

Reporte ObsCore

Cristopher Arenas*

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

I. OBJETIVOS OBSCORE

El documento “*Observation Data Model Core Components and its Implementation in the Table Access Protocol*” (*ObsCore*) persigue dos objetivos:

- Definir un modelo de datos simple para describir los datos de las observaciones.
- Definir una forma estándar de exponer este modelo a través del *Table Access Protocol* para proveer una interfaz uniforme y así descubrir productos de datos científicos observacionales de cualquier tipo.

II. RESUMEN OBSCORE

A. Casos de Uso

El foco primario de *ObsCore* está en el descubrimiento de datos. Se definen una serie de casos de uso con el objetivo de encontrar *observational data products* en el dominio del OV entregando una misma consulta a múltiples archivos. Los *data providers* deben poder soportar consultas, tales como:

- Soporte multi-longitud de ondas (búsquedas de posición y tiempo).
- Soporte de cualquier tipo de *data product* científico (imagen, cubo, espectro, etc).
- Soporte directo del ordenamiento del contenido típico encontrado en archivos (FITS, VOTable, etc).

B. Elementos del modelo de datos

1) *Data Product Type*: Es un atributo que describe una clasificación científica de alto nivel acerca del *data product* que está siendo considerado. Está codificado como un string en minúscula que transmite una idea general del contenido y la organización del conjunto de datos. Los valores posibles son:

- **image**: una imagen astronómica, típicamente una imagen 2D con dos ejes espaciales.
- **cube**: una imagen astronómica multidimensional con 3 o más ejes.
- **spectrum**: cualquier conjunto de datos para los cuales la cobertura espectral es el atributo primario.
- **sed**: distribución de energía espectral, un *data product* avanzado, producido por la combinación de datos de múltiples observaciones.
- **timeseries**: arreglo de una dimensión que presenta alguna cantidad como función del tiempo.
- **visibility**: relacionado con el radio de visibilidad de algún tipo. Típicamente son datos instrumentales. Puede contener datos con información espacial, espectral, temporal

y de polarización. Puede ser usada para producir *data products* de mayor nivel.

- **event**: conjunto de datos de un recuento de eventos de algún tipo. A menudo corresponde a un conjunto de datos de un objeto complejo que contiene múltiples archivos u otras subestructuras. Puede contener datos con información espacial, espectral y temporal. Puede ser usada para producir *data products* de mayor nivel.

El *data provider* debe elegir la categoría primaria más apropiada para el conjunto de datos. En caso de que no se encuentre un tipo apropiado se permite el valor NULL y más información deberá ser proporcionada en el campo subtipo del *data product* que incluye el modelo. La combinación de *data product types* no está permitida.

2) *Calibration Level*: El nivel de calibración provee información del estado de procesamiento o reducción de los datos. Puede tomar un valor entero de acuerdo a los siguientes niveles:

- **Nivel 0**: datos instrumentales en bruto, en un formato definido por su propietario o un *data provider* interno. Necesita herramientas específicas de los instrumentos para ser manejados.
- **Nivel 1**: datos instrumentales en formato estándar (FITS, VOTable, SDFITS, ASDM, etc) los cuales pueden ser manejados por paquetes astronómicos estándar.
- **Nivel 2**: datos científicos calibrados, con la firma del instrumento removida.
- **Nivel 3**: *data products* mejorados. Puede representar la combinación de datos de múltiples observaciones.

3) *Observation*: Una observación puede estar compuesta por múltiples *data products* individuales. En estos casos, todos los *data products* que comprenden la observación deberían compartir el mismo identificador de observación. Pueden tener distintos valores de *data product type*, *calibration level*.

4) *File Content and Format*: Se refiere al contenido del archivo y a lo que es en realidad el archivo. Es importante conocer esto en caso de decidir si descargar o no un archivo en determinado formato y contenido.

C. Implementación de ObsCore en un Servicio TAP

El modelo *ObsCore* debe ser implementado en los servicios de TAP de manera que las consultas válidas puedan ser ejecutadas sin cambios en cualquier servicio que implementa el modelo. Se permiten columnas adicionales siempre y cuando se especifiquen todas las columnas obligatorias.

El modelo se especifica explícitamente mediante tablas de una base de datos relacional. En el contexto de TAP esto significa que se especifican tablas lógicas como las expuestas

en el TAP_SCHEMA. No implica que estas tablas tendrán la misma estructura que TAP, per la relación con la descripción TAP_SCHEMA es sencilla.

1) *Tipo de data product (dataprodut_type)*: Compuesto por una cadena simple que describe la naturaleza primaria del *data product*. Debería asumir uno de los siguientes valores en minúscula: **image**, **cube**, **spectrum**, **sed**, **timeseries**, **visibility** o **event** según lo explicado anteriormente. El valor NULL es permitido especificando un sub tipo que entregue mayor detalle en el campo opcional *dataprodut_subtype*.

