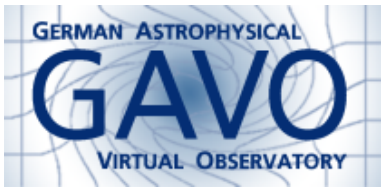
GUMS-10 in the Virtual Observatory

Gabriel Stöckle, Florian Rothmaier,
Stefan Jordan and Markus Demleitner

with contributions from Svea Proft

Astronomisches Recheninstitut Heidelberg

June 22, 2012



Outline

1. Introduction
   - GAVO Data Centre in Heidelberg
   - GUMS-10 in Our Data Centre
   - Data Centre Helper Suite
   - TAPShell
2. Visualising GUMS-10 in the VO
3. Scientific Use Cases
   - Common Proper Motion
   - Find the Brightest Supernova
   - Quasar Studies
4. Summary and Outlook

The GAVO Data Centre in Heidelberg

- accessible at `http//dc.g-vo.org`
- about ~ 13 TB of astronomical data stored at the ARI
  - currently still many terabytes of free disc space available

- data sets cover a large variety of astronomical subjects, e.g.

The GAVO Data Centre in Heidelberg

- accessible at `http//dc.g-vo.org`
- about ~ 13 TB of astronomical data stored at the ARI
  - currently still many terabytes of free disc space available

- data sets cover a large variety of astronomical subjects, e.g.
  - astrometry: catalogues ARIGFH ("ARI's Geschichte des Fixsternhimmels"), PPMX, PPMXL, ...
  - astrophotography: digitised photo plates from Heidelberg and Calar Alto
  - dark matter/quasars: images of lensed quasars from various sources, e.g. from the Maidanak and the Apache Point Observatory

The GAVO Data Centre in Heidelberg

- accessible at `http//dc.g-vo.org`
- about ~ 13 TB of astronomical data stored at the ARI
  - currently still many terabytes of free disc space available

- data sets cover a large variety of astronomical subjects, e.g.
  - astrometry: catalogues ARIGFH ("ARI's Geschichte des Fixsternhimmels"), PPMX, PPMXL, ...
  - astrophotography: digitised photo plates from Heidelberg and Calar Alto
  - dark matter/quasars: images of lensed quasars from various sources, e.g. from the Maidanak and the Apache Point Observatory
  - light pollution measurements: data from the "Lightmeter Network"
  - extrasolar neutrino candidates observed by the AMANDA-II and IceCube neutrino telescopes
  - computer simulations: "TheoSSA" interface with theoretical spectral energy distributions; **GUMS-10** data set

GUMS-10 in GAVO's Data Centre

- $\sim$ 730 GB GUMS-10 input data
- six output tables created by our data centre:

Table Name	N_{obj}	Query Performance
`gums.mw`	$\sim 2.1 \times 10^9$	very time-consuming
`gums.galaxies`	$\sim 3.8 \times 10^7$	time-consuming
`gums.lmc`	$\sim 7.6 \times 10^6$	fast
`gums.smc`	$\sim 1.3 \times 10^6$	fast
`gums.quasars`	$\sim 1.0 \times 10^6$	fast
`gums.sn`	150	very fast

