

# Data Model for Astronomical DataSet Characterisation



**LIRAE**  
Astro Engineering  
Lab

Cristopher Arenas

`carenas@csrg.inf.utfsm.cl`

Computer Systems Research Group,  
Universidad Técnica Federico Santa María

August 5, 2013



# Overview

## Objetivos

## Aspectos Generales

## Aspectos Generales

## Estructura y estrategia de desarrollo

- El enfoque de los Ejes

- El enfoque de las Propiedades

- Presentación de la información por capas

## El modelo de datos

## Serialización XML



# Objetivos

- Definir un modelo de datos que permita describir observaciones astronómicas y satisfacer requerimientos de descubrimiento, análisis y procesamiento de datos.



# Aspectos Generales

Al considerar una observación astronómica se identifican varias *Propiedades*:

- **Cobertura** (*Coverage*).
- **Precisión de Muestreo** (*Sampling Precision*).
- **Resolución** (*Resolution*).
- **Precisión** (*Accuracy*).



# Aspectos Generales

## **Cobertura** (Coverage)

- Describe la dirección en que estaba apuntando el telescopio, en cuales longitudes de onda y cuando y/o la región cubierta por cada eje.
- Descrito por niveles crecientes de detalle:
  - Ubicación (Location)
  - Límites (Bounds)
  - Soporte (Support)
  - Sensibilidad (Sensitivity)
  - Factor de llenado (Filling factor).



# Aspectos Generales

## **Precisión de muestreo** (Sampling Precision)

- Describe los intervalos de muestreo de cada eje.

## **Resolución** (Resolution)

- Describe la resolución física efectiva (PSF, LSF, etc).

## **Precisión** (Accuracy)

- Describe la precisión de la medición.



# Estructura y estrategia de desarrollo

Tres direcciones independientes:

- El enfoque de los ejes.
- El enfoque de las propiedades.
- Presentación de la información por capas.



# Estructura y estrategia de desarrollo

El enfoque de los ejes

Dimensiones físicas para los datos descritas por ejes tales como:

- SPATIAL.
- SPECTRAL.
- TIME.
- VELOCITY.
- VISIBILITY.
- POLARISATION.
- OBSERVABLE.





# Estructura y estrategia de desarrollo

## El enfoque de los ejes

- Se recomiendan los nombres anteriores.
- Nombres FITS pueden ser usados también.
- El proveedor de datos deberá proporcionar un UCD y unidades para cada eje.
- No hay límite en la cantidad de ejes y pueden ser dependientes o superpuestos.



# Estructura y estrategia de desarrollo

## El enfoque de los ejes

- Algunos ejes son implícitos al ser presentados con una palabra clave.
- Se utilizan *flags* para indicar propiedades de calificación booleanas:
  - Variable dependiente.
  - Estado de calibración.
  - Datos bajo muestreo.
- La precisión caracteriza alguna incertidumbre asociada con cada eje.



# Estructura y estrategia de desarrollo

El enfoque de las propiedades

Las propiedades principales necesarias para la descripción y recuperación de datos están categorizadas por:

- Cobertura
- Límites
- Resolución
- Precisión de Muestreo



# Estructura y estrategia de desarrollo

## El enfoque de las propiedades

- La cobertura está descrita por cuatro capas, las cuales entregan una visión jerárquica de detalle incremental:
  - *Location*: posición de un punto en un parámetro espacio N-dimensional.
  - *Bounds*: un rango en cada parámetro, proporcionando el límite inferior y superior de una caja N-dimensional.
  - *Support*: describe cuantitativamente el subconjunto de espacio, tiempo, frecuencia u otros dominios en los cuales hay datos válidos. Puede incluir uno o varios rangos en cada eje.
  - *Sensitivity*: provee valores numéricos indicando la variación de la función de respuesta en cada eje.



