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

S. Koposov[1,2,3], O. Bartunov[2], S. Karpov[4] et al[2].

Russian Virtual Observatory

[1]Max Planck Institute for Astronomy, Heidelberg, Germany
[2]Sternberg Astronomical Institute, Moscow, Russia
[3]Institute of Astronomy, Cambridge, UK
[4]Special Astrophysical Observatory, Nijniy Arhyz, Russia

IVOA Interop. meeting &
Small projects meeting
Moscow 2006

http://vo.astronet.ru
http://vo.astronet.ru/preview/cas

Outline

- Existing data-centers, and catalogue access providers.
- Why we want to build our own Catalogue Access Service
- The requirements for our CAS system
- The importance of the Database for the CAS (PostgreSQL & Q3C)
- The technical realization of CAS
- What has been already done
- Example of work
- TODO

SAI CAS: ConeSearch - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://vo.sai.msu.ru/cas/conesearch.php

PG Hackers BBC Mylinks РБК Lenta ON ЕЖ G AP Э PGp » xmm catalogue

T-Mobile - HotSpot | file:///hom...to_show.xml | SAI CAS: ConeSearch | SAI CAS: Cross-catalo... | Google Mail - Inbox (5)

[logo] SAI VO | SAI CAS

SAI CAS | ConeSearch | About

Catalogs

- denis
  - main [?]
- usnoa2
  - main [?]
- twomass
  - psc [?]
  - xsc [?]
- usnob1
  - main [?]
- gsc1_2
  - main [?]
- tycho2
  - main [?]
  - suppl_1 [?]
  - suppl_2 [?]
- nomad
  - main [?]
- ucac2
  - main [?]

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

Selected tables:

Remove **main** of **Third release of DENIS data (20 September 2005)**

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

Cone Search: Query Show in Aladin!

Third release of DENIS data (20 September 2005)

Table **main**:

Name	Description	Units	UCD	Info
denis	DENIS name, identifier		ID_MAIN	
image	DENIS image number		OBS_ID	The image number is incremented at each image taken during the survey. In case a source results from the merging of two consecutive overlapping images (N and N+1) in the same strip (Flag_mult set to 1), the image number corresponds to the first image N.

http://vo.sai.msu.ru/cas/conesearch.php | Open Notebook | 5 | Proxy: None | Adblock

Storing and Accessing Astronomical Catalogues
Existing projects

- CDS
- OpenSkyQuery
- CASjobs with SDSS

They are all really great! But there are problems:

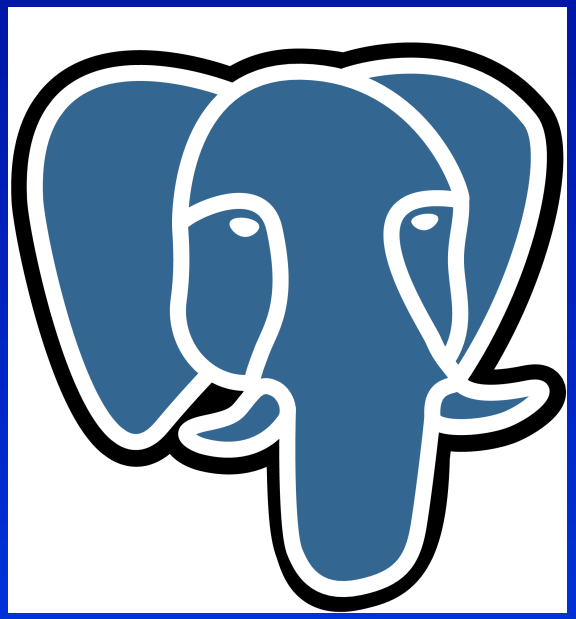
- Not OpenSource
- No good API to work with
- Not prepared for the work with VO
- No possibility of combining private datasets with large catalogues (cross-matches, etc...)