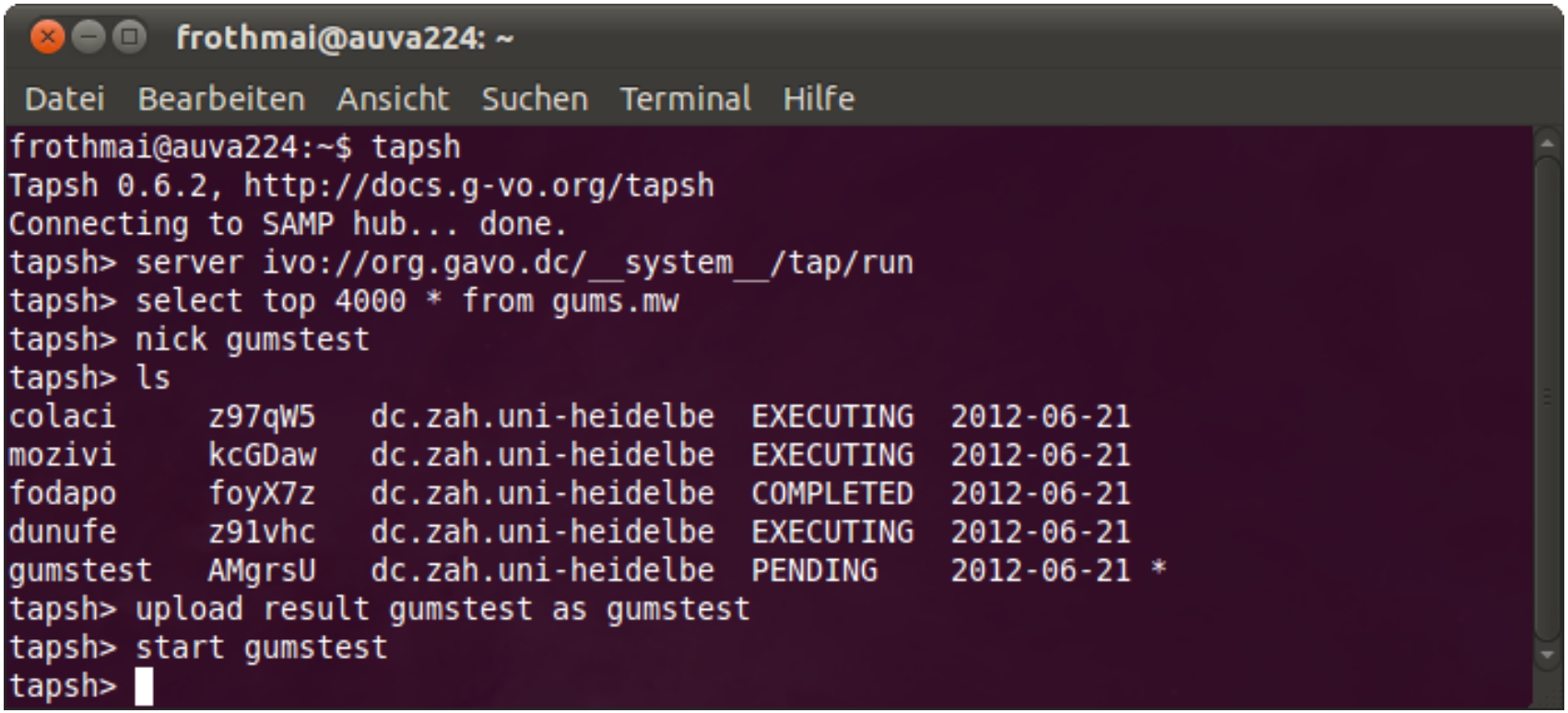
- all these tables are available by ADQL ("Astronomical Data Query Language") queries and by the TAP ("Table Access Protocol") service

Data Centre Helper Suite ("DaCHS")

- data publishing infrastructure for the VO providing...
  - pre-defined grammars for mapping input data to output tables
  - support for data ingestion
  - tools for metadata handling
  - support for many VO protocols and standards, e.g. TAP and Simple Cone Search (SCS)
- mostly written in Python the DaCHS software acts on PostgreSQL databases

TAPShell (I)

- TAP client for the shell ("command-line")



- very nice tool for larger ADQL/TAP queries

- user-friendly syntax

TAPShell (II) – How to Use?

- start `TAPShell`:
  - `tapsh`
- choose a server, e.g.:
  - `server ivo://org.gavo.dc/__system__/tap/run`
- enter your query, e.g.:
  - `select top 10000000 * from gums.mw`

TAPShell (II) – How to Use?

