Storing and Accessing the largest astronomical catalogues with the SAI CAS project

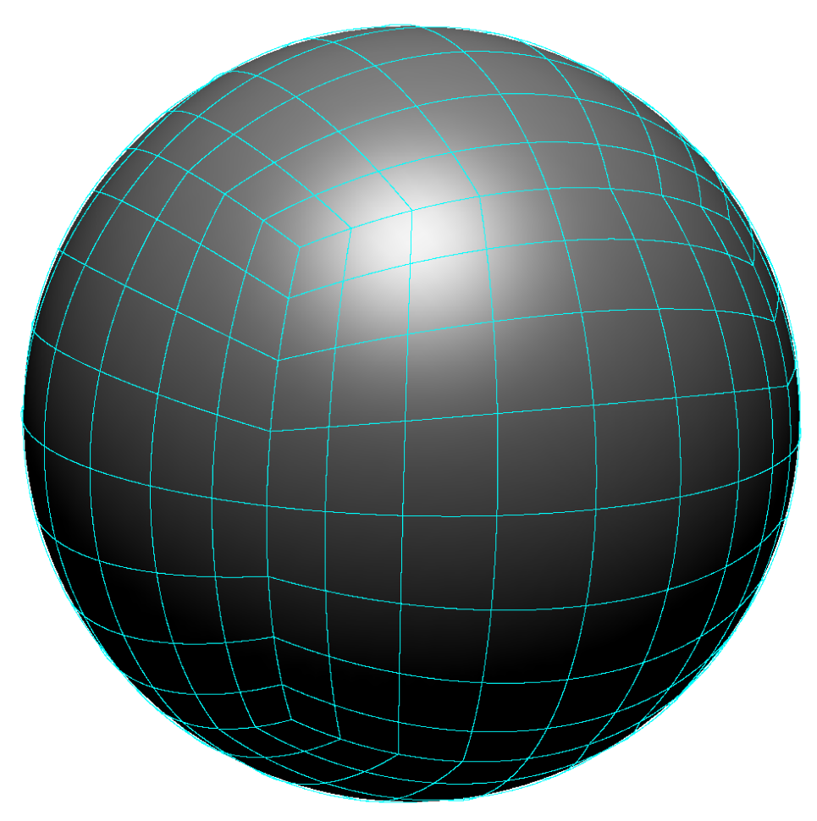
Sergey Koposov[1,2,3,*], Oleg Bartunov[2], Sergey Karpov[4], Aleksandr Belinskiy[2]
[1] Max Planck Institute for Astronomy, Konigstuhl, 17, Heidelberg, Germany
[2] Sternberg Astronomical Institute, Universitetskiy 13, Moscow, Russia
[3] Cambridge Institute of Astronomy, Madingley road, Cambridge, UK
[4] Special Astrophysical Observatory, p. Nijniy Arhyz, Russia
[*] math@sai.msu.ru

http://vo.astronet.ru/

We present the new software system for storing and accessing the large astronomical catalogues. The system is fully based on the open-source software like Axis, Tomcat, PostgreSQL and is available for download. The SAI CAS project provides users with a large set of capabilities and services such as the ConeSearches in the catalogues in the system, the crossmatches of the user data with the hosted catalogues, the SkyNode service, the different WS to manipulate the data and metadata in the catalogues, upload the user's data into the system and others. The SAI CAS system deployed in Sternberg Astronomical Institute already hosts the largest data collection in Russia, and one of the largest in the world and includes USNO (A2, B1), 2MASS, DENIS, NOMAD, SDSS, Tycho, XMM, UCAC2.

The motivation of our project:

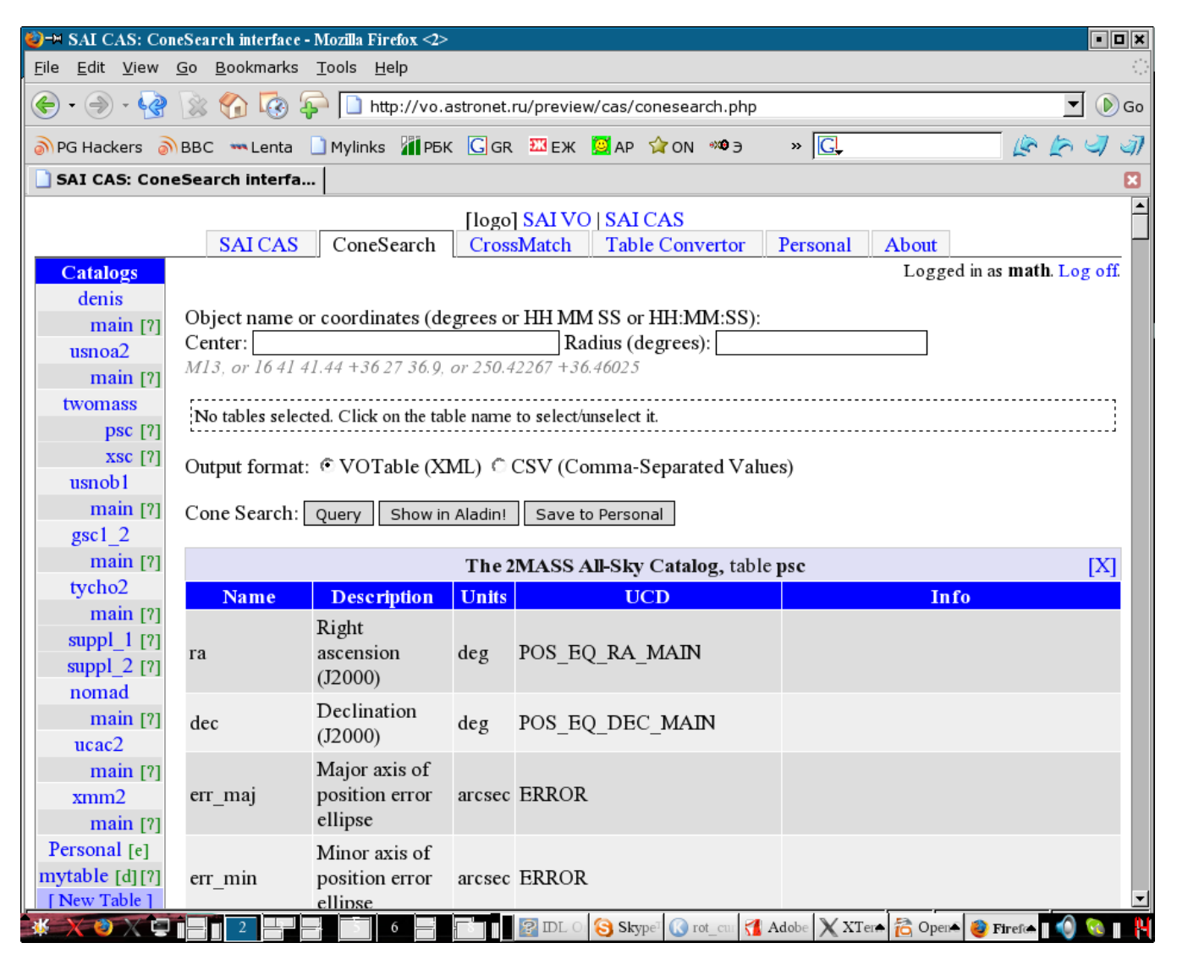
- *Create the infrastructure for storing of the large astronomical catalogues and their metadata.*
- The catalogues and their metadata should be queriable.
- The system should be able to output/input the data both in simple formats like ASCII or CSV and in the VOTable formats
- The catalogue system should support the fast conesearches & crossmatches
- The system should provide the functionality similar to SDSS CASjobs (uploading the data into the system, and perform the queries on the datasets).
- The system should by VO aware and ready for being deployed as part of the VO services such as SkyNode.
- The system should be able to be operated through the WS, HTTP POST/GET and simple web interfaces.



The heart of the SAI CAS project is the PostgreSQL database and the Q3C plugin for PostgreSQL performing the spatial queries & crossmatches.

SAI CAS system running in Sternberg Astronomical Institute:

SAI CAS now is a fully working system deployed in Sternberg Astronomical Institute.
We host the largest astronomical datasets: USNO-B1, USNO-A2, 2MASS, XMM, NOMAD, GSC 1.2, Tycho, UCAC2, DENIS, SDSS (not fully exposed yet) These catalogues can be ConeSearch queried, can be crossmatched with the user's tables. The registered users have the ability to upload their data into the system in different formats. The data and metadata of both system and user tables can be queried through the set of Web Services, and HTTP GET/POST. The CAS can be easily explored with a user friendly web interface .
The screenshots demonstrating basic functionality and the current state of SAI CAS are shown on Fig. 1 (the conesearch interface together with catalogue/table browsing interface & the crossmatch interface with the user's table).