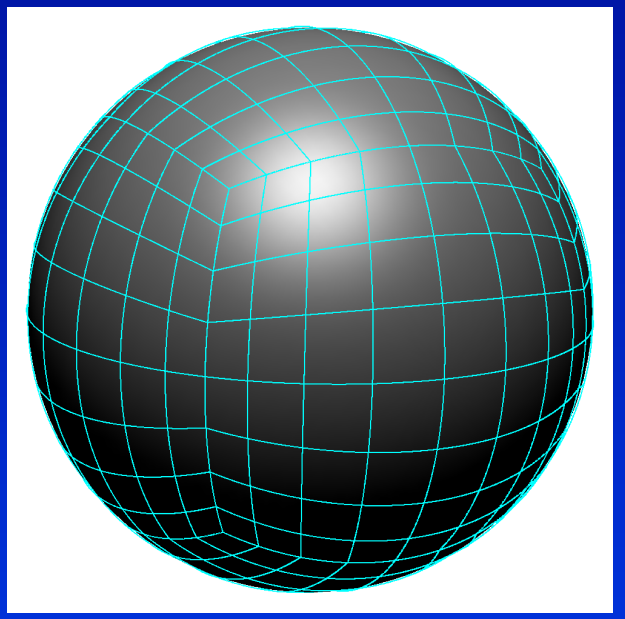
Requirements for the CAS

- Relational storage of the Data
- The ability to attach different metadata to the data (including different VO attributes). That metadata should be also stored in the relational database.
- OpenSource
- API allowing to work with it and allowing put miscellaneous WS on top of it.
- Good architecture (separation of the DB, WS software and frontend).
- The support of spatial queries and cross-matches.
  Easy uploading, editing of the catalogues (through the WS).
- The support for separate user storage area, allowing to work with all datasets.

Importance of the Database for the Catalogue Storage & Access System

- Large Data volumes
- Spatial searches
- Cross matches





PostgreSQL + Q3C (Quad Tree Cube)

Q3C sky indexing scheme for PostgreSQL

- C plugin for PostgreSQL
- Support of cone-searches, ellipse-searches, polygonal searches
- Circle & ellipse cross-matches (with constant cross-match radius and variable cross-match radius).
- Do not require additional columns in the table – only one index.
- Easy queries: SELECT * FROM table WHERE
  q3c_radial_query(ra, dec, 0.001)

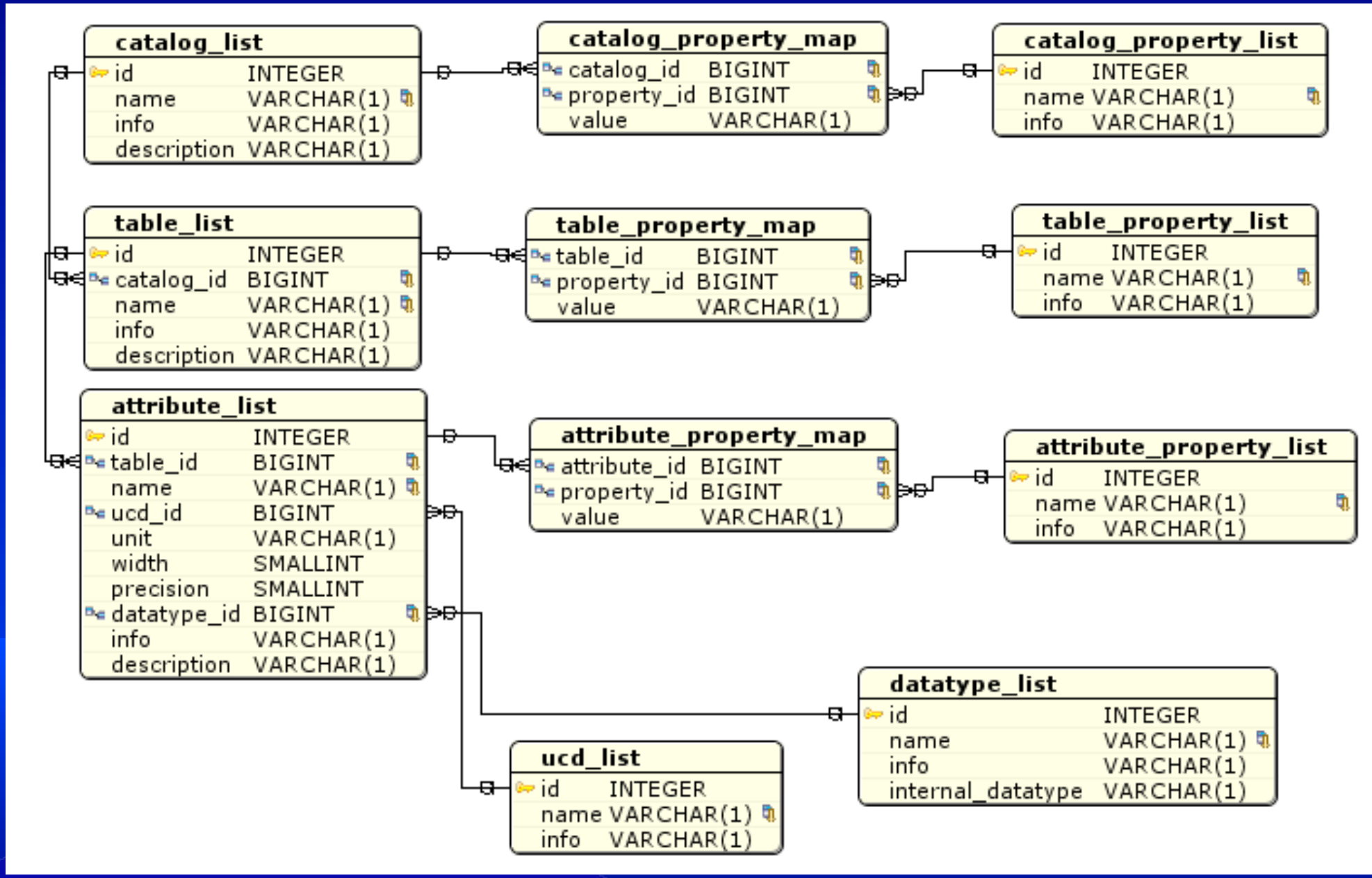
SELECT * FROM a, b WHERE
  q3c_join(a.ra, a.dec, b.ra, b.dec, 0.001)