- start `TAPShell`:
  - `tapsh`
- choose a server, e.g.:
  - `server ivo://org.gavo.dc/__system__/tap/run`
- enter your query, e.g.:
  - `select top 10000000 * from gums.mw`
- get a status list (including job nicknames and "real" names)
  - `ls`

➥ status of your job will be `PENDING`

- start your job
  - `start <nickname>`

➥ status of your job will change to `EXECUTING`

TAPShell (III) – Scripting

- TAPShell supports a limited form of scripting
  - ➥ write your queries in files
- write in the first line
  - `#!/usr/bin/env tapsh`
- make your script executable... and run it

TAP-Query Workflow in TOPCAT

workflow example:

- open TOPCAT
- click on "VO" $\rightarrow$ "Table Access Protocol (TAP) Query"
- enter `http://dc.g-vo.org/tap` (GAVO's TAP service) in the "TAP URL" field and press "Enter Query"
- use the Astronomical Data Query Language (ADQL) to express your query

Visualising GUMS-10 in the VO

see Gabriel's slides

GAIA-Use Case: Common Proper Motion

- Scenario: "I would like to retrieve all pairs of stars with common proper motion."
  - cf. `http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess/`

➥ use VO tools

GAIA-Use Case: Common Proper Motion

- Scenario: "I would like to retrieve all pairs of stars with common proper motion."
  - cf. `http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess/`

➥ use VO tools

- ADQL query:

```
SELECT
  a.sourceid, a.mualpha, a.mudelta,
  b.sourceid, b.mualpha, b.mudelta
FROM (
  SELECT * FROM gums.mw
  WHERE 1=CONTAINS(POINT('ICRS',alpha,delta),
    CIRCLE('ICRS',10,15,1))) AS a
JOIN gums.mw AS b ON (
  a.sourceid<b.sourceid
  AND SQRT(POWER(a.mualpha-b.mualpha,2)
    + POWER(a.mudelta-b.mudelta,2))<1e-7)
```

Use Case: Brightest Supernova (I)

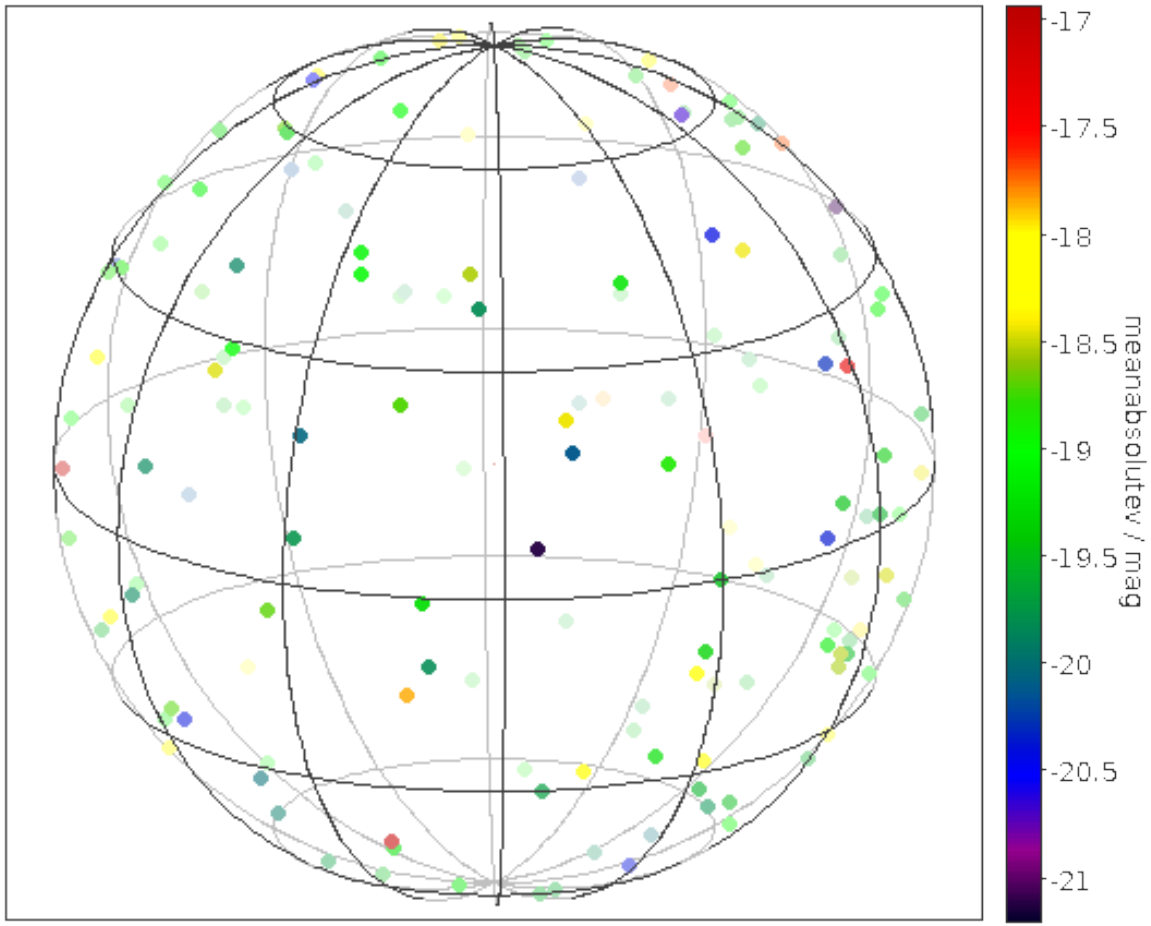
- Scenario: "I want to find the brightest supernova in the database."

