Fig. 1

1. Making product-type a Vocabulary

Markus Demleitner
msdemlei@ari.uni-heidelberg.de

SimpleDALRegExt needs a vocabulary of product types (image, spectrum, ...).

Let's use Obscore's product_type!

However: We want a strict concept tree (see Vocabularies 2).

That needs a little work.

(cf. Fig. 1)

2. Guiding Use Case

```
SELECT TOP 30 access_url
FROM ivoa.obscore
WHERE 1=gavo_vocmatch(
  'product-type',
  'spectrum',
  dataproduct_type)
```

should return SEDs as well.

3. Baseline

The vocabulary originates in Obscore. To see where I'm making the concepts a bit more precise (and mutually distinct), cf. page 17 of

https://ivoa.net/documents/ObsCore/

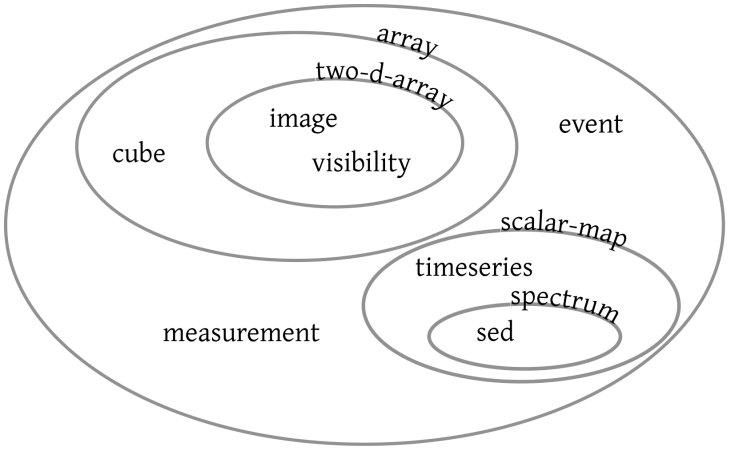

Fig. 2

4. Proposed Concept Map

(cf. Fig. 2)

5. Scalar Maps

Problem: What's a dynamical spectrum (i.e., a time series of spectra)? A `#spectrum`? A `#time-series`?

Realisation: We cannot have per-axes terms for things with more than 1D without solid prioritisation. That is, we would have to say "time before spectrum" or something like that. Perhaps we will want to do that some day; but for now it's not clear there is a use case for that.

Spectrum and time series clients right now quite certainly expect scalar values, too.

Hence, let's define:

`#scalar-map`: Data that maps one scalar to another scalar.

6. Sorts of scalar maps

- `#spectrum`: Flux or magnitudes given as a function of a spectral coordinate. `#spectrum` ⊂ `#scalar-map`
- `#sed`: A spectral energy distribution, i.e., a set of calibrated fluxes over a typically wide part of the spectrum. An SED is usually produced by combining data from multiple observations. `#sed` ⊂ `#spectrum`
- `#timeseries`: Any sort of scalar observable given as a function of time. `#timeseries` ⊂ `#scalar-map`

7. Arrays

There is a fundamental distinction between relational (event, measurement, scalar maps) and array-like data.

At least for the arrays we ought to make clear what we mean by array-like in a common top-level term:

`#array`: A dataset primarily consisting of numbers in a rectangular grid addressed by integer indices; the number of such indices is called the array's dimension here. A single data product may contain more than one such array, usually all of the same size (e.g., for per-pixel error estimates).

There is a bit of a trap here in that IRAF has traditionally represented its spectra as 1D FITS image arrays. Ill-meaning persons could construe this as "it's not a spectrum, it's an array". If this were legal text, we would have write some exception clause for that into the definition. Given that nobody has doubts that these are spectra, maybe we can agree that's a serialisation oddity?

A related issue is that in FITS (and, for instance, astropy), relational data is conventionally represented in record arrays; but these, at least, are not arrays in the sense presented here.

8. 2D data

I'm not aware of useful and true 1D arrays in obscore right now. So, we'll start with 2D arrays.

`#two-d-array`: Array-like data with two dimensions. `#two-d-array` ⊂ `#array`

We already have two narrower terms:

`#image`: Fluxes (calibrated or uncalibrated) as a function of two spatial coordiates. `#image` ⊂ `#two-d-array`

`#visibility`: An interferometric radio observation given in Fourier space. `#visibility` ⊂ `#two-d-array`

Slit spectra, dynamical spectra, etc. would go here, too. But before inventing terms for them, I think we should see who wants to use them for what.

9. Cube

For all arrays with > 2 dimensions, there is just one term:

`#cube`: An array with 3 or more axes, e.g., a spectral image cube, a polarization cube, a full Stokes radio data cube, a time image cube, etc.

I expect there will be narrower terms later; but let's first collect use cases.

10. Relational Types

`#event`: A collection of some sort of observed events, such as high-energy particles observed. An event is typically characterised by spatial, spectral, and time information.

and

`#measurements`: Generic tabular data not fitting any of the other terms. Because of its lack of specificity, this term should generally be avoided, and new, more precise terms should be introduced instead.

We could have a common superconcept "relational" for them, but I'd rather wait and see how they work out in the wild.

11. Comments? Suggestions?

While this vocabulary is still preliminary, it's easy to change things. Let's make this as good as we can while it's still malleable!

Talk to me or post to semantics@ivoa.net.

Thanks!