# Interop Sep 2006

Spectrum Data Model

Jonathan McDowell

Spectrum Overview

- Spectrum model represents a set of fluxes versus a set of spectral coordinates

- Use UCD to specify type of flux and type of spectral coordinate (wavelength, freq. etc)

- Support for describing errors, bin size, data quality for each point

- Metadata definitions grouped in mini-data models (we will reuse in other data models)

- Same model usable for time series, photometry

- Aggregatable into SED, echelle, etc. model
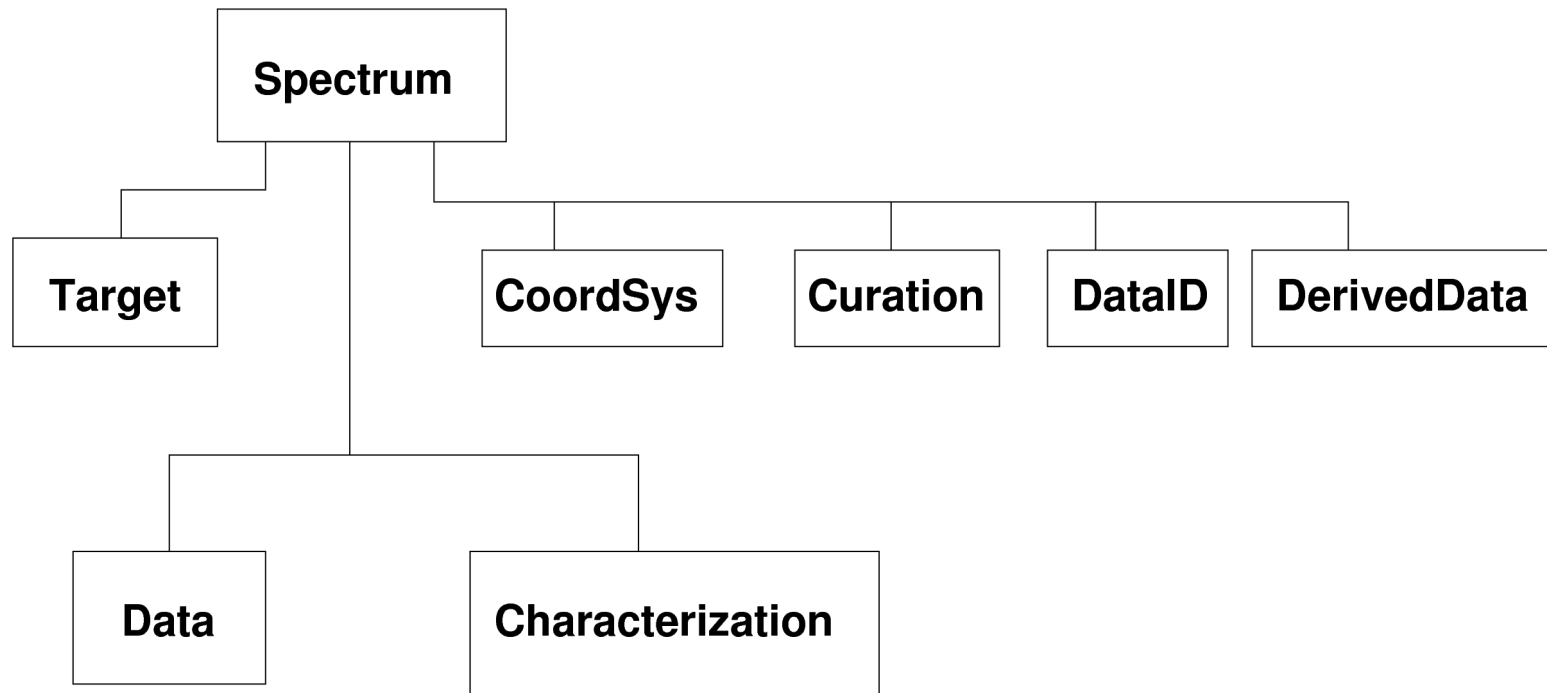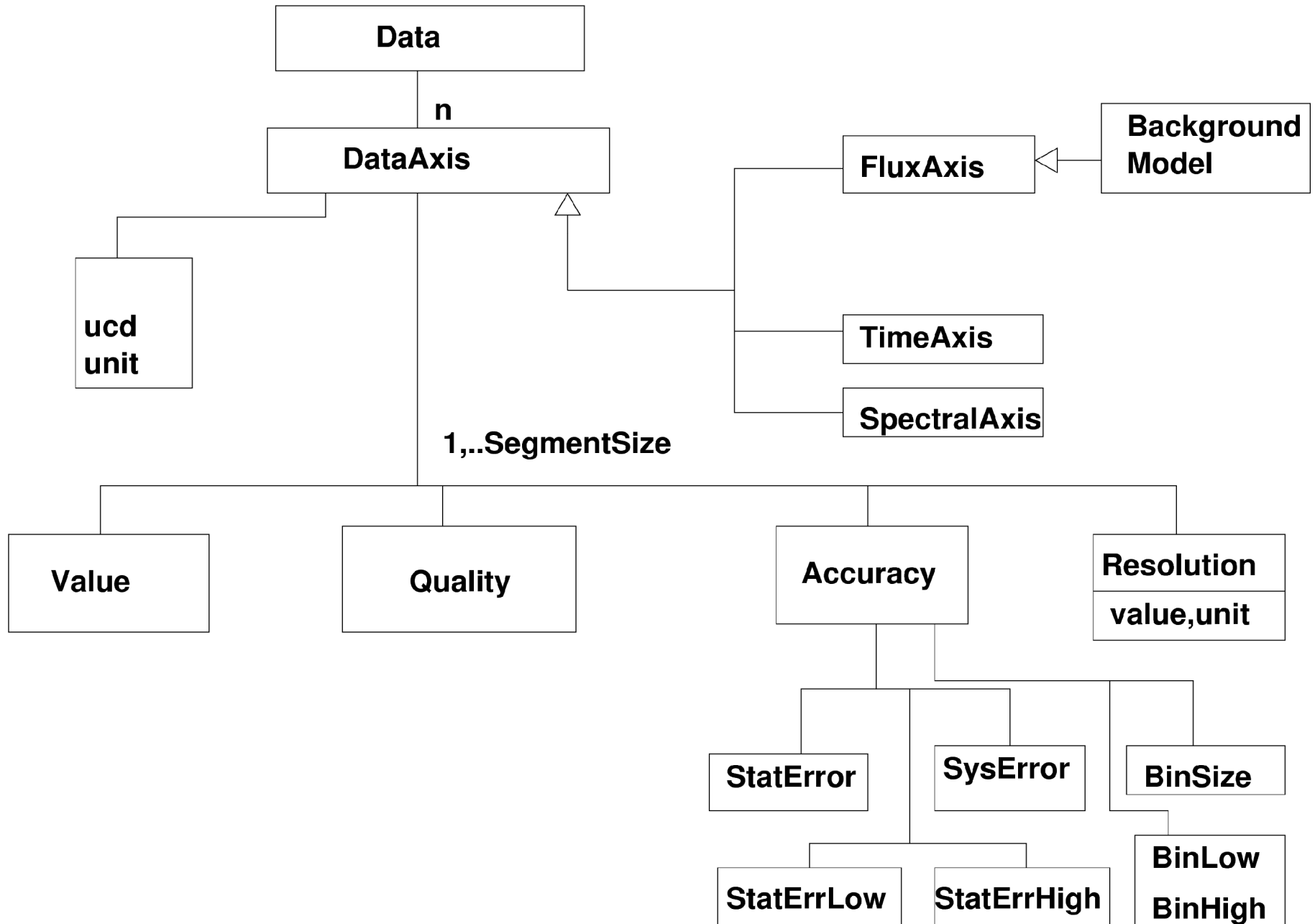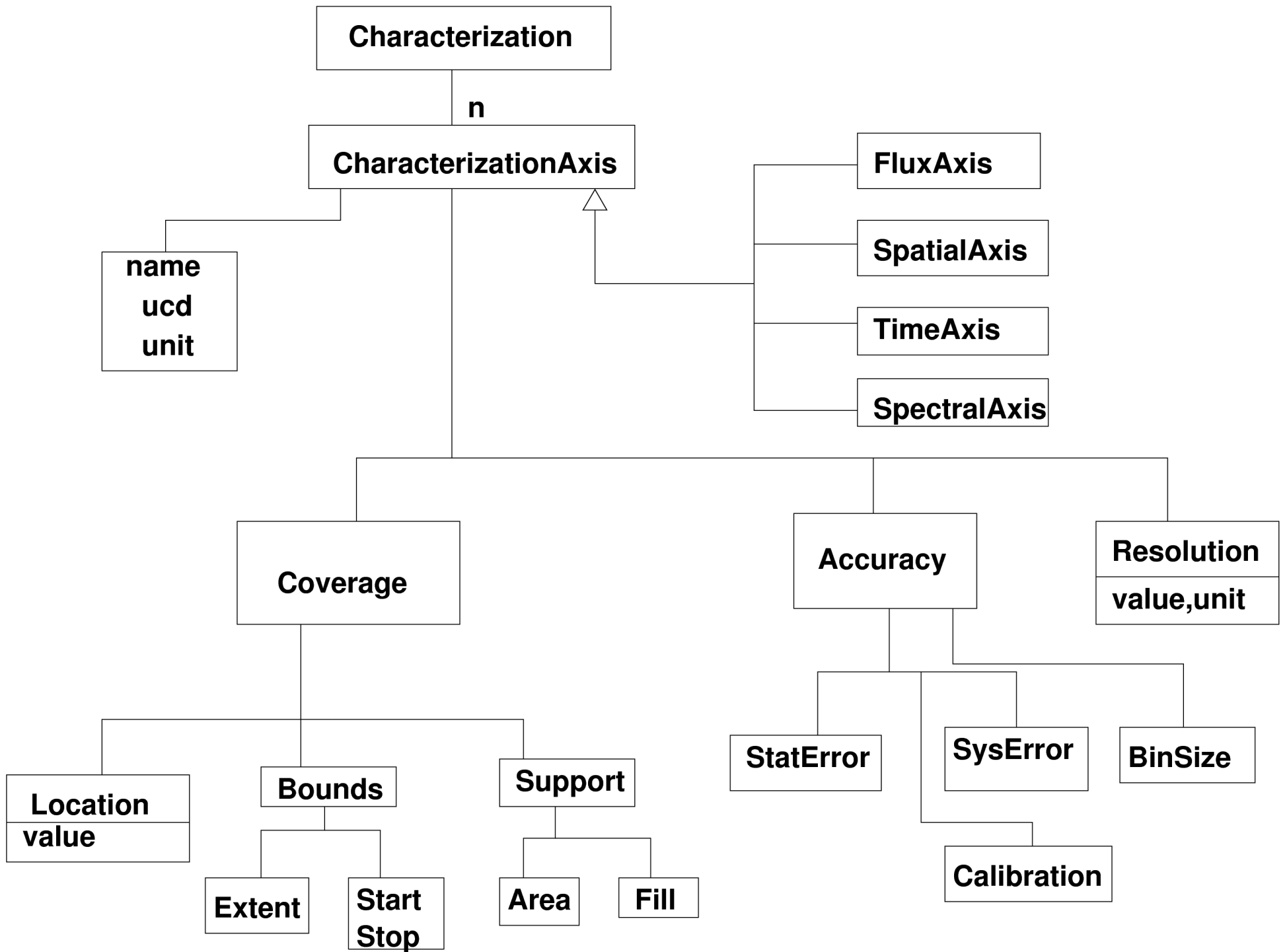
Spectrum