➥ do it in TOPCAT by using TAP

```
SELECT * FROM gums.sn
ORDER BY meanabsolutev
```

Use Case: Brightest Supernova (II)

- Scenario: "I want to find the brightest supernova in the database."



Use Case: Quasar Studies (I)

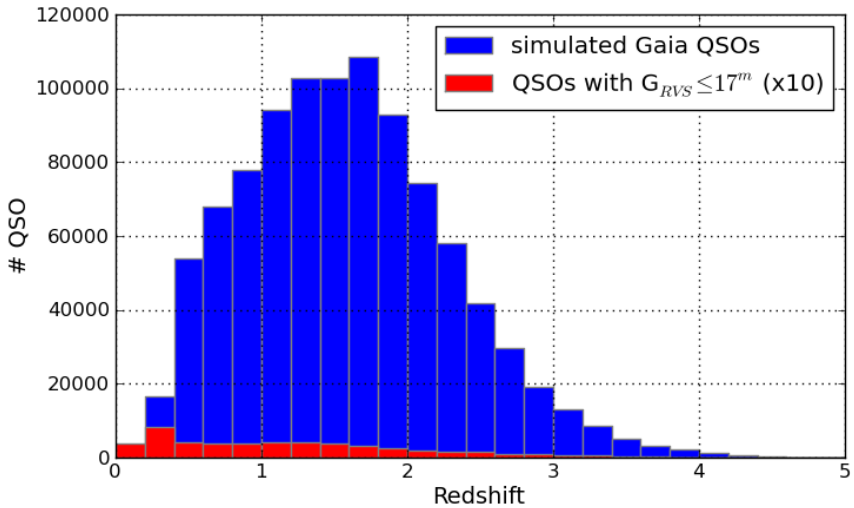
- quasar simulation studies with the Radial Velocity Spectrometer (RVS) by S. Proft
- RVS operating mode: creation of high resolution spectra for observations
  - in the $847 - 874$ nm region
  - up to $G_{\mathrm{RVS}} = 17^m$
- quasars show broad emission lines in their spectra which may match the RVS range (according to z)

