Fig. 1

Fig. 2

1. The Virtual Observatory

Your science in the Net

Markus Demleitner *(msdemlei@ari.uni-heidelberg.de)*

- Why VO?
- What does it look like?
- Getting started
- Contribute!

(vgl. Fig. 1)

(vgl. Fig. 2)

2. Why VO?

Fringe Benefits

- Locate data useful to you (like ADS search)
- . . . relatively easily
- Access it (like PDF download)
- . . . with as little manual work as possible

∩ If you consider that "data" usually is something quite complex with a complex history that needs to be known to some extent to make the data useful, it should be clear that even such a program is quite ambitious. Add in the century-old troubles of curation, preservation, etc., and you have
∪ a rather messy picture. But this is the right time to tackle this mess, because of:

The real #1 reason

Because we must – CCDs and CPU grow faster than the bandwidth at your desk: Data can no longer move to programs but the programs must go to the data:

<div align="center">"Smarts must follow data"</div>

∩ Basically, SDSS has shown how such a thing can work, and the VO is to some extent an effort to generalize what SDSS offers to a world of multiple data centers that have to communicate and offer somewhat consistent interfaces.

Whatever one thinks of the current VO effort, a system allowing some kind of "remote job
∪ submission" in a standard way is very much needed.

3. Move the Mouse

Let's see the (fledgling) VO in action; example: we suspect some interesting property of infrared SEDs of carbon stars.

We will be using:

- VO Desktop and Topcat (desktop applications by Astrogrid UK)
- The Registry (an index of the services in the VO)
- A "custom" service at GAVO's data center (as a shortcut)

∩ To reproduce the demo, try:

1. Start VODesktop, New smart resource "Title contains carbon" and capability is Cone.
2. Choose "Carbon stars from the Hamburg. . .", inspect metadata, "Query"
3. Get them all (cone search around 0,0 with 180 degrees radius)
4. Start topcat, send table over, view it, plot on sky (UCDs!), show column metadata
5. Now add 2MASS magnitudes: Click multi query – in the near future, choose a service and run the multicone; right now, go by foot:
6. Go to http://vo.uni-hd.de, select "Custom Uploading Crossmatcher", paste the "Table URL from the Multi Position search window into the Remote URL field, select twomass.data in the crossmatcher's Target Table
7. The result page has a form at the very bottom. Select VOTable as Output format, open an "Open" dialog in topcat and drag the "Result link" from the crossmatcher there.
8. Inspect metadata, etc., again. For kicks, do a spherical plot again, add a radius (say, Jmag) and a color (say, Kmag-Hmag).

∪
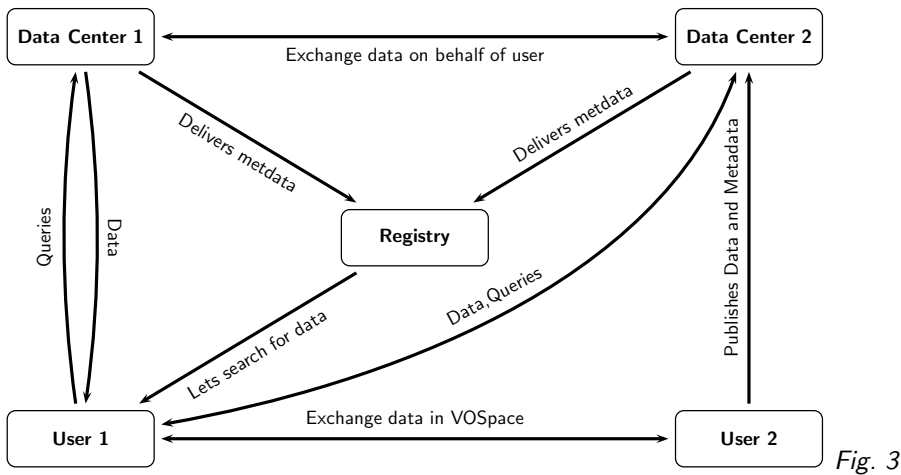
Fig. 3

4. The VO is . . .

. . . a network of interoperating computers.

(vgl. Fig. 3)

∩

So, the user can ask a "Registry" (of the services and data within the VO) for data relevant to her and can then send queries to the data centers (or other service providers) to be returned data.

