Data Models and Interoperability at High Resolution

Jonathan McDowell

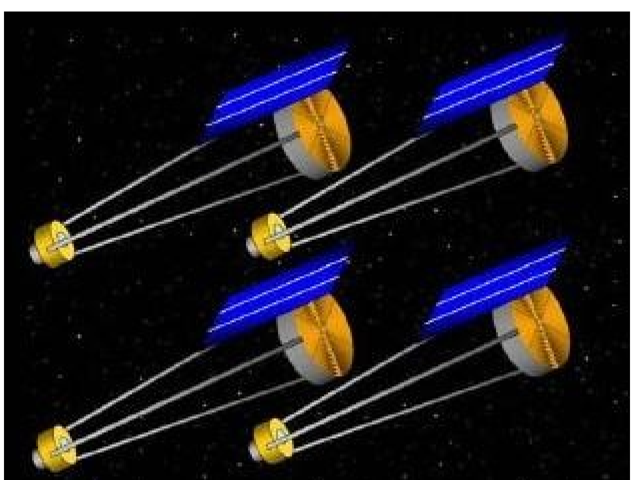
COSPAR Jul 2006

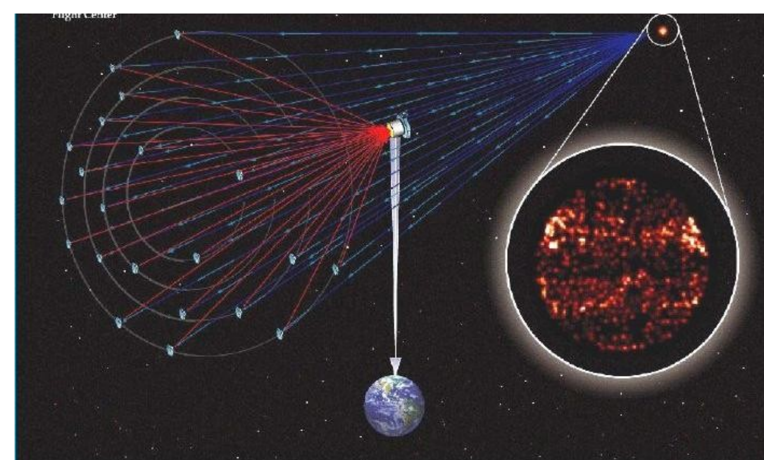
High Res Space Astronomy

- Forced to data analysis which is
  - Multiwavelength
  - Multiscale
  - Multi-instrument
  - and multigigabyte!
- Common data models can help us integrate these data
- Make the algorithms and the archives work together

High Res Space Astronomy

- Con-X: few “ spatial and high spectral res.
- Gen-X: 0.1” and high spectral resolution
- Maxim: microarcseconds in X-rays
- Optical Interferometers: Stellar Imager, 100 microarcsec
- Compare near term and current missions:
  - Herschel: 20” in submm - plus high res. spectroscopy
  - JWST, HST 0.1”, Chandra 0.5”