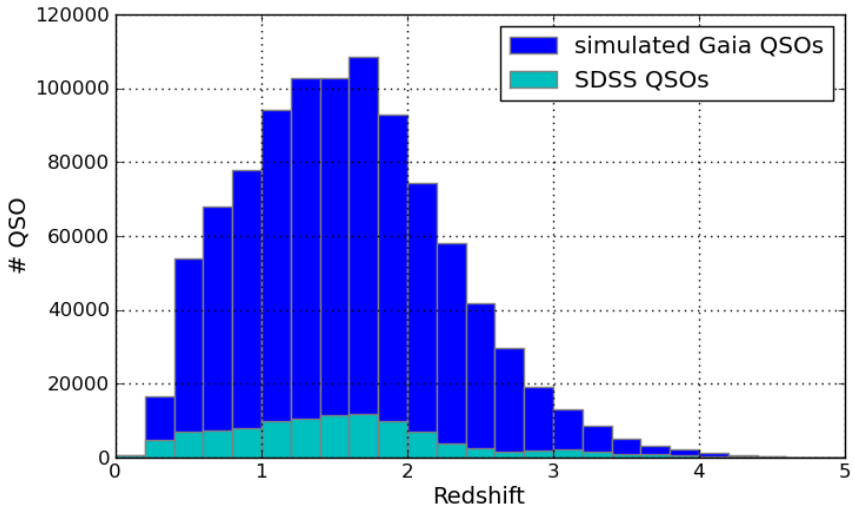
➥ find all quasars with $G_{\mathrm{RVS}} \leq 17^m$

```
SELECT * FROM gums.quasars
WHERE maggrvs<17
```

Use Case: Quasar Studies (II)

- quasar simulation studies with the Radial Velocity Spectrometer (RVS) by S. Proft





↪ looking forward to the data simulation which takes the instrument into account

Summary and Outlook

What you can do...

- DaCHS helps you to
  - bring your data into the GAVO Data Centre
  - handle your database queries

- TAPShell lets you
  - access a variety of data services in the VO
  - perform sophisticated ADQL/TAP queries (that may take some time but it's feasible...)
  - send your query results ("votables") to other clients, e.g. TOPCAT, in order to visualise, analyse your data sets

Summary and Outlook

What you can do...

- DaCHS helps you to
  - bring your data into the GAVO Data Centre
  - handle your database queries
- TAPShell lets you
  - access a variety of data services in the VO
  - perform sophisticated ADQL/TAP queries (that may take some time but it's feasible...)
  - send your query results ("votables") to other clients, e.g. TOPCAT, in order to visualise, analyse your data sets

What you can't do...

- not all ADQL/TAP queries can be handled by every TAP client
  - limits on the number of rows to be processed

➥ choose a VO client appropriate to execute your type of query

Thank you...

...for your attention!!

References

A. C. Robin et al.: Gaia Universe Model Snapshot, A statistical analysis of the expected contents of the Gaia catalogue, arXiv:1202.0132v2 [astro-ph.GA].

S. Jordan et al.: Some ideas for GAP/CU9 contributions from the ARI. Technical note (2012).

GAVO Data Center, *Zentrum für Astronomie Heidelberg*, http://dc.zah.uni-heidelberg.de.

GAVO web page offering a short course on ADQL and TAP, http://docs.g-vo.org/adql/html/.

TOPCAT – Tool for OPerations on Catalogues And Tables, http://www.star.bris.ac.uk/~mbt/topcat/.

