

Résumé de la discussion "mapping metadata CTA dans les protocoles VO"
Strasbourg 10/10/2022

<https://indico.obspm.fr/event/1489/>

Participantes, participants : Paula Kornecki, Pierre Cristofari, Catherine Boisson, Ada Nebot-Gomez, Mireille Louys, Mathieu Servillat, Bruno Khelifi, François Bonnarel

Première partie : retour sur les réunions de mars (CEVO) et Juin (homework français)

- Prototype HESS de service mappant des keywords gamma Open data format vers ObsCore
- [Synthèse d'une future note IVOA](#)
- Cas d'utilisation à affiner
- Caractérisation des axes à compléter
- Provenance logicielle (analysis, releases)
- Provenance : observation
- Modèle d'un Event list (réunion ASOV du 11/10)

Cas d'utilisation

Nous avons évoqué en juin un grand nombre de critères de sélection souhaités .

Extraits :

Use cases HE qui fonctionnent avec ObsCore

Cone Search

attention : e_min, e_max en unités m dans ObsCore

exemple service australien (ajout post meeting, mireille) Data Central

<https://datacentral.org.au/search/cone/>

facility name, instrument name permettent aussi des recherches simples

Use cases HE qui nécessiteraient une extension

recherche temporelle

GTI, décrire le support sur l' axe des temps (au sens de Characterisation)

recherche spectrale

e_min, e_max en TeV ??? intérêt pour le provider (peut être ds une table annexe pour la construction de la base)

cf Exemple d'affichage données HESS dans TAP Handle

Ce sont des uses cases très génériques. Nous aimerions aussi avoir des scenarii d'utilisation composés de successions de combinaisons de requêtes dans un objectif complexe précis. Nous avons donc modifié en séance des