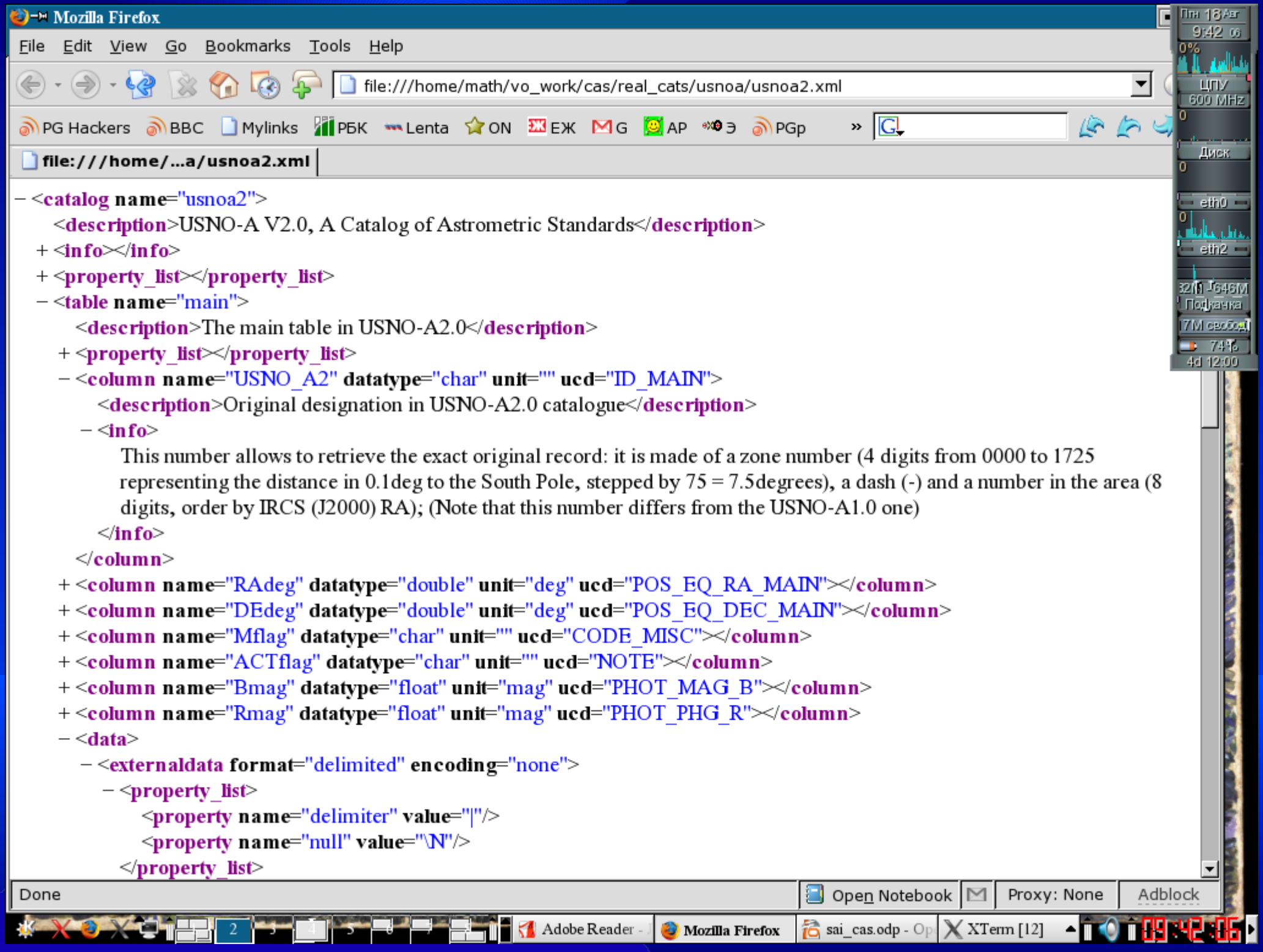
- Extremely fast (cone-searches on catalogues with ~ 10^9 objects take ~ 1 second, cross-matches of 2MASS with USNO-B1 takes ~ 16 hours)

