Practical Interoperability in the Virtual Observatory

Markus Demleitner
Universität Heidelberg, Astronomisches Rechen-Institut, Mönchhofstraße 12-14, Germany
msdemlei@ari.uni-heidelberg.de

German Astrophysical Virtual Observatory
Federal Ministry of Education and Research

The Virtual Observatory (VO) is an international effort to run and develop a federated data infrastructure in Astronomy that is held together by a set of data and protocol standards. Consisting of a Registry, some 20000 interoperable services (which roughly match data collections) comprising hundreds of millions of datasets (spectra, images, and the like) and hundreds of billions of table rows, and a set of clients and libraries consuming these services, it is widely used in the astronomical community. In this poster, we will give a very short overview of what the VO is, how it works, and why it was built the way it is.

What's a VO Service?

VO creed: Users should not normally interact with services – machines should do that on their behalf.

This is because humans won't fill in 100 different forms on 100 different web pages. In contrast, a computer program ("**client**") can easily query 100 different services as long as they all share parameters and semantics.

A **service** thus is a network-accessible endpoint defined by our **standards**, giving access to one or more data collections.

Suppose a user requests "images of Barnard's star in x-rays". This is how this proceeds in the VO:

1. A client asks a searchable registry: Give me resources that
   - serve images.
   - have data in the x-ray part of the spectrum.
   - have data around $\alpha = 269.45$, $\delta = 4.693$ (that's where Barnard's star is these days).
2. The Registry responds with metadata for the services matching these criteria.
3. The client now goes to each service returned and asks it for data...
   - covering the position $\alpha = 269.45$, $\delta = 4.693$,
   - intersecting the spectral range $0.1 \cdots 120$ keV of photon energy.
4. Each server responds with one set of metadata per matched image. The client turns this into some representation for the user.
5. The user picks images based on the metadata (e.g., observation date, sensitivity...).
6. The client retrieves image (or parts of them) and makes them available for further processing.

Why no single "Image Discovery Platform"?

Well, nobody would pay for it, and they wouldn't want their neighbour to host it.

More importantly: With multiple **interoperable** (i.e., they can be used in a well-defined, uniform way by machines) services the VO can grow from the edges: **Users control** their end of processing, operators can adapt services their needs and **evolve the standard**. No single part can dictate what happens.

In Action

The figure on the right shows a VO session: Using 1 TOPCAT as a TAP client, a query in ADQL 2 was used to retrieve information on colour, brightness, and distances of objects in the vicinity of the beautiful galaxy M101, obtained using space astrometry. This data is then displayed in tabular 3 and graphical 4 form.

In the plot, the user spotted an interesting – e.g., outlying – object and selected it. TOPCAT was configured to broadcast the position of the currently selected object to all VO clients on the desktop, which made Aladin 5, a different client specialised in dealing with images, display the vicinity of the object using archival images, centered on the object of interest 6. In this case, the imagery is for visible light, but many other sorts of data are available.

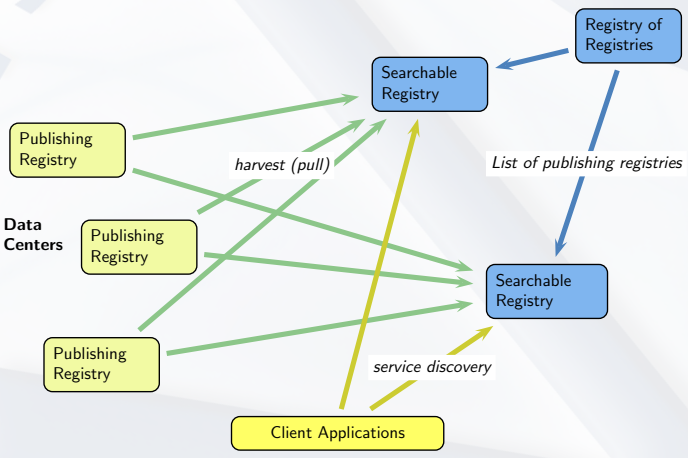
Aladin was then used to discover spectra 7 in the field shown. It renders the VO Registry in its discovery tree 8. Using the SSAP standard, services actually providing data in the field shown have been highlighted, and one particular service (in this case, providing time series letting the user investigate variabilty behaviour) has been queried. Positions where data is available are marked in Aladin's image zone.

Note that none of the services and clients used have any special knowlege of each other except that they know what standards they implement. This is, in the end, the meaning of the term interoperability: Users freely assemble components provided by different authors and data centers.

Federated Registry

Service discovery happens using **searchable registries**. While anyone can run one, in practice there are three major operators (GAVO, ESA, STScI) that clients can go to as they choose.

These searchable registries get their data by OAI-PMH harvesting **publishing registries** operated by the data providers. They know where to go because all (currently about 45) publishing registries are listed in the one central infrastructure of the VO, the **RofR** (Registry of Registries) – that in itself isn't critical infrastructure as the searchable registries chug along nicely even if the RofR fails for a while (no new publishing registries could come online for that time, though).



Standards

The International Virtual Observatory Association defines the standards required to make all this work. For instance:

Searching for data: Images (SIAP), spectra (SSAP), objects (SCS), spectral lines (SLAP), generic datasets (ObsCore).