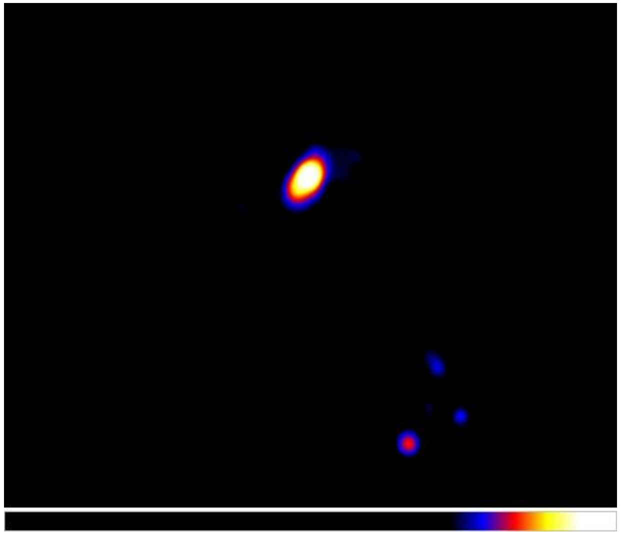


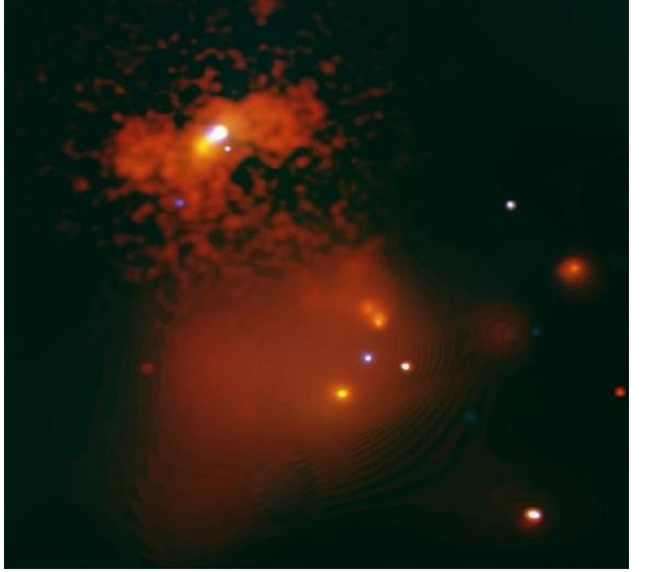
Role of the IVOA

- The International Virtual Observatory Alliance (IVOA) is a collaboration of many national VO projects
- It is also a standards organization
- We must have standards if we are to work interoperably with the new generation datasets
- Data Models group defines COMMON METADATA to describe our data

Challenges

- Multiscale data: what are the PSFs of these two datasets? What are the spectral responses?
- What if one dataset has a detailed PSF and the other only has an approximate resolution?
- IVOA will not solve the algorithmic problem of combining the data but we will address standard interfaces those algorithms can make use of.





Characterization

- The IVOA Characterization model allows you to specify 'what is this data' to multiple levels of subtlety. First we discuss “Coverage” - where is the data?
- Simple level: “Location” - where is the data in parameter space (RA, Dec, wavelength, time)
- Almost as simple - “Bounds” - we observed within these max and min values
- Next: “Support”: exact field of view shape, perhaps including bad pixels/columns; sequence of on/off times; spectral range; visibility cuts
- Finally “Sensitivity”: depth as a function of RA, Dec; transmission curve, sensitivity vs time