2) *Nivel de calibración (calib_level)*: Muestra el nivel de calibración que ha sido aplicado para crear el *data product*. Puede tener asumir los valores enteros 0, 1, 2 y 3. No puede tener valores nulos.

3) *Nombre de la Colección (obs_collection)*: Identifica la colección de datos a la que pertenece el *data product*. Una colección de datos (*data collection*) es un conjunto de datos que es semejante de alguna manera. El nombre puede estar compuesto por el instrumento que genera los datos y por el *facility name*, por ejemplo en el formato siguiente CHANDRA/ACIS-S. No es un atributo definido. Lo importante es que sea familiar para los astrónomos y discriminativo para apundar a un conjunto de datos de interés. No puede tener el valor NULL.

4) *Identificador de la Observación (obs_id)*: Define un identificador único para la observación. En caso de que múltiples *data products* estén disponibles para una observación, cada uno de sus componentes debe llevar el mismo valor para *obs_id*. Este valor debe mantenerse inalterable a través del tiempo para futuras referencias. No puede tener el valor NULL.

5) *Identificador del conjunto de datos del publicador (obs_publisher_did)*: Contiene el identificador del conjunto de datos IVOA para el *data product* publicado. Debe ser un valor único dentro del espacio de nombres controlado por el conjunto de datos del publicador (*data center*). Debe mantenerse inalterable con el paso del tiempo. No puede ser nulo.

6) *Acceso URL (access_url)*: Contiene la URL que puede ser usada para descargar el *data product* como archivo de algún tipo. No está garantizado para permanecer válido de manera indefinida.

7) *Formato de Acceso (access_format)*: Especifica el formato del *data product* si es descargado como un archivo. Es importante para el *data discovery* y para el cliente, pues evaluarán si son capaces de usar los datos si son descargados. El valor para esta columna puede ser un tipo MIME, MIME estándar, un MIME extendido o un nuevo tipo MIME personalizado definido por el *data provider*.

8) *Tamaño estimado de descarga (access_estsize)*: Contiene el tamaño aproximado (en KiloBytes) del archivo disponible vía *access_url*.

9) *Nombre del objetivo (target_name)*: Contiene el nombre del objetivo de la observación, si lo hay. Típicamente corresponde al nombre del objeto astronómico, pero podría ser el nombre de un *survey field*.

10) *Coordenadas centrales (s_ra, s_dec)*: El sistema de coordenadas usado es ICRS. La columna *s_ra* especifica el *ICRS Right Ascension* del centro de la observación. La columna *s_dec* especifica el *ICRS Declination* del centro de la observación.

11) *Extensión espacial (s_fov)*: Contiene el tamaño aproximado de la región cubierta por el *data product*. En regiones circulares corresponde al diámetro, en otros *data products* es un valor lo suficientemente grande como para contener toda el área de observación. Si no están bien definidos los límites se utiliza un valor característico.

12) *Cobertura espacial (s_region)*: Puede ser usada para especificar precisamente la cobertura espacial de un *data product*. Es a menudo una representación exacta o casi exacta de la región iluminada de una observación dada.

13) *Resolución espacial (s_resolution)*: Especifica un valor de referencia escogido por el *data provider* para la resolución espacial estimada del *data product* en arco segundos. Corresponde al elemento espacial más pequeño en la señal observada o a un valor típico de resolución espacial estimado en el caso de que el muestreo de la frecuencia espacial sea complejo.

14) *Límites de tiempo (t_min, t_max)*: Contiene el tiempo inicial y el tiempo final de la observación, especificados en MJD. En caso de combinación de múltiples estructuras se considera el menor *t_min* y el mayor *t_max*.

15) *Tiempo de exposición (t_exptime)*: Para exposiciones simples corresponde a *t_min-t_max*. En casos donde el detector no está activo en todo momento *t_exptime* ; *t_max-t_min*. Si no es constante en el tiempo se utiliza el tiempo medio de exposición. En algunos casos es un indicador de la sensibilidad relativa dentro de un *data collection*.

16) *Resolución de tiempo (t_resolution)*: Es el mínimo intervalo interpretable entre dos puntos a lo largo del eje temporal.

17) *Límites espectrales (em_min, em_max)*: Especifican el menor y el mayor de los valores espectrales observados, expresados como una longitud de onda de vacío en metros.

18) *Poder de resolución espectral (em_res_power)*: Es un valor adimensional definido como $\frac{\lambda}{\delta\lambda}$.

19) *Descripción de ejes observable (o_ucd)*: Especifica un UCD describiendo la naturaleza de lo observable dentro del *data product*. Cualquier valor UCD válido es permitido, si no hay uno apropiado el campo se deja nulo.

20) *Columnas adicionales*: *Service providers* pueden incluir columnas adicionales en la tabla `ivoa.ObsCore` para exponer metadatos adicionales. Estas columnas deben ser descritas en el la tabla TAP_SCHEMA.columns y en la salida de *VOSI-tables resources*.

III. IMPLEMENTACIONES

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