SAI CAS: ConeSearch interface - Mozilla Firefox <2>
File Edit View Go Bookmarks Tools Help
http://vo.astronet.ru/preview/cas/conesearch.php
[logo] SAI VO | SAI CAS
SAI CAS ConeSearch CrossMatch Table Convertor Personal About
Catalogs
denis main [?] usnoa2 main [?] twomass psc [?] xsc [?] usnob1 main [?] gsc1_2 main [?] tycho2 main [?] suppl_1 [?] suppl_2 [?] nomad main [?] ucac2 main [?] xmm2 main [?]
Personal [e] mytable [d][?] [New Table]
Query
Show in Aladin!
Save to Personal

Logged in as **math**. Log off.

Object name or coordinates (degrees or HH MM SS or HH:MM:SS):
Center: Radius (degrees):
M13, or 16 41 41.44 +36 27 36.9, or 250.42267 +36.46025

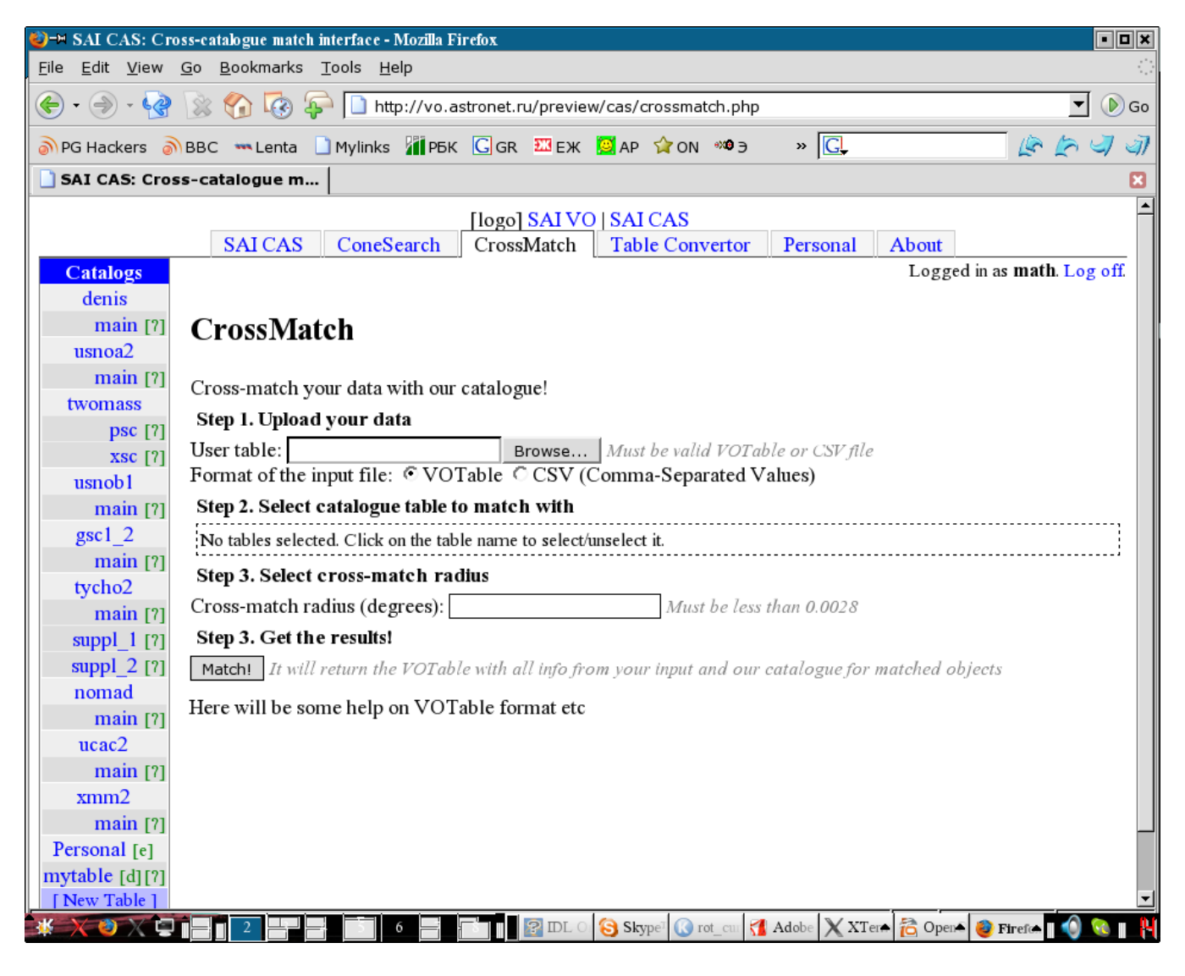
No tables selected. Click on the table name to select/unselect it.