Q3C can be downloaded from
http://q3c.sourceforge.net/

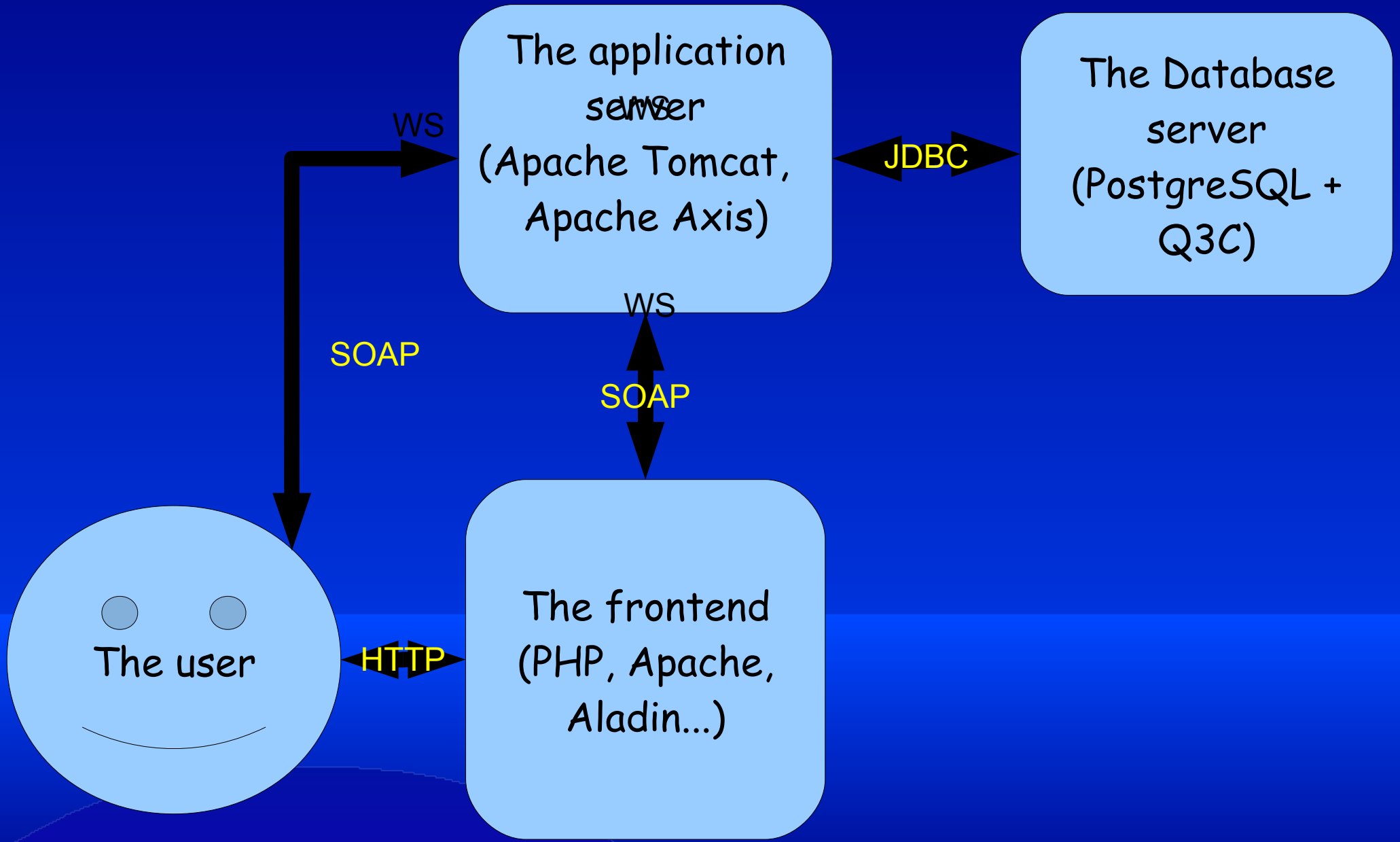
Ref: Koposov, Bartunov, ASPC 2006, vol. 351, p. 735

