

Linear STC

An attempt at defining a linear syntax that allows simple specification of STC metadata

Arnold Rots, CXC/SAO

2005-02-24T17:39:00

Scope and Basic Principles

The STC metadata design for the Virtual Observatory has been implemented in XML Schema. For documentation, see <http://hea-www.harvard.edu/~arots/nvometa>. That design is complete and self-consistent, but may seem bewilderingly complex to the novice user; and, anyway, XML certainly does not lend itself to easy human readability.

There is an obvious need for a syntax that allows humans to specify and interpret STC metadata objects in a way that is as intuitive as possible without being sloppy – essentially a minimalist approach. This is especially true for queries and resource descriptions. A linear representation is highly desirable to achieve that purpose. We propose that a complete STCDescription (AstroCoordSystem, AstrCoords, and AstroCoordArea) be represented by a single phrase consisting of one or more sub-phrases, with each sub-phrase specifying the STC information for one coordinate frame. Each sub-phrase starts with the type of the CoordArea element assuming the additional role of Frame Identifier. If there is no CoordArea element, the Coord element type fulfills this role. It is followed by the associated coordinate frame specification (ReferenceFrame, ReferencePosition, etc.), the remaining CoordArea elements, and the Coords elements. Since the Frame Identifiers are unique, no additional punctuation or parenthesizing is required.

The syntax proposed in this document is based on a simple recipe:

- Write out all elements and values of the STC element from an XML document, separated by whitespace, and ordered by coordinate frame
- Within each coordinate frame, start with the identification of the CoordArea element, immediately followed by the CoordFrame, then the CoordArea elements, and finally the Coords elements; if there is no CoordArea element, start with the identification of the Coords element instead
- Allow omission of all words in the resulting sentence that does not change the information content (with the adoption of some common defaults)

As will become apparent, this second rule leads to very considerable economy (and thereby transparency) in what otherwise would be a very verbose construct.

One should note that this definition allows transforming XML into linear STC and vice versa. A transformation and a subsequent inverse transformation are not guaranteed to produce a result that is identical to the original, but it should be equivalent.

Query Example

Let us take the query example from Appendix B.5 of the STC Working Draft version 1.20:

Verbose version	Economy version	Comments
TimeInterval TimeFrame Name Time TimeScale TT BARYCENTER StartTime TimeScale TT ISOTime 1900-01-01T00:00:00	StartTime TT BARYCENTER 1900-01-01	StartTime indicates that we are talking about time and a TimeInterval, TT is uniquely a TimeScale, and the format implies ISOTime
Region SpaceFrame Name Equatorial ICRS BARYCENTER SPHERICAL 2 Circle unit=deg Center 148.9 69.1 Radius 2.0 Position2D unit=deg Name RA,Dec unit=deg Resolution2 0.0001 0.0001 Resolution2 0.0003 0.0003 Size2 0.5 0.5 Size2 0.67 0.67 PixSize2 0.00005 0.00005 PixSize2 0.00015 0.00015	Circle ICRS BARYCENTER 148.9 69.1 2.0 Resolution 0.0001 0.0001 0.0003 0.0003 Size 0.5 0.5 0.67 0.67 PixSize 0.00005 0.00005 0.00015 0.00015	Circle uniquely implies a SpaceFrame and a Region, and the order of its parameters is fixed; ICRS implies spherical 2-D coordinates; and the two pairs of numbers for Resolution, Size, and PixSize implies upper and lower limits
SpectralInterval SpectralFrame Name Wavelength BARYCENTER unit=Angstrom LoLimit 4000 HiLimit 7000 Spectral unit=Angstrom Name Lambda Resolution 300 Resolution 600	SpectralInterval BARYCENTER A 4000 7000 Resolution 300 600	SpectralInterval implies a SpectralFrame; its unit implies Wavelength

If the reference positions and time scale do not matter, and one does not care about the Resolutions, Sizes, and PixSizes, the sum-total becomes simply:

```
StartTime 1900-01-01 Circle ICRS 148.9 69.1 2.0
SpectralInterval A 4000 7000
```

It may be of interest to note that, had there also been a stop time (say, 2000-01-01), the phrase could have been composed as follows:

```
TimeInterval 1900-01-01 2000-01-01 Circle ICRS 148.9 69.1
2.0 SpectralInterval A 4000 7000
```

A few Conventions

A missing required element, if its value or name is not implied by anything else in the phrase, should be interpreted as UNKNOWN or *nil*; i.e., “don’t care” in the context of a query.

Default units (“deg”, “m”, “s”, “m/s”, “Hz”) may be omitted.

To conform with common abuse, “J2000” and “FK5” will both be interpreted as “FK5 J2000”; “B1950” and “FK4” will be interpreted as “FK4 B1950”.

I have added a Box shape to the region definition – because it is so convenient and such a pain to figure out in terms of vertices on a sphere. The linear syntax will be:

Box <Xcenter> <Ycenter> <Xsize> <Ysize>

In spherical coordinates the sides of the box will be **great circles**. It is defined in terms of the size of a cross through the center, parallel to the axes, with great circle sides perpendicular to the cross at its end points.