Output format: (•) VOTable (XML) ( ) CSV (Comma-Separated Values)

Cone Search:

The 2MASS All-Sky Catalog, table **psc** [X]

Name	Description	Units	UCD	Info
ra	Right ascension (J2000)	deg	POS_EQ_RA_MAIN	
dec	Declination (J2000)	deg	POS_EQ_DEC_MAIN	
err_maj	Major axis of position error ellipse	arcsec	ERROR	
err_min	Minor axis of position error ellipse	arcsec	ERROR	


SAI CAS: Cross-catalogue match interface - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://vo.astronet.ru/preview/cas/crossmatch.php
[logo] SAI VO | SAI CAS
SAI CAS ConeSearch CrossMatch Table Convertor Personal About
Catalogs
denis main [?] usnoa2 main [?] twomass psc [?] xsc [?] usnob1 main [?] gsc1_2 main [?] tycho2 main [?] suppl_1 [?] suppl_2 [?] nomad main [?] ucac2 main [?] xmm2 main [?]
Personal [e] mytable [d][?] [New Table]
Browse...
Match!

Logged in as **math**. Log off.

CrossMatch

Cross-match your data with our catalogue!

Step 1. Upload your data
User table: *Must be valid VOTable or CSV file*
Format of the input file: (•) VOTable ( ) CSV (Comma-Separated Values)

Step 2. Select catalogue table to match with
No tables selected. Click on the table name to select/unselect it.

Step 3. Select cross-match radius
Cross-match radius (degrees): *Must be less than 0.0028*

Step 3. Get the results!
It will return the VOTable with all info from your input and our catalogue for matched objects

Here will be some help on VOTable format etc

Fig.1: Two screenshots of the SAI CAS web interface.

The design of the SAI CAS system

The components of the system:

- Database
- Backend
- Frontend

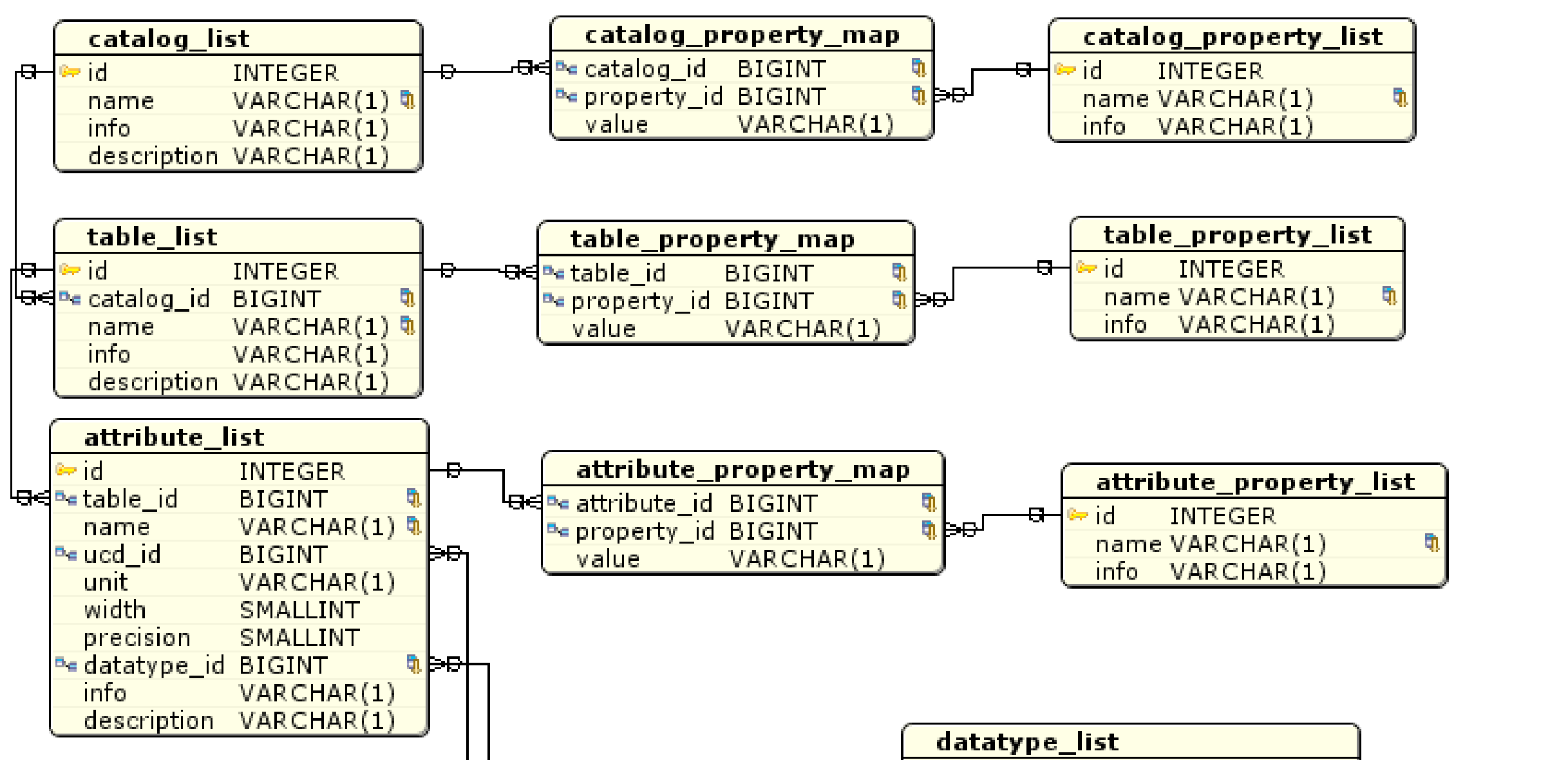


Fig.2: The structure of the metadata storage in the database. It allows to describe in details all the catalogues, tables and columns in the system and effectively produce the output like VOTable.

The main database in the system is PostgreSQL 8.1. The role of the database is :

- It stores all the data from the catalogues
- It stores all the metadata of the catalogues (descriptions, infos, units, UCD's, structure of the catalogues etc.). The structure of the metadata storage is shown on Fig. 1
- It provides separate storages for the user tables and for their metadata.
- It performs the set of spatial queries via Q3C plugin (Q3C support conesearches, polygonal searches, crossmatches, ellipse crossmatches, crossmatches with variable match radius) (http://q3c.sf.net).
- The database is connected with the backend with the JDBC connection.

The Backend is working on Apache Tomcat with Apache Axis. The main role of the backend is to implement all the business logic in it. So its functionalities:

- It talks to the database with JDBC (no other component of the system can talk directly to the DB) (the connections to the DB are organized in the connection pool).
- It provides the set of Web Services to access the metadata of the stored catalogues (to retrieve the list of catalogues, list of tables, columns and all other metadata).
- It provides several Web Services for uploading of new system catalogues and user catalogues.
- It provides the set of servlets for ConeSearch, Crossmatch with user supplied data, Upload of the user's data to the system.
- It provides some of the Skynode WS functionality (metadata queries) (the ADQL queries of the catalogues are being implemented right now...).