# Estructura y estrategia de desarrollo

## El enfoque de las propiedades

- Los límites pueden incluir como sub-nivel el factor de llenado.
- La resolución es usualmente el mínimo intervalo independiente de medición en cada eje.
- El muestreo, también llamado pixelación, precisión o cualtización, describe el truncamiento de valores de datos.



# Estructura y estrategia de desarrollo

Presentación de la información por capas

- Permite tareas para recuperar solo los metadatos requeridos.
- Los niveles más bajos pueden estar muy detallados. Esto puede tener varias formas:
  - Un valor simple o un rango.
  - Una función analítica de otros valores de *propiedad*.
  - Un *variance map* para datos 2D.
  - Una tabla de consulta para la corrección de paso de banda para datos espectrales 1D.
- Propiedades más complejas pueden ser proporcionadas utilizando punteros a datos auxiliares.



# El modelo de datos

- Diagramas UML describen la organización de los metadatos de *Characterisation* siguiendo la perspectiva Propiedades/Ejes/Niveles.
- Otros elementos incluyen el número de *bins* presentes en el eje *Observable* o ejes que representen fenómenos medidos por medio de otros ejes.
- Para la calibración, se provee un *flag* de estado *CharacterisationStatus* por cada eje que entrega información acerca de la calibración de los datos con alguno de los valores:
  - UNCALIBRATED
  - CALIBRATED
  - RELATIVE
  - NORMALIZED



# El modelo de datos

- Para el estado de muestreo se proveen *flags* para *Undersampling* y para el muestreo regular.
- La clase *Accuracy* provee *flags* que sirven como indicadores de calidad con múltiples niveles de detalle, análogo a *Coverage*.
- *Characterisation* es el elemento raíz del modelo. Puede ser serializado utilizando dos conjuntos alternativos de elementos primarios:
  - Propiedades, con las clases correspondientes para cada eje adjuntas.
  - Ejes, factorizando cada descripción en los niveles multi-capas de las propiedades.



# El modelo de datos

Implementación utilizando elementos de STC

- *Location* implementa STC: AstroCoords.
- *Bounds* implementa tipos básicos de STC. Algunos elementos de STC: Interval y STC: Coords en una estructura similar a STC: AstroCoordArea.
- *Support* usa STC: AstroCoordArea.
- *Resolution* ResolutionRefVal puede ser implementado usando elementos STC: Resolution.
- *SamplingPeriod* y *SampleExent* encapsulan elementos de STC: CPixSize.



# Serialización XML

Esquema XML de *Axis First*

- Serialización basada en un árbol.
- Elementos apropiados son tomados de STC.
- El elemento raíz, *Characterisation*, es la agregación de un conjunto de elementos *CharacterisationAxis* por cada eje.
- *CharacterisationAxis* contiene información de los ejes, coordenadas del sistema, unidades, etc.
- *Coverage* implementa distintos elementos acorde a los cuatro niveles de descripción.



# Serialización XML

Esquema XML de *Axis First*

- Los niveles más bajos de estas propiedades pueden reusar parámetros de ejes de niveles superiores para redefinir sus propios parámetros de ejes localmente.
- La mayor parte de coordenadas de información reutiliza definiciones y estructuras de las coordenadas STC.
- Para construir el esquema se considera una jerarquía desde el concepto más general bajando a los más específicos.
- Las clases agregadas son traducidas en elementos agregados.
- Los atributos de una clase UML son codificados como elementos de subniveles.

# Serialización XML

## Utypes

- Son identificadores que permiten reconocer un concepto en varios modelos o serializaciones.
- Se construyen mediante la navegación en la representación del esquema XML, siguiendo los enlaces lógicos provistos.
- La idea es que puedan ser entendidos por cualquier herramienta de un OV fuera del modelo.



# Data Model for Astronomical DataSet Characterisation



**LIRAE**  
Astro Engineering  
Lab

Cristopher Arenas

`carenas@csrg.inf.utfsm.cl`

Computer Systems Research Group,  
Universidad Técnica Federico Santa María

August 5, 2013

