

ObsCore-ObsTap-Datalink

Laurent Michel - Strasbourg

L'Observatoire Virtuel

Journée « L'Observatoire Virtuel pour les débutants »
Vendredi 17 mars 2017



Avec le soutien de l'AS OV France

Obscore: les modèles de données

- **Pourquoi des modèles**

- **Travail de réflexion pour décrire de manière univoque les grandeurs portées par les données circulant au sein du VO**
 - STC: description générale des principaux axes de mesure
 - Espace, temps, polarisation, énergie
 - Non destiné à être implémenté
- **Documentation des champs des réponses retournées par les services**
 - SpectrumDM: description de spectres
 - Les réponses SSA sont dérivées de ce modèle
 - Peut être utilisé pour annoter des données
- **Annotation de données exposées**
 - Obscore: description d'un produit d'observation
 - Pas de service retournant spécifiquement des données Obscore, mais SIAV2 utilise le modèle TAP dans ses réponses
 - Utilisé comme vue sur des données natives

Obscore

- **Les services simples exposent des jeux de données homogènes**
 - Spectres pour SSA
 - Images pour SIA
 - Une requête sur un service SIA ne peut être appliquée à un service SSA
 - Difficulté à récupérer des données de nature différentes mais répondant aux mêmes critères.
- **Et TAP vint...**
 - TAP est taillé pour des données hétérogènes
 - Exposition de multiples tables de données
 - Mise à disposition d'un langage de requêtes avancé
 - Auto-description du service
- **... mais**
 - Difficulté pour retrouver des données de format différents hébergées dans des tables différentes
 - Demande une analyse très poussée du méta-schéma pour de requêtes très compliquées

ObsTap: l'idée

- **Proposer un moyen de découvrir des données dans un service TAP**
 - Méthode simple
 - Méthode universelle
 - La même requête doit pouvoir fonctionner sur tout site TAP proposant une interface Obscore
 - Une seule table avec des colonnes bien définies
 - Les noms et la localisation de cette table sont bien définies
 - Cette table regroupe des observations de natures différentes.
- **Exemple de requête**
 - `Trouver tous les spectres et toutes les images`
`Couvrant les longueurs d'onde 1.e-6 à 2e-6 m`
`Faites par l'instrument 'INSTRUMENT'`
`De l'observatoire 'OBSERVATOIRE'`
`Entre le 1/1/2004 et le 1/5/2005`

Obscore Vs ObsTap

- **ObsCore est un modèle de données**

- Description homogène des axes de mesures suivants
 - Espace, temps, energie, polarisation, observable
- Description de l'observation
 - Observation
 - Cible
 - Fichier d'observation
 - Données (axe DATA)
 - Curation
 - Acces
 - Provenance

- **ObsTap est l'aplatissement de ce modèle sur une table relationnelle.**

- La hiérarchie des classes disparaît
- Chaque ligne correspond à une observation
 - Observation Obscore
 - Un fichier de données à usage scientifique.
 - Ne correspond PAS aux données obtenue lors d'un pointage d'un télescope
- La table se nomme **obscore** et doit se trouver dans le schéma **ivoa**

Obscore: la partie observation

| Observation | Cible | Data | curation | Access | Provenance |
|--------------------------|---------------------|-------------------|-------------------|----------------|-----------------|
| data_product_type | target_name | obs_id | obs_release_date | access_url | facility_name |
| data_product_subtype | target_class | obs_title | obs_publisher_did | access_format | instrument_name |
| | | obs_collection | publisher_did | access_estsize | proposal_id |
| | | obs_creation_data | bib_reference | | |
| | | obs_creator_name | data_rights | | |
| | | obs_creatort_did | | | |

Obscore: la partie mesures

| spatial | temps | spectral | observable | Polarisation |
|------------------|----------------|-------------------|----------------|--------------|
| s_xele1/xele2 | t_xele | em_xele | | |
| s_ucd | t_ref_pos | em_ucd | o_ucd | pol_states |
| s_unit | t_exp_time | em_unit | o_unit | |
| s_calib_status | t_calib_status | em_calib_status | o_calib_status | |
| s_ra/dec | t_min | em_min | o_stat_error | |
| s_region | t_max | em_max | | |
| s_fov | | | | |
| s_pixel_size | | em_res_power | | |
| s_resolution_min | | em_res_power_min | | |
| s_resolution_min | | em_res_power_maxe | | |
| s_resolution | t_resolution | em_resolution | | |
| s_stat_error | t_stat_error | em_stat_error | | |

DataLink

Datalink

- **Découverte de données par un service ObsCore**
 - Sélection sur les colonnes exposées
 - Récupération de l'observation par l'**access_url**
- **Les limites du download**
 - Le fichier peut être trop gros
 - Il peut être disponible sous plusieurs formes
 - Previews
 - Version dégradées
 - Il peut être généré à la volée par un service nécessitant des paramètres
 - Il peut bénéficier d'un traitement scientifique
- **Questions**
 - Comment attacher plusieurs URLs à une observation contenue dans la table Obscore
 - Comment décrire l'utilisation de ces URLs
- **Réponse**
 - DataLink

