- full SIS backup, including two copies of the data, service files and software.

Migration of radio observations

- 6 local archives for each type of radiometric complex and secondary reflector
- ref1: 1982-01 – 2016 (more 30 years)
- ref2: 2011-11 – 2016
- ref3: 2013-02 – 2016
- ref4: 2011-07 – 2013-01
- ref5: 2008-09 – 2008-10
- ref6: 1988-02 – 1988-04

volume: ~74GB compressed data, ~540000 files

reliability of digital carriers

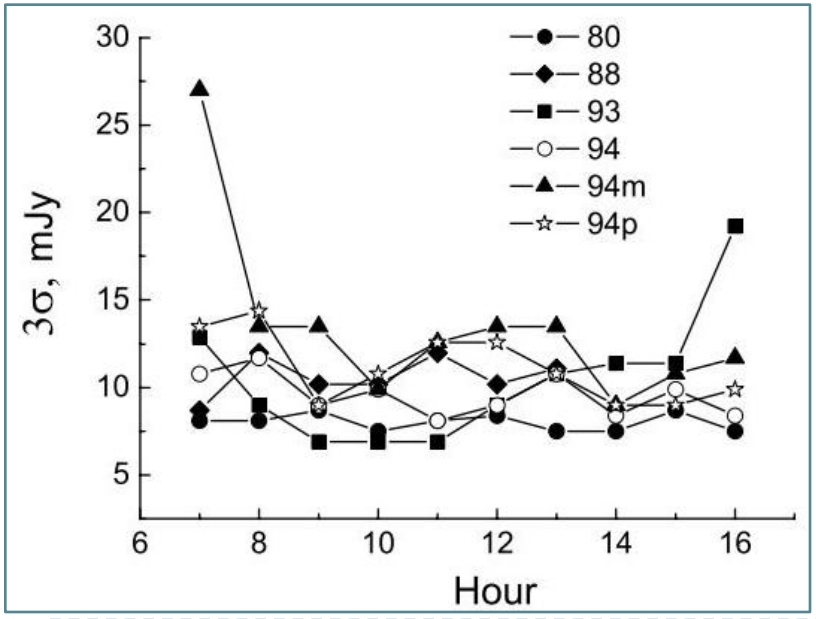


expected frequency of migration

RATAN-600 «Cold» blind surveys 1980-1994

survey	H	date	t, days	N	σ, mK	$F^{lim}_{3.94GHz}$, mJy
1980	51°08'	15.03.1980-06.06.1980	84	25-50	0.7	8.0±0.5
1988	51°09'	16.12.1987-12.01.1988	28	25	1.1	10.6±1.3
1993	51°10'	17.09.1993-01.11.1993	46	46	1.6	10.4±1.2
1993	51°22'	01.04.1994-25.05.1994	55	40	1.2	11.1±2.0

(Soboleva+, 2010AstBu..65…44)

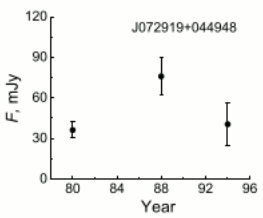


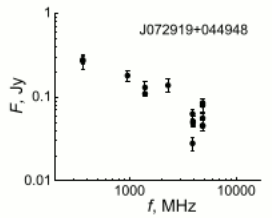
1) The ability to detect variable sources based on blind RATAN-600 surveys (Majorova&Zhelenkova, 2012AstBu..67..318);
2) Search candidates of variables sources in the range 7h<RA<17h (Majorova&Zhelenkova, 2013AstBu..68..371) and in 2h<RA<7h (Majorova+, 2015AstBu..70…34);
3) Search sources which were observed only in one survey, including the transient events (Zhelenkova&Majorova, 2016);