Remote manipulation: SODA lets you do cutouts, rescaling, etc., to avoid pulling data you don't need.

Interacting with databases: Access using TAP, common query language ADQL.

Formats: Table exchange using VOTable, complex spherical geometries with MOC, multiscale images with HiPS.

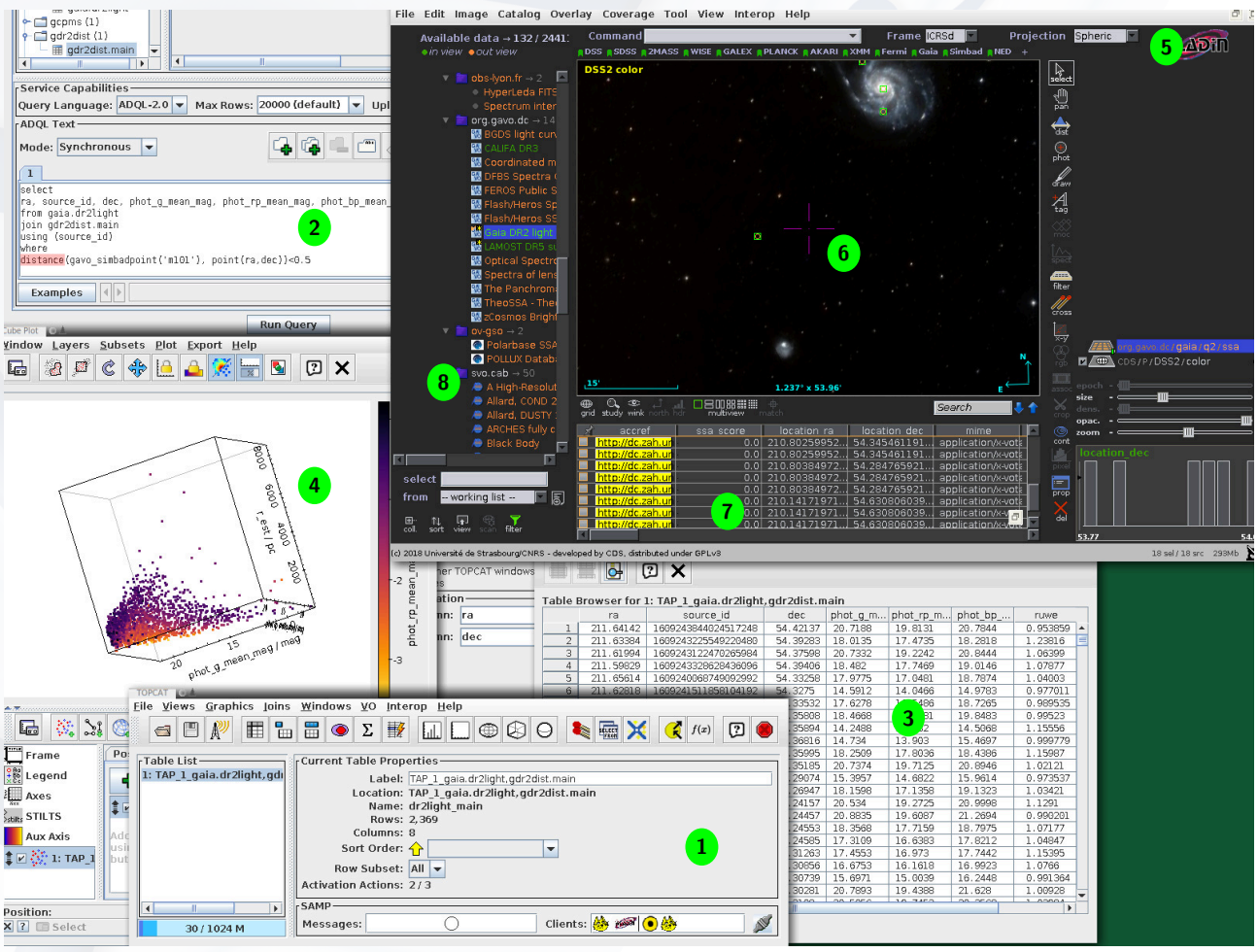
Registry: Registry Interfaces for the architecture, VOResource, VODataService, TAPRegExt, SimpleDALRegExt for the metadata format, RegTAP for how to search it.

Semantics: Light semantics of physical quantities (UCD), Unit syntax, Vocabulary maintenance.

SAMP: Assembling complex environments from simple building blocks.

Does it need to be that complex? Well... almost. Federated data is difficult.

All the standards are available at: `http://ivoa.net/documents`



Further Reading

Much material on the VO is a few clicks away from `http://ivoa.net`; for instance, a directory of VO-enabled software is at `http://ivoa.net/documents`. Data center operators might be interested in `https://wiki.ivoa.net/twiki/bin/view/IVOA/WebHome?topic=PublishingInTheVO`.

Two special volumes on the VO were published in Astronomy and Computing, volumes 10 and 11.

`http://dc.g-vo.org/VOTT` lists many tutorials, courses, and worked-out use cases for VO techniques (and is itself a Registry client).

Acknowledgement: This work was supported by the e-inf-astro project (BMBF FKZ 05A17VH2).