

Reporte Characterisation

Cristopher Arenas*

* Universidad Técnica Federico Santa María, CSRG. Santiago, Chile

I. OBJETIVOS DE CHARACTERISATION

El documento “*DataModel for Astronomical DataSet Characterisation*” (*Characterisation*) tiene como objetivo definir un modelo de datos que permita describir observaciones astronómicas y satisfacer requerimientos de descubrimiento, análisis y procesamiento de datos.

II. RESUMEN DE CHARACTERISATION

A. Aspectos generales

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

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

B. Estructura y estrategia de desarrollo

Characterisation provee una estructura para presentar los metadatos necesarios para especificar un conjunto de datos en un formato estándar y para realizar cualquier interrelación explícita. Se permite la evolución del modelo en tres direcciones independientes.

1) *El enfoque de los Ejes*: Las dimensiones físicas para los datos están descritas por ejes tales como: SPATIAL, SPECTRAL, TIME, VELOCITY, VISIBILITY, POLARISATION, OBSERVABLE. Se recomienda que los *data providers* usen estos nombres, pero no es obligatorio (nombres FITS pueden ser usados). El *data provider* deberá proporcionar un UCD y las unidades para cada eje. No hay límite en el número de ejes y pueden ser dependientes o superpuestos.

Algunos ejes no pueden ser presentados por los datos, pero son implícitos al presentarlos solo con una palabra clave en la cabecera o en otro lugar.

Se usan *flags* para indicar propiedades de calificación booleanas. Incluye si el eje representa una variable dependiente, el estado de calibración y si los datos están bajo muestreo.

La precisión caracteriza alguna incertidumbre asociada con cada eje.

2) *El enfoque de las Propiedades*: Las propiedades principales necesarias para la descripción y recuperación de datos están categorizadas por Cobertura, Resolución y Precisión de Muestreo.

La cobertura está descrita por cuatro capas, las cuales entregan una visión jerárquica de detalle incremental:

- a. *Location*: es la posición de un punto en un parámetro espacio N-dimensional.
- b. *Bounds*: es un rango en cada parámetro, proporcionando el límite inferior y superior de una “caja” N-dimensional. Los límites garantizan encerrar todos los datos válidos.
- c. *Support*: describe cuantitativamente el subconjunto de espacio, tiempo, frecuencia otros dominios en los cuales hay datos válidos. Puede incluir uno o varios rangos en cada eje.
- d. *Sensitivity*: provee valores numéricos indicando la variación de la función de respuesta en cada eje.

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 o precisión o cuantización, describe el truncamiento de valores de datos como parte de la adquisición o procesamiento de datos.

3) *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.

C. El modelo de datos

Se utilizan diagramas UML para describir la organización de los metadatos de *Characterisations* siguiendo la perspectiva Propiedades/Ejes/Niveles.

La información relacionada con un eje es reunida dentro de la clase *CharacterisationAxis*. Contiene el nombre del eje, unidades, UCD, así como un soporte para el marco de coordenadas STC.

Otros elementos pueden incluir 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.

Para el estado de muestreo se proveen flags para *Under-sampling* y para el muestreo regular.

La clase *Accuracy* provee flags que sirven como indicadores de calidad con múltiples niveles de detalle, análogo a *Coverage*.

Como elemento raíz del modelo se encuentra la clase *Characterisation*. El modelo puede ser serializado utilizando dos conjuntos alternativos de elementos primarios:

- Propiedades (*Properties*), con las clases correspondientes para cada eje adjuntas.
- Ejes (*Axes*), factorizando cada descripción en los niveles multi-capas de las propiedades.

1) Implementación del modelo usando elementos de STC:

Se utilizan elementos de STC de nivel intermedio como bloques de construcción del modelo *Characterisation*.

El objeto `STC:AstroCoordSystem` es necesario como referencia para el eje *Coverage*. Subestructuras pueden ser usadas de la siguiente forma:

- *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 implementada usando elementos `STC:Resolution`.
- *SamplingPeriod* y *SampleExent* encapsulan elementos de `STC:CPixSize`.

D. Serialización XML

1) *Esquema XML de Axis First*: La serialización XML de *Characterisation* está basada en un árbol. Elementos apropiados son tomados de STC. El elemento raíz, llamado *Characterisation* es la agregación de un conjunto de elementos *CharacterisationAxis* por cada eje. Este último elemento contiene la información de los ejes, coordenadas del sistema, unidades, etc. *Coverage* implementa distintos elementos acorde a los cuatro niveles de descripción. Los niveles más bajos de estas propiedades a través de un *CharacterisationAxis* puede reusar parámetros de ejes de niveles superiores para redefinir sus propios parámetros de ejes localmente.

La mayor parte de las coordenadas usuales de información reutilizan definiciones y estructuras de las coordenadas STC.

Para contruir el esquema se considera una jerarquía desde el concepto más general bajando a los más específicos. Las clases agregadas son traducidas en subelementos agregados. Los atributos de una clase UML son codificados como elementos de subniveles.

2) *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.

III. IMPLEMENTACIONES

Describir implementaciones realizadas, o que hayan encontrado hechas por algún observatorio virtual (se exige poner las correspondientes referencias)