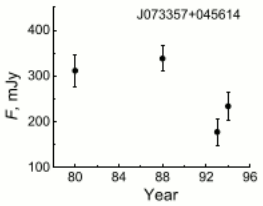
Кандидаты в переменные источники

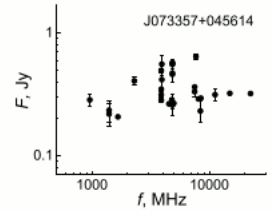
RCR RA_{2000} Dec_{2000}	V	V_R	V_F	$\overline{F}$, mJy	σ^{set}, mJy	RMS^{set}	$\overline{dH}$, arcmin	α	Notes
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
J072919.57+044948.7	0.189	2.09	2.61	51	22	0.428	5.5	−0.67	
J073357.46+045614.1	0.199	1.90	3.99	265	73	0.276	0.6	0.11	*
J075314.02+045129.4	0.025	1.35	1.64	183	25	0.138	4.2	−0.34	*
J080757.60+043234.6	0.275	2.68	3.12	374	142	0.380	−23.0	−0.30	
J081218.14+050755.5	0.097	1.92	1.91	118	32	0.266	11.5	−0.75	
J081626.62+045852.8	0.106	1.67	2.42	53	14	0.272	3.5	−0.88	
J083148.89+042938.5	0.130	1.89	2.32	949	247	0.261	−25.5	0.04	*

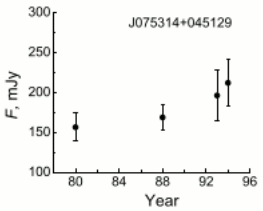
RCR RA_{2000} Dec_{2000}	p_{df}	p_{df-1}	$\overline{p}$,	$\langle F \rangle$, mJy	ΔF, mJy	V_χ	$\langle \sigma \rangle$, mJy	$\langle \sigma \rangle^{otn}$	χ^2	df
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
J072919.57+044948.7	0.966		0.966	42	16	0.385	5	0.124	6.82	2
J073357.46+045614.1	1	0.999	0.999	266	103	0.388	15	0.056	18.80	3
J075314.02+045129.4	0.653	0.796	0.725	173	10	0.057	10	0.059	3.31	3
J080757.60+043234.6	0.991	0.995	0.993	302	179	0.594	37	0.124	10.67	3
J081218.14+050755.5	0.738	0.906	0.823	110	21	0.194	12	0.112	4.01	3
J081626.62+045852.8	0.974	0.990	0.983	51	17	0.337	4	0.078	9.25	3
J083148.89+042938.5	0.960	0.984	0.972	859	316	0.367	79	0.092	8.31	3