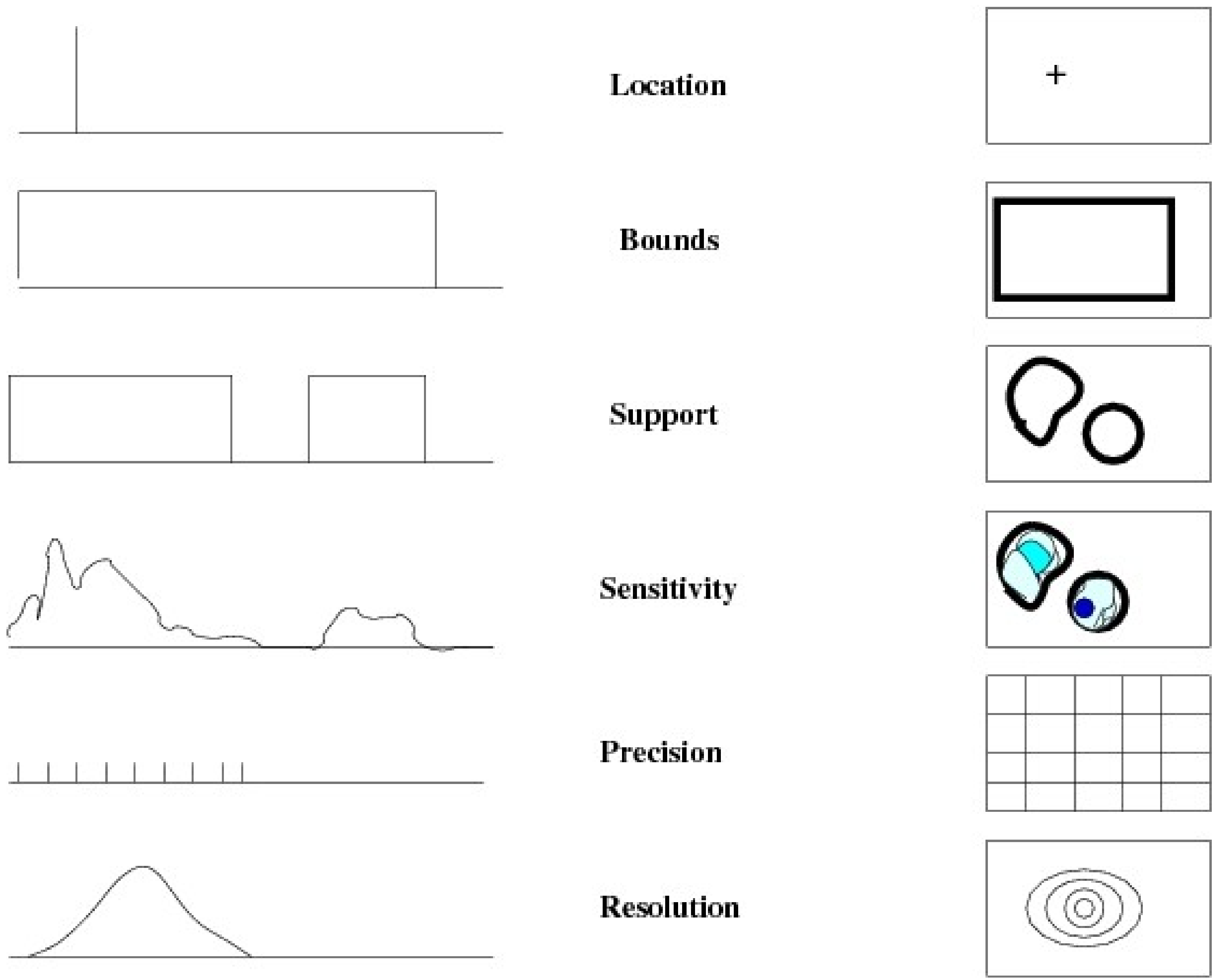
Formats

- Current IVOA effort focuses on XML representation
- Abstract model can also be mapped to other formats

```
<Characterization coord_sys=”id02”>
<Coverage>
<SpatialAxis>
<Support>
 <Region>
<Polygon>148.2312,-
85.1321,148.2331,-84.9812,....
</Polygon>
</Region>
</Support>
</SpatialAxis>
<TimeAxis>
<Support>
```

```
SPA_REG = 'Poly.reg'  / Spatial cha
TIMEREG = 'GTI.fits'   / HDU with
times
...

EXTNAME = 'GTI'
TTYPE1 = 'START'
TTYPE2 = 'STOP'
TUNIT1 = 's'
TUNIT2 = 's'
```


Location
Bounds
Support
Sensitivity
Precision
Resolution

Characterization - 2

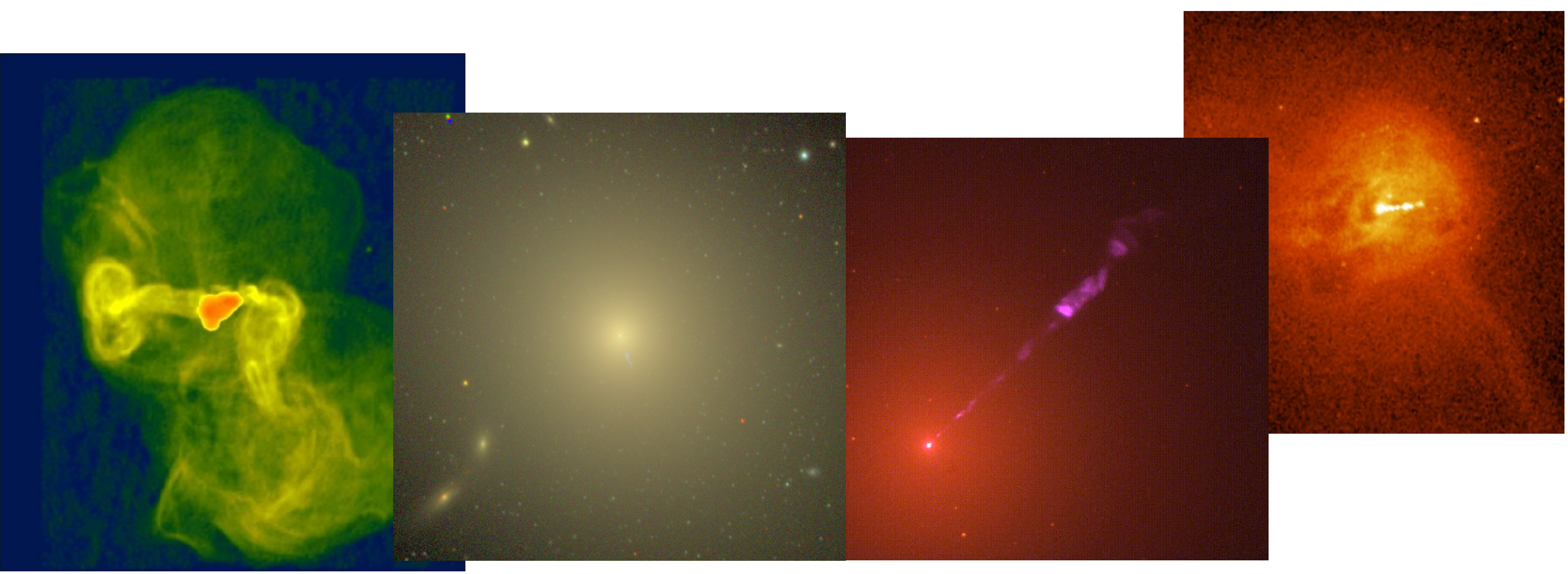
- Similar approach to Resolution
- Simple level: single number for dataset for each of time, spatial, spectral resolution
- or you can give bounds for the values
- or the full PSF, line spread function etc.
- Same story for errors, with support for stat/sys errors, quality
- Key ideas:
  - provide a standard way to express your data,
  - provide a place for detailed metadata but
  - make sure small projects can provide simpler metadata

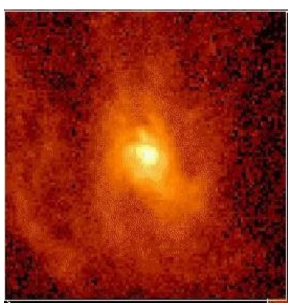
More Challenges

- Point sources don't exist! (if you look hard enough)
- Plus, the sources all have proper motions!
- Source catalogs keyed on position don't work
- Extended source analysis is hard to automate:

More Challenges

- Problem of source identification - and source identity - on multiple scales
  - Active nucleus; VLBI core/jet, knot
  - Broad emission line region
  - Host galaxy bulge and halo
- Some of these have sharp boundaries, some blend into each other. All have the same central coordinates. How do we characterize, how do we do the bookkeeping?





stolen from Bill Keel, F. Owen, Bill Forman, H Ford

Catalog models

- The IVOA is developing source catalog models
- We need to go beyond the traditional 'sort on RA' approach
- Standards for characterizing complex sources
- Standards for distinguishing different components with the same nominal RA and Dec
- Fluid query software, that can cope with similar objects even if they are divided up differently

And more challenges

- In astronomy, we make an observation. We see some objects and measure their properties. Then we compare these observational data with a model.
- Right?

Low resolution observations

And more challenges

- In astronomy, we make an observation. We see some objects and measure their properties. Then we compare these observational data with a model.
- Right?
- er... not at the limit of resolution we don't!
  - Deconvolution ambiguities
  - Aliasing in some imaging techniques
  - Extrasolar planet detection: very indirect (now that SIM is RIP)
  - none of this is news to radio-astronomers!
- Models and data interpretation are TIGHTLY COUPLED: a problem for archival data products

Encoding Assumptions

- You can avoid some of this by forward folding a la X-ray, but for large surveys it's not practical to do a joint forward fit to many sources folded through many instruments.
- We need a way to record our assumptions about model algorithms and parameters,
- and make someone else's software understand them - and change them
- Bayesian approach may help: use the language of priors
- We will need small component physics models
  - cosmology
  - galaxy spatial profiles, spectral line libraries...

Astronomical Semantics

- UCD (Uniform Content Descriptors
  - Already deployed
  - phot.flux;em.freq   specifies f-nu
  - Dictionary for relatively precise description of the physical concepts
  - Key to reliably connecting the output of one piece of software to the input of another
  - UCD and units specify what a parameter means, don't have to guess from the parameter name

Priors and assumptions

- We can record a set of fixed model assumptions
  - we used H0=73 km/s/Mpc
  - took a King model with a core radius of 2 kpc
- Still work to be done to make a standard language to say this
- Harder: prior distributions
  - The model code implies a gaussian temp. dist'n
  - We initially assume a uniform dist'n of sources
- Should be possible to make a standard way of representing such statements using UCDs and a language of Bayesian priors

Astrophysics mini-models

- In each subfield, can define a model for standard representation of a problem
- Jets (radio AGN, young stars, etc..)
- Must allow arbitrary extension – always new science

<jet>

<flux>.....</flux> (Reusing)

<power unit=”erg/s” type=”bol”>1.e38</power>

<jetshape>

<opening_angle unit=”deg”>12.3</opening_angle>

<pos_angle>102.1<pos_angle>

<length unit=”arcmin”>1.2</length>

</jetshape>

<morph>FR2</morph>

Conclusions

- The era of pointlike well-defined sources with constant RA/Dec whose properties we know in an unambiguous model-independent way is coming to an end
- To extract the maximum science from the next generations of space observatories we must embrace a more subtle astronomy where we are careful to characterize hierarchies of sub-sources within a source, and capable of handling model-laden interpretations of a source based on combining data at different resolutions and wavelengths. (and given the budget cuts, we'd better plan to live a long time...)
- The IVOA can help by providing standards for encapsulating these problems in observation and pipeline metadata