Fig. 1

1. The Virtual Observatory

Markus Demleitner
msdemlei@ari.uni-heidelberg.de

- What is the Virtual Observatory?
- Findable?
- Accessible?
- Interoperable?
- Re-usable. . . Reproducable? I'm focusing at the Reproducability here because at least in Astronomy I think that's the biggest challenge in that field.

(cf. Fig. 1)

2. What's the VO?

It is. . .

not a website ("platform"),

not a bunch of websites,

not a programme that does all things astronomy.

Instead. . .

3. The VO is...

Standards for finding, accessing, using, and describing data (more on which in a moment)

plus

$\sim$ 50 data centers worldwide adhering to these standards (which includes almost all the major players like ESO, NASA, ESA, etc)

plus

a few volunteers operating some $\pm$ central infrastructure (these are things like searchable Registry endpoints, our document repository, and the various bodies working on the standards)

plus

authors of client software, libraries, and web pages making these resources available to astronomers and the public (you may have heard of TOPCAT and Aladin on the desktop, pyVO and STIL as libraries, or ESA Sky or Aladin lite in the browser).

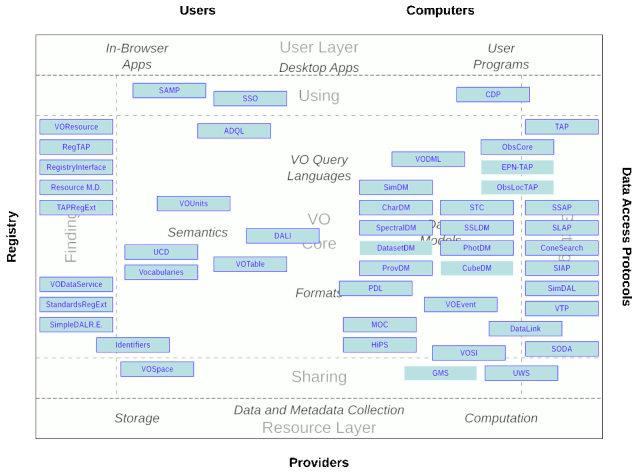

Fig. 2

4. Standards

(cf. Fig. 2)

Yes, that's quite a mouthful – don't bother trying to understand what all this is for at this point. We will be happy to provide a comprehensive introduction to our ecosystem in a much longer workshop.

Dealing with scientific data is a complicated business.

5. Mapping to FAIR

The IVOA standards landscape maps rather nicely to the FAIR principles.

Specifically. . .

(cf. Fig. 3)

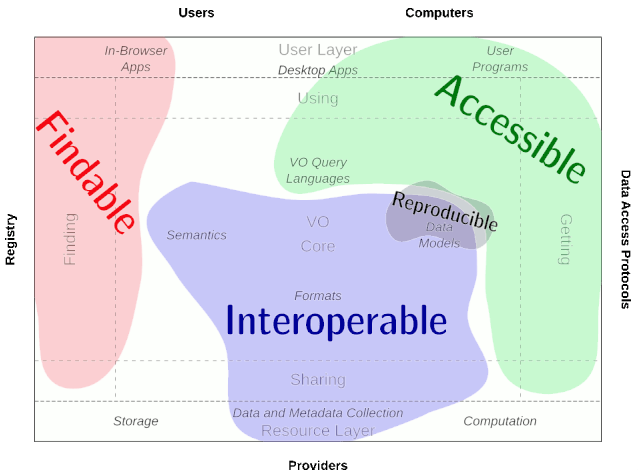

Fig. 3

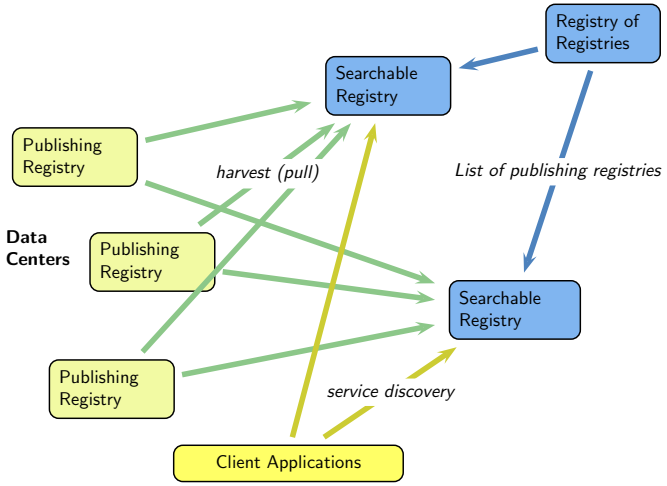

Fig. 4

6. Findable

To find resources, the VO has the Registry.

Conceptually, that's just a big bunch of detailed metadata.

Technically, data centers operate publishing registries that are harvested by full registries which are then queried by client programmes.

(cf. Fig. 4)

7. Accessible

In the VO, people access data by first discovering services that *might* have data they want and then have a machine query these services.

Query protocols can be

- "typed": SIAP for images, SSAP for spectra, SCS for catalogue data...
- "untyped": ObsTAP for observational data, SimDAL for theory data... Most of these are based on transporting real database queries (TAP/ADQL).

8. Interoperable

Interoperability has many faces. Among them:

- Rich metadata in VOTable
- Light semantics to mark up "physics" in UCDs
- Machine-readable unit strings
- Agreed-upon vocabularies, possibly with hierarchy
- Data models... have been painful for us so far
- Authentication and Authorisation become really tricky (who runs the central parts?)
- ...and lots more

9. Reproducible

That's a hard one, depending on what one is aiming for.

Straightforward answer: Provenance DM, to at least have a chance to trace problems when they surface.

Takeup is... an uphill battle. The underlying problem is as usual with FAIR: it's work that the original data provider typically doesn't profit from very much, at least not during the initial data exploitation, which is what they usually really care about. And properly doing Provenance is a lot of work.

10. That's it

...unless you're curious about just *why* we're doing things in this way.

In case you haven't touched the VO yet, try any research-level tutorial from https://dc.g-vo.org/VOTT.

Thanks.

11. Why?

Why has the VO been built this way? In the world out there, everyone does "platforms", no?

Well...

- Funding: Nobody would fund one central platform. And having multiple platforms interoperate is at least as hard as doing it right.
- User freedom: As long as you're building your code against standards, you can easily switch from one resource to the next; and you decide what to write with which tools.
- Operator efficiency: No re-implementing the same functionality on dozens of platforms; rather, implement it in re-usable clients and libraries that new services can use right ways.
- Evolvability: People with good ideas can extend the system "from the edges".