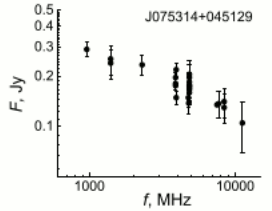


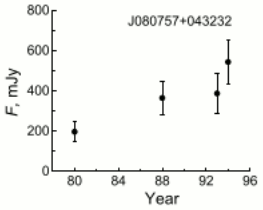


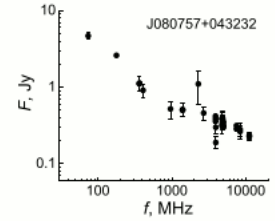






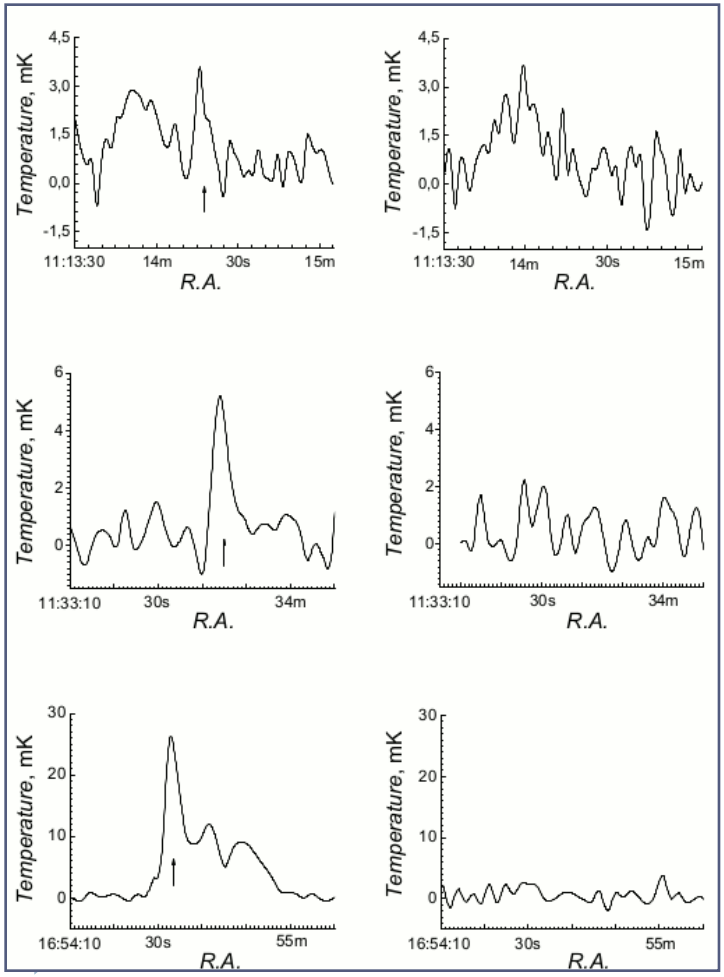




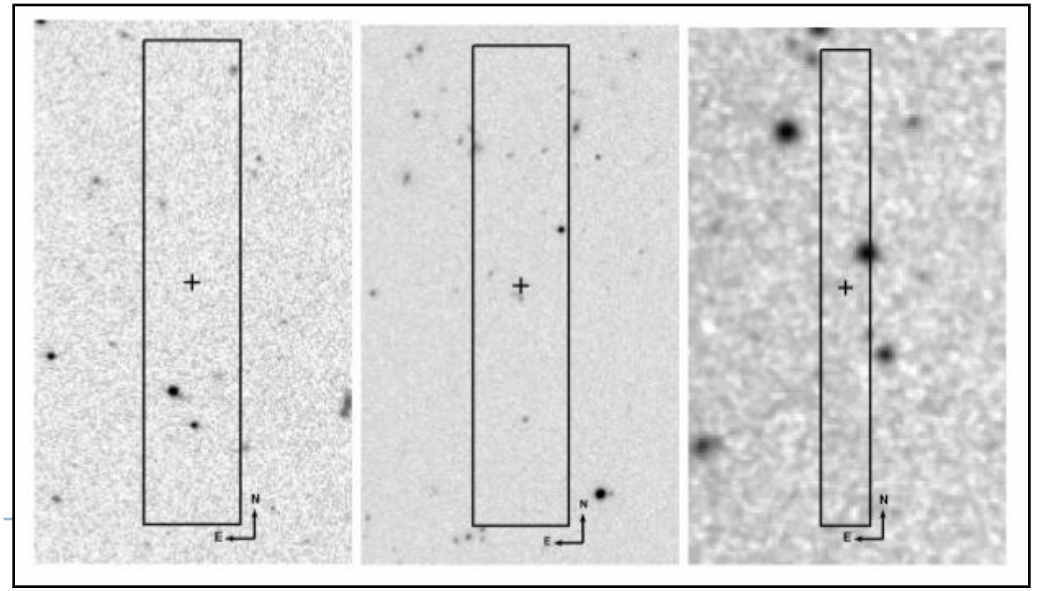


Sources detected only in one survey, including transient events

A criterion of the transient nature of the source:
absence it in NVSS and other catalogs,
detection on only one of the surveys despite the fact that the sensitivity of at least a further survey, it was enough to detect it.



Имя	*R.A.* hhmmss.s	*Dec.* sddmmss	$F_{3.94\,ГГц}$ мЯн	T_a/σ 1988–1994 гг.
(1)	(2)	(3)	(4)	(5)
J111417+045530	111416.7±0.6	+045530±45	21.0±2.0	5.2÷7.5
J113344+045030	113344.1±0.6	+045030±45	24.3±2.5	3.4÷4.0
J165433+045457	165433.1±0.3	+045457±45	88.2±8.5	11.1÷16.6



Thanks for attention!

- The works supported by RFBR grants:
07-07-00415, 10-07-00412, 11-07-00108, 14-07-00361