Structure of the metadata storage in the DB




Mozilla Firefox
File Edit View Go Bookmarks Tools Help
file:///home/math/vo_work/cas/real_cats/usnoa/usnoa2.xml
PG Hackers BBC Mylinks РБК Lenta ON ЕЖ G AP Э PGp
file:///home/...a/usnoa2.xml
- <catalog name="usnoa2">
<description>USNO-A V2.0, A Catalog of Astrometric Standards</description>
+ <info></info>
+ <property_list></property_list>
- <table name="main">
<description>The main table in USNO-A2.0</description>
+ <property_list></property_list>
- <column name="USNO_A2" datatype="char" unit="" ucd="ID_MAIN">
<description>Original designation in USNO-A2.0 catalogue</description>
- <info>
This number allows to retrieve the exact original record: it is made of a zone number (4 digits from 0000 to 1725 representing the distance in 0.1deg to the South Pole, stepped by 75 = 7.5degrees), a dash (-) and a number in the area (8 digits, order by IRCS (J2000) RA); (Note that this number differs from the USNO-A1.0 one)
</info>
</column>
+ <column name="RAdeg" datatype="double" unit="deg" ucd="POS_EQ_RA_MAIN"></column>
+ <column name="DEdeg" datatype="double" unit="deg" ucd="POS_EQ_DEC_MAIN"></column>
+ <column name="Mflag" datatype="char" unit="" ucd="CODE_MISC"></column>
+ <column name="ACTflag" datatype="char" unit="" ucd="NOTE"></column>
+ <column name="Bmag" datatype="float" unit="mag" ucd="PHOT_MAG_B"></column>
+ <column name="Rmag" datatype="float" unit="mag" ucd="PHOT_PHG_R"></column>
- <data>
- <externaldata format="delimited" encoding="none">
- <property_list>
<property name="delimiter" value="|"/>
<property name="null" value="\N"/>
</property_list>
Done
Open Notebook
Proxy: None
Adblock
Adobe Reader
Mozilla Firefox
sai_cas.odp
XTerm [12]

Technical implementation of CAS.



Currently working services

- ConeSearch with output in different formats (VOTable, CSV) through HTTP GET query, but also through the SOAP WS
- Cross-match with user's VOTable
- The miscellaneous WebServices retrieving the information about the catalogues, their info, contents etc.
- The WS of new catalogue creation
- Some authorization WS are beeing built right now
- Basic SkyNode in testing regime

Currently loaded catalogues

- USNO-A2.0 (5×10^8 objects)
- USNO-B1.0 (10^9 objects)
- NOMAD catalogue (> 10^9 objects)
- 2MASS PSC (2.5×10^8 objects)
- 2MASS XSC (2×10^6 objects)
- Tycho 2
- UCAC 2
- GSC 1.2
- 2XMM
- SDSS DR5 is beeing downloaded

Small demonstration of capabilities of SAI CAS

TODO

- Enable the full SkyNode (it requires only wrapping of SkyNode WS calls around already existing WS API of SAI CAS. (the Basic Skynode will be available before ADASS).
- We are finishing the WCS matching web-service using the catalogues to WCS calibrate user images.
- The user's storage and upload of personal catalogues. There is a working prototype, but the authorization mechanisms are still under development.
- Start to enable more VO stuff on top of CAS (registry, SkyNode etc.)

Conclusions

- We presented the new Catalogue Access system the first large project made by Russian Virtual Observatory.
- It already has a collection of the largest existing astronomical catalogues
- You can work with it using the URLs:

http://vo.astronet.ru

http://vo.astronet.ru/cas/preview