The Registry knows of services and data in data centers because they send their metadata to the registry.

The data centers may exchange information between themselves if the user told them to do so (this happened in our little example when the crossmatcher retrieved the data directly from vizier).

Users can, directly or indirectly, exchange data using VO protocols as well (but don't throw away your mail client just yet. . .).

And, of course, users can and should publish their data to the VO. There are tools that let them do this without the help of a data center (e.g., Astrogrid DSA), but it's usually more convenient and sustainable to just dump data to people paid for worrying about this mess.

∪   The good news: Ideally, you don't notice too much of these complexities.

5. Representing Data

Within the VO, data is preferably represented in **VOTable**s. VOTables are basically like FITS tables, but they are more structured, a bit more flexible, and it's XML.

To let machines figure out data (to some extent), VOTables for each column may declare

- **UCD**, Unified Content Descriptors, for saying "what is this?", and

- **utype**s for saying "what's it for?".

For example, a column could have the UCD pos.eq.ra to say it's a right ascension. Another column could have phot.mag;em.opt.U to say it's a magnitude roughly in the area of Johnson U. Thus, you or a computer can decide from table descriptions what kind of physical quantity is included in tables. This information is available in the Registry.

utypes exist with respect to specific data models, i.e., descriptions of domains. Such a data model might define concepts like "center of light" (in a data model for extended sources) or "minimum magnitude" (in a data model for variable stars). One has to admit that not terribly many of these data models have been successfully defined so far.

A popular program to manipulate VOTables (and convert them to other formats) is topcat.

6. Transferring Data

Much of the VO is about transporting data between computers. To make this work, one has to define protocols.

Some useful protocols you may want to use include:

- SCS – search in object catalogs (VODesktop)

- SIAP – search for images (VODesktop, Aladin)

- SSAP – search for spectra (splat, VOSpec)

- TAP – remote querying (soon VODesktop)

- PLASTIC – exchange data between programs

- VOSpace – storing and transmitting data server-side

You should not need to know too much about these protocols, except that knowing what they are good for will help you locating relevant services.

TAP and the language transported by it has been highly contentious but is being passed into a "Recommendation" (which, in the VO world, means: Go ahead and write software) right now. Once there is good tool support and robust service implementations out there, the VO will come a lot closer to living up to the challenge of "transporting smarts".

7. ADQL

The Astronomical Data Query Language is just SQL with some extensions. For the time being, it will probably be the main vehicle to transmit "programs" – it is TAP's main language, for starters. Queries look like this:

```
SELECT * FROM (
  SELECT ALL q.name, q.raj2000, q.dej2000, p.alphaFloat,
    p.deltaFloat, p.vmag FROM (
    SELECT TOP 100 raj2000, dej2000, name, z
    FROM veronqsos.data
      WHERE notRadio!='*'
      AND z BETWEEN 0.5 AND 1
      AND dej2000<-40) AS q JOIN
    ppmx.data AS p ON (1=CONTAINS(
      POINT('ICRS', q.raj2000, q.dej2000),
      CIRCLE('ICRS', p.alphaFloat, p.deltaFloat, 0.3))))
  WHERE vmag BETWEEN 10 and 11
```

This is a query that could find guide stars for certain quasars for a telescope sitting on the south pole: Select radio-loud quasars with a redshift between 0.5 and 1 and a declination that is guaranteed to be well observable from the south pole, and then see if there are stars with a visual magnitude between 10 and 11 in a 0.3 degrees radius around them.

Such queries look horrible when you first see them, but so does FORTRAN. From there, SQL is somewhat of a progress.

Anyway, this is the really bad news of this talk: To do *really* interesting things in the VO, learning to write ADQL queries will probably be necessary.

Then again, even without ADQL, there's much to discover. . .

8. Next steps

- NVO Registry Interface[1] – Discover VO services from your Browser
- DataScope[2] – Cone search from your browser
- VO software for your desktop[3] – Collected by EuroVO
- GAVO data center[4] – Publish your data to the VO
- Check out this afternoon's workshop.

. . . Thanks for your attention

---

[1] http://nvo.stsci.edu/vor10/index.aspx
[2] http://heasarc.gsfc.nasa.gov/cgi-bin/vo/datascope/init.pl
[3] http://www.euro-vo.org/pub/fc/software.html
[4] http://vo.uni-hd.de