SED Data Model and Access Protocols

Doug Tody, NRAO, USVAO

SED Data Model

- Topics
  - Status
  - SED data model
  - Open issues

SED Data Model

- Status
  - In Pune identified need for core SpectralDM
    - Defines core model for SED segment
    - Needed for Spectrum, TimeSeries as well as SED

  - SED Data Model
    - Updated to be based upon SpectralDM
   - Promoted to working draft

  - SED Data Access Protocol
    - Not yet addressed aside from VAO/NED prototype

SED Data Model

- **Uniform SED**
  - Uniform view of photometric data points from multiple observations
    - Presents overall/combined SED with uniform spectral and flux units
    - Can be viewed as a type of Spectrum
  - *Rebinned* SED is a special case of a uniform SED
    - Combines any overlapping segments/observations
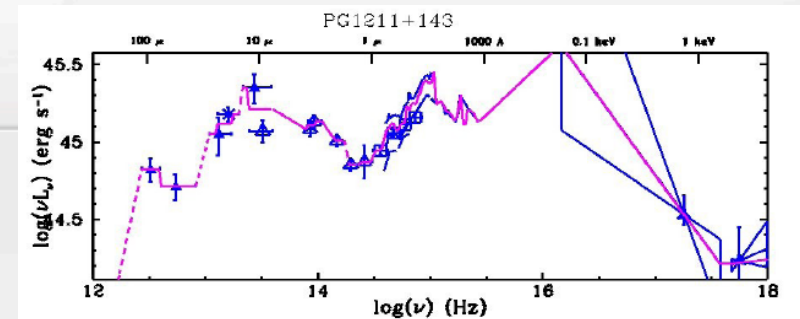
- **Aggregate SED**
  - A set of "segments", one for each contributing observation
    - Permits full characterisation of each observation
  - Used by a SED builder tool to build the overall/uniform SED
  - Allows editing of an existing uniform SED

SED Data Model

- A SED dataset is
  - A single "segment" (the uniform or binnned SED)
    - Must always be present to define standard dataset metadata
    - Is a uniform/binned SED only if data is present
  - An aggregation of segments
    - Mapping of SED data points to segments can be complex
  - Or both within the same dataset

- Segments of aggregate SED can be
  - Single photometry point
  - Time series or light curve
  - Spectrum

SED Segments

- ## PhotometryPoint
  - Limiting case of light curve with single data point
  - Often derived from an image
    - hence addition of aperture correction support in SpectralDM

- ## Spectrum
  - Spectrum is described directly by SpectralDM 2.0
  - SpectralDM integrates Photometry DM (allows fluxing)

- ## TimeSeries
  - Time series is an extension of SpectralDM
  - Light curve is a time series with one photometric band

SED Metadata

- Dataset metadata
  - Describes entire VO Dataset, e.g. for discovery, characterisation
    - Generic Dataset metadata for overall SED
    - Unchanged from SpectralDM, same for Spectrum, TimeSeries, etc.
  - SED-specific metadata
    - Dataset *type*, *DataModel*, SED *length*, *nSegments*, *sedAggregrate* URI, etc.

- Data point metadata
  - Included directly in table data
  - Describes a data point in the uniform SED
  - Some summary segment-related metadata

- Custom metadata
  - Provided by data provider or SED builder application
  - Describes specific algorithm/process used to compute uniform SED

Dataset SED Metadata

Dataset.dataModel	"SED-1.0"
Dataset.type	"SED" (the overall Dataset type as well as the segment type)
Dataset.length	Number of data points in the uniform SED
SED.nSegments	Number of segments in the aggregate SED
SED.sedAggregate	URI to aggregate SED if applicable (FITS keyword: SEDORIG)

Uniform SED Datapoint Metadata

Data.segmentType	One of "PhotometricPoint", "TimeSeries", "Spectrum", or "Composite".
Data.segments	The segment number, a comma-delimited list of segment numbers, or NULL if a more complex combination of segments are used to compute the given photometric value. The uniform SED is segment zero.
Data.SpectralAxis.value	Spectral coordinate for data points.
Data.FluxAxis.value	Computed flux value for data point.
Data.comments	Additional information about the computed data value or segment.

SED Serializations

- ## Science data formats
  - VOTable, FITS binary table
  - Segment serialization mostly defined by SpectralDM

- ## Either serialization can contain
  - Uniform SED
    - single table
  - Aggregate SED
    - set of tables, one per segment
  - Both Uniform and aggregate SED in one file
    - useful to not only view but understand or edit uniform SED

Time Series Data

- TimeSeries DM
  - Dataset type is "TimeSeries"
  - By default inherits entire SpectralDM
    - Dataset, DataID, Curation, Char, etc.

- New "TimeSeries" component model
  - Allows us to add whatever time series data requires
    - e.g., period, periodIsFolded, trend removal, classification etc.
  - Work with time series community to define this (CRTS, AAVSO, etc.)

- Issues
  - Require multiple instances of FluxFrame for multiband photometry

Open Issues

- Utypes
  - Namespace usage, inheritance
    - Same approach as in SSA/Spectrum for now

  - Metadata extension
    - As in current Utype draft ("sed.ned:" example)
      - sed:Data.FluxAxis.value
      - sed.ned:Data.FluxAxis.publishedValue

  - Multiple Instances
    - TBD - container issue or possibly extend Utype syntax
      - e.g., "ts:CoordSys.FluxFrame.XX;<instance>"

Open Issues

- ## SED Data Access Protocol

  - Concept

    - Based upon SSA query interface
      - Registered as a "sed" service
      - Service capabilities specific to SED (uniform/aggregate etc.)
    - Inherit basic interface, define what is different/new
      - e.g. query by object name is important for SED data
    - SpectralDM and SSA define the core for Spectrum,SED,TimeSeries

  - Status

    - VAO has prototyped this for NED SEDs
    - Include simple version based upon SSA in SED spec?