This mechanism is primarily intended to make simple cases easy to express; initially it will be restricted to those easy cases and will not support:

- Custom coordinate reference frames and custom reference positions
- Spatial frames with offset positions
- Error, resolution, sizes in 2 or 3 dimensions other than the 2 or 3 element vectors
- Relocatable frames
- Planetary reference frames
- For ellipses: position angle references other than ‘X’ and units other than “deg”
- Geodetic reference spheroids other than IAU 1976

Formal Definition

We shall refer to a Linear STC expression as a *phrase*. A *phrase* is a linear concatenation of one or more *sub-phrases*. Each sub-phrase pertains to a particular coordinate frame and they shall be concatenated in the following order:

Time Space Spectrum Redshift

A *sub-phrase* is a linear concatenation of one, two, or three *sub-phrase components* in the order:

CoordinateFrame CoordinateArea Coordinates

A *sub-phrase component* is a linear concatenation of one or more *words*, the first of which is the *sub-phrase component identifier*. However, the CoordinateArea identifier (or, if CoordinateArea is absent, the Coordinates identifier) is moved in front of the CoordinateFrame sub-phrase component and serves as the *sub-phrase identifier*.

Words are the names of elements or attributes, or their values

Sub-phrase component identifiers are unique within each *sub-phrase* and *sub-phrase identifiers* are unique within each *phrase*. Words, components, and sub-phrases are separated by whitespace.

Time Sub-phrase

```
TimeInterval [fillfactor <fill>] [<timescale>] [<refpos>]
  [<start> <stop> ...]
  [Time <time>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [PixSize <pixsiz> [<pixsiz>]]
```

```
StartTime [fillfactor <fill>] [<timescale>] [<refpos>] <start>
  [Time <time>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [PixSize <pixsiz> [<pixsiz>]]
```

```
StopTime [fillfactor <fill>] [<timescale>] [<refpos>] <stop>
  [Time <time>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [PixSize <pixsiz> [<pixsiz>]]
```

```
Time [<timescale>] [<refpos>] [<time>]
  [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [PixSize <pixsiz> [<pixsiz>]]
```

Notes:

- *<timescale>* is one of the following: TT, TDT, ET, TAI, IAT, UTC, TEB, TDB, TCG, TCB, LST, nil; the default value is nil.
- *<refpos>* is one of the following: GEOCENTER, BARYCENTER, HELIOCENTER, TOPOCENTER, GALACTIC_CENTER, EMBARYCENTER, MOON, MERCURY, VENUS, MARS, JUPITER, SATURN, URANUS, NEPTUNE, PLUTO, UNKNOWN; the default value is UNKNOWN.
- *<time>*, *<start>*, and *<stop>* are one of the following: *<ISOstring>*, JD *<JDvalue>*, MJD *<MJDvalue>*; there is no default.
- *<unit>* only applies to the arguments that follow it and is one of the following: s, d, a, yr, cy; the default value is s.
- *<error>*, *<resln>*, and *<pixsiz>* are doubles; two doubles indicate a range.
- [...] indicates that the preceding parameter set may occur more than once, to specify a “support” area that consists of multiple intervals.
- *<fill>* (single interval only) is a number between 0 and 1; the default value is 1.0.

Space Sub-phrase

```

PositionInterval [fillfactor <fill>] <frame> [<refpos>] [<flavor>]
  [<lolimit> <hilimit> ...]
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

AllSky [fillfactor <fill>] <frame> [<refpos>] [<flavor>]
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

Circle [fillfactor <fill>] <frame> [<refpos>] [<flavor>] <pos> <radius>
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

Ellipse [fillfactor <fill>] <frame> [<refpos>] [<flavor>] <pos> <radius>
  <radius> <posangle>
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

Box [fillfactor <fill>] <frame> [<refpos>] [<flavor>] <pos> <bsize>
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

Polygon [fillfactor <fill>] <frame> [<refpos>] [<flavor>] [<pos> ...]
  [Position <pos>] [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

Position <frame> [<refpos>] [<flavor>] [<pos>]
  [unit <unit>] [Error <error> [<error>]]
  [Resolution <resln> [<resln>]] [Size <size> [<size>]]
  [PixSize <pixsiz> [<pixsiz>]]
  [<Velocity sub-phrase>]

```

Notes:

- **<frame>** is one of the following: ICRS, FK5, FK4, J2000, B1950, ECLIPTIC, GALACTIC[_II], SUPER_GALACTIC, GEO_C, GEO_D, UNKNOWN; the default value is UNKNOWN.
- **<refpos>** is one of the following: GEOCENTER, BARYCENTER, HELIOCENTER, TOPOCENTER, GALACTIC_CENTER, EMBARYCENTER, MOON, MERCURY, VENUS, MARS, JUPITER, SATURN, URANUS, NEPTUNE, PLUTO, UNKNOWN; the default value is UNKNOWN.
- **<flavor>** is one of the following: SPHER2, UNITSPHER, CART1, CART2, CART3, SPHER3; the default value is SPHER2.
- **<unit>** applies to all numerical arguments except **<fill>** and is one of the following:

Linear STC

deg, arcmin, arcsec, m, mm, km, AU, pc, kpc, Mpc; the default value for spherical coordinates, except GEO, is deg; for GEO: deg deg m; for Cartesian coordinates: m.