Datalink

Définition d'une déclinaison particulière de VOTable contenant des liens plutôt que des données

| Valeur de <code>access_format</code> | Ce qui est retourné |
|---|-----------------------------------|
| <code>image/png</code> | Une image PNG |
| <code>application/fits</code> | Un fichier FITS |
| <code>application/x-votable+xml</code> | Une VOTable avec des données |
| <code>application/x-votable+xml;content=datalink</code> | Une VOTable avec des liens |

Datalink

- **Datalink est un service**

- Il est déclaré comme tout service VO (VOSI, DALI, ...)
- Il s'applique à un seul (ou à une liste) fichier d'observation(obs_publisher_did)
- Le fichier à traiter est référencé par son identifiant VO

- **Exemple d'URL**

- Service:

<http://my.data.link/online>

- Pour appliquer ce service à l'observation identifiée par **AZERTY**

<http://my.data.link/online?ID=AZERTY>

Le service retourne une VOTable dont chaque ligne de données correspond à un lien

Datalink: definition d'un lien

| name | description | required | UCD |
|----------------|---|----------|---------------------|
| ID | Input identifier | yes | meta.id;meta.main |
| access_url | link to data or service | one only | meta.ref.url |
| service_def | reference to a service descriptor resource | | meta.ref |
| error_message | error if an access_url cannot be created | | meta.code.error |
| description | human-readable text describing this link | no | meta.note |
| semantics | Term from a controlled vocabulary describing the link | yes | meta.code |
| content_type | mime-type of the content the link returns | no | meta.code.mime |
| content_length | size of the download the link returns | no | phys.size;meta.file |

Datalink: Types de liens (RFD)



| | | | |
|----------------|--------------|-----------------------------|---|
| #this | | the data itself | the primary (as opposed to related) data of the identified resource |
| #progenitor | | Progenitor | data resources that were used to create this dataset (e.g. input raw data) |
| #derivation | | Derivation | data resources that are derived from this dataset (e.g. output data products) |
| #auxiliary | | Auxiliary | auxiliary resources |
| #weight | #auxiliary | Weight map | resource with array(s) containing weighting values |
| #error | #auxiliary | Error map | resource with array(s) containing error values |
| #noise | #auxiliary | Noise map | resource with array(s) containing noise values |
| #calibration | | Calibration data | resource used to calibrate the primary data |
| #bias | #calibration | Bias calibration data | used to subtract the detector offset level |
| #dark | #calibration | Dark calibration data | used to subtract the accumulated detector dark current |
| #flat | #calibration | Flat field calibration data | used to calibrate variations in detector sensitivity |
| #preview | | Preview | low fidelity but easily viewed representation of the data |
| #preview-image | #preview | Image preview | preview of the data as a 2-dimensional image |
| #preview-plot | #preview | Plot preview | preview of the data as a plot (e.g. spectrum or light-curve) |
| #proc | | Processing | server-side data processing result |
| #cutout | #proc | Cutout | a subsection of the primary data |

Datalink: Les services paramétrés



- **Certains liens peuvent référencer des services paramétrés**
 - Il peut 'agir de services VO
 - SIA, SODA
 - Ou de services libres
 - Cutout, fits de spectres,...
 - Ce type de service doit être décrit de manière à ce que le client sache comment l'utiliser
 - Génération d'un URL valide
 - Saisie de valeurs plausibles de paramètres
- **Les services paramétrés sont décrits dans des ressources spécifiques**
 - Ces ressources sont identifiées par un ID
 - Elles sont référencé par le champs service_def des liens les utilisant
 - Les paramètres sont décrits à la manière des VOTables
 - Échelle de validité
 - Énumération de valeurs possibles
 - Valeurs par défaut
 - Un paramètre particulier, **ID**, contient l'identifiant du produit à traiter
 - Cette ressource est un **service descriptor**

Exemple d'interface générée à partir d'une réponse Datalink

Link #science

  *FITS, scaled by 1/4*

Link #proc

Processing server-side data processing result

ID *The publisher DID of the dataset of interest*

DEC *The latitude coordinate*

Value **-2.505**

RA *The longitude coordinate*

Value **80.446**

POLYGON *A polygon (as a flattened array of ra, dec pairs) that should be covered by the cutout.*

FREE

CIRCLE *A circle (as a flattened array of ra, dec, radius) that should be covered by the cutout.*

FREE


PIXEL_1 *Pixel coordinate along axis 1*

Value **11239**




PIXEL_2 *Pixel coordinate along axis 2*

Value **14298**

KIND *Set to HEADER to retrieve just the primary header, leave empty for data.*





Link #this

   *the primary (as opposed to related) data of the identified resource*

The full dataset.

Link #preview

  *low fidelity but easily viewed representation of the data*

A preview for the dataset.

Datalink: Décrire un Datalink dans une VOTable

- **Un service Datalink peut être attaché à une VOTable de données**
 - Le service DataLink es décrit par un **service descriptor**
 - Le paramètre ID, identifiant un produit particulier, contient la référence à la colonne de la VOTable utilisée pour identifier les produits