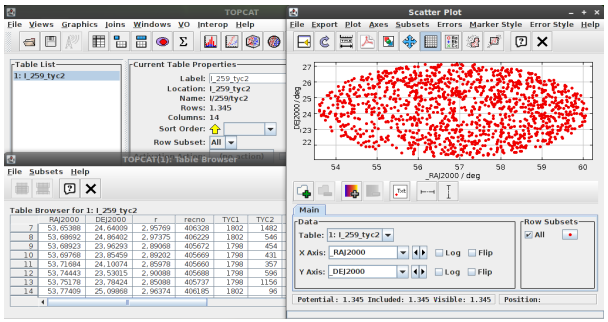
What is the Virtual Obervatory (VO)? (I)

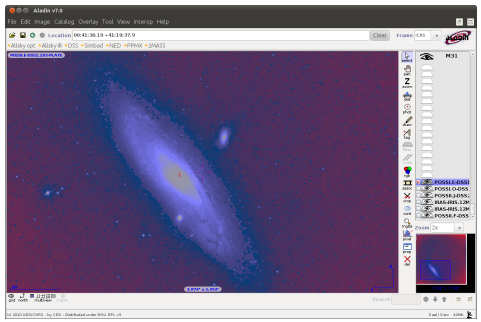
- international astronomy project
  - 19 member organisations on 5 continents

What is the Virtual Obervatory (VO)? (I)

- international astronomy project
  - 19 member organisations on 5 continents

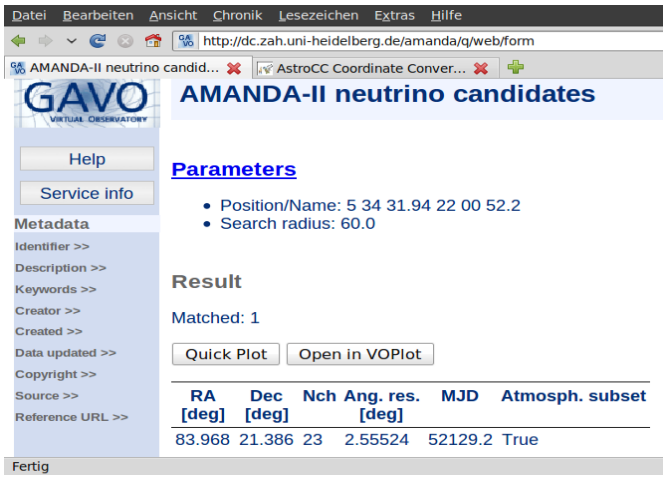
- collection of interoperating software tools which aim at
  - providing global electronic access to astronomical data archives
  - visualising and analysing data
  - comparing data sets taken by different telescopes
  - combining data from different wavelength regions





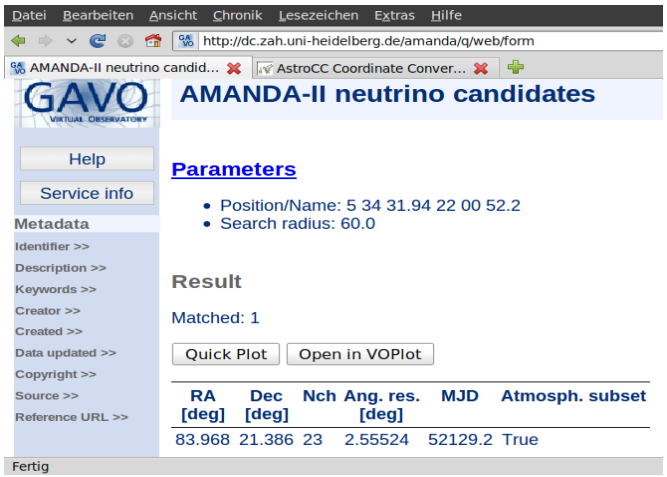
What is the Virtual Obervatory (VO)? (II)

- collection of data centres which ensure that
  - new data sets get published
  - published data are preserved
  - data is properly described and accessible



What is the Virtual Obervatory (VO)? (II)

- collection of data centres which ensure that
  - new data sets get published
  - published data are preserved
  - data is properly described and accessible



- development and implementation of standards (similar to the web) to
  - facilitate data access and storage
  - ensure the unambiguous communication between applications