- `<pos>`, `<lolimit>`, `<hilimit>`, `<center>`, `<bsize>`, `<error>`, `<resln>`, `<size>`, and `<pixsiz>` are lists of 1, 2, or 3 doubles, depending on the dimensionality of the space frame; the last four may be pairs of vectors, indicating ranges.
- `<radius>` and `<posangle>` are doubles; position angle reference is 'X', unit=deg.
- [...] indicates that the preceding parameter set may occur more than once, to specify a "support" area that consists of multiple intervals.
- `<fill>` (single interval only) is a number between 0 and 1; the default value is 1.0.
- The Velocity sub-phrase:
[VelocityInterval [fillfactor `<fill>`] `<lolimit>` `<hilimit>` ...]
[Velocity `<vel>`] [unit `<unit>`] [Error `<error>`] [`<error>`]
[Resolution `<resln>`] [`<resln>`] [PixSize `<pixsiz>`] [`<pixsiz>`]
with m/s as default unit.

Spectral Sub-phrase

```
SpectralInterval [fillfactor <fill>] [<refpos>]  
  [<lolimit> <hilimit> ...]  
  [Spectral <freq>] [unit <unit>] [Error <error>] [<error>]  
  [Resolution <resln>] [<resln>] [PixSize <pixsiz>] [<pixsiz>]
```

```
Spectral [<refpos>] [<pos>] [unit <unit>] [Error <error>] [<error>]  
  [Resolution <resln>] [<resln>] [PixSize <pixsiz>] [<pixsiz>]
```

Notes:

- `<refpos>` is one of the following: GEOCENTER, BARYCENTER, HELIOCENTER, TOPOCENTER, LSR[K], LSRD, GALACTIC_CENTER, LOCAL_GROUP_CENTER, EMBARYCENTER, MOON, MERCURY, VENUS, MARS, JUPITER, SATURN, URANUS, NEPTUNE, PLUTO, UNKNOWN; the default value is UNKNOWN.
- `<unit>` applies to all numeric arguments except `<fill>` and is one of the following:
Hz, MHz, GHz, m, mm, um, nm, Angstrom, eV, keV, MeV;
the default value is Hz.
- `<freq>`, `<lolimit>`, `<hilimit>`, `<error>`, `<resln>`, and `<pixsiz>` are doubles; the last three may be pairs of doubles, indicating ranges.
- [...] indicates that the preceding parameter set may occur more than once, to specify a "support" area that consists of multiple intervals.
- `<fill>` (single interval only) is a number between 0 and 1; the default value is 1.0.

Redshift Sub-phrase

```
RedshiftInterval [fillfactor <fill>] [<refpos>] [<type>] [<dopplerdef>]  
  [<lolimit> <hilimit> ...]  
  [Redshift <rs>] [unit <unit>] [Error <error>] [<error>]  
  [Resolution <resln>] [<resln>] [PixSize <pixsiz>] [<pixsiz>]
```

```
Redshift [<refpos>] [<rs>] [<type>] [<dopplerdef>]  
  [unit <unit>] [Error <error>] [<error>]  
  [Resolution <resln>] [<resln>] [PixSize <pixsiz>] [<pixsiz>]
```

Linear STC

Notes:

- `<refpos>` is one of the following: `GEOCENTER`, `BARYCENTER`, `HELIOCENTER`, `TOPOCENTER`, `LSR[K]`, `LSRD`, `GALACTIC_CENTER`, `LOCAL_GROUP_CENTER`, `EMBARYCENTER`, `MOON`, `MERCURY`, `VENUS`, `MARS`, `JUPITER`, `SATURN`, `URANUS`, `NEPTUNE`, `PLUTO`, `UNKNOWN`; the default value is `UNKNOWN`.
- `<unit>` applies to all numeric arguments except `<fill>` and is `km/s` for velocities, `nil1` for redshifts.
- `<type>` is `VELOCITY` or `REDSHIFT`; the default value is `VELOCITY`.
- `<dopplerdef>` is `OPTICAL`, `RADIO`, or `REDSHIFT`; the default value is `OPTICAL`.
- `<rs>`, `<lolimit>`, `<hilimit>`, `<error>`, `<resln>`, and `<pixsiz>` are doubles; the last three may be pairs of doubles, indicating ranges..
- `[...]` indicates that the preceding parameter set may occur more than once, to specify a “support” area that consists of multiple intervals.
- `<fill>` (single interval only) is a number between 0 and 1; the default value is 1.0.