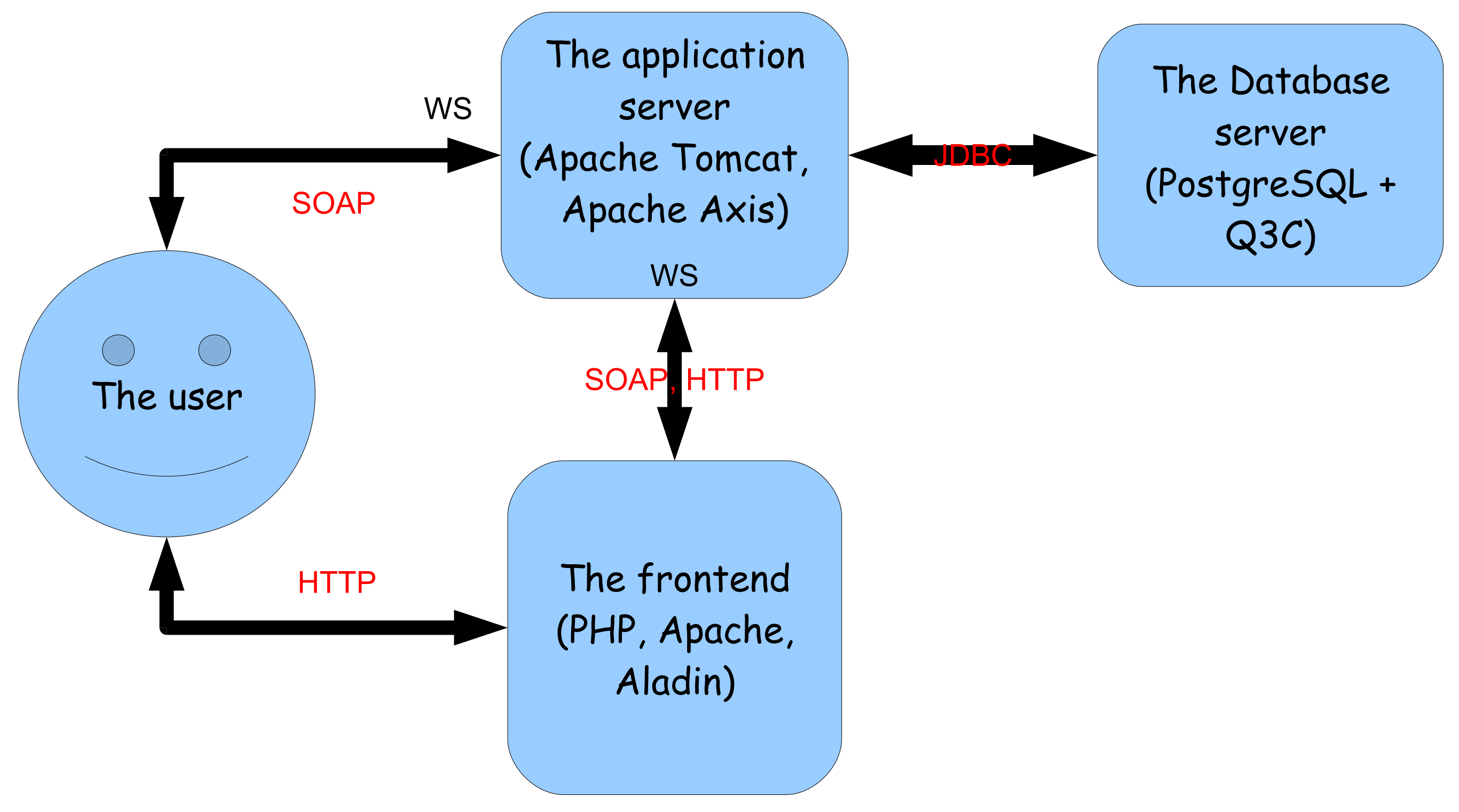

The application server (Apache Tomcat, Apache Axis)
WS
WS
SOAP
JDBC
The Database server (PostgreSQL + Q3C)
The user
SOAP
HTTP
HTTP
The frontend (PHP, Apache, Aladin)

Fig 3: The internal 3-tier structure of the SAI CAS system.

The frontend is the only part of the system directly seen by the end user (it works on Apache web server & PHP). It's role:

- Provide the user friendly interface to interact with SAI CAS
- It allows to browse the catalogues in SAI CAS, to see their metadata, to list tables of catalogues etc, (to do that, the backend calls the metadata WS exposed by the backend)
- It provide the user friendly form interface to perform the conesearches, crossmatches, select input/output formats, upload the data into the system, edit the metadata of the catalogues.

It give the ability to analyze the retrieved results with Aladin.

Sergey Koposov

Acknowledgments: We are grateful to Hewlett-Packard Russia for the donation of the Itanium 2 Database Server and of the 5 Tb data storage.
This work was partially supported grant of Russian Fund for Basic Research

References:
- Q3C: Koposov, Bartunov, 2006, Asp. Conf. Ser. 351, p.753, ADASS XV, http://q3c.sf.net
- SAI CAS: Koposov, Bartunov, IAU Special Session 3, http://vo.astronet.ru/files/sai_cas.pdf