- New document version 0.98c

- http://hea-www.harvard.edu/~jcm/vo/docs/spec98/spec98c.pdf

- Changes since last Interop:

  - Made changes agreed in Victoria

  - Major reorganization of document:

    - consistency of nomenclature
    - clarification of characterization vs data
    - consolidated list of utypes
    - overall clarification

  - Minor modifications to XML and FITS serializations

- Doug and I believe this is ready for PR with minor edits

- R. Thompson and A. Richards, others submitted comments

Feedback – Lazlo Dobos

- XML serialization/schema:

  - For web services, better use array of simple types rather than array of structs, to work around SOAP serializer limitations for optional data items (end up with 0s instead of absent items)

  - I would welcome feedback from schema/xml experts

Feedback - various

- SSAP protocol (and secondarily, DM)
  - Must specify type of serialization
  - need uniform VO mechanism
  - extension of MIME types
- Note: currently Spectrum DM identifies 3 specific formats as well as 'foreign':
  - Spectrum DM version 1.0 / FITS
  - Spectrum DM version 1.0 / VOTABLE
  - Spectrum DM version 1.0/ XML
  - these three are unambiguous

Feedback – Alberto Micol

- MUST/SHOULD/MAY instead of REQUIRED/OPTIONAL (Anita also said this)
  - Complication is that it depends on what the data is (theoretical, spectrum, time series)
  - Proposal: MUST can always include UNKNOWN as a valid value (and maybe N/A as well? or this is effectively the same)
  - Accuracy must default to UNKNOWN, not zero
- Disentangle coverage and sampling (extent vs bins)
  - but they **are** tangled (bounds must include error bin width)

Feedback – Randy Thompson

- Change remaning SED references to 'spectral associations'
  - e.g. Echelle has extra requirements compared to SED; want support for different kinds of grouping of spectra, time series.
  - actual support for these is a post-1.0 task
- FITS serialization: have SPEC... and SPCO.. keywords – make more consistent
- Give more FITS examples
- Proposal: TUCDn, TUTYPn
  - concern: filtering software may lose columns

Feedback – Myron Smith, MAST

- SNR: advocates a standard definition (and suggests one) in addition to data-provider-specific SNR.
  - S/N = mean flux / median (diff)
  - where diff is difference between adjacent flux values
- Redshift: MAST objects (ambiguous, offensive to galactic astronomers, prefers + to – shifts)
  - proposes WAVSHIFT, in km/s but not presuming doppler: (lam/lam0-1)*c as an SSAP param
  - I propose keeping Redshift, but specifying that it's the "optical" definition, or providing UCD to distinguish.

Feedback – Anita Richards

- Units: allow any units and any SI prefix
  - We earlier agreed restricted set for V1.0, expand later
  - Compromise?: MUST have unit; SHOULD be in restricted set
- Units for data axes: MUST if the axis is present
- Axes (in Char): are generic other axes (e.g. CMB multipole) allowed?
- Coordinate conversion: any positional coords allowed? Use of full STC vs just ICRS?
  - Victoria decision: don't wait for STC
  - Agree must evolve to full STC when supported

Feedback – more from Anita

- SpectralAxis char. should be optional for timeseries
  - I respectfully disagree
- Bounds.Extent objected to, not fully equivalent to Bounds.Min/Max
  - Yes it is if I define it carefully enough?
- Concern about use of specified formats vs native data provider format
  - foreign (data provider) formats allowed by protocol
  - must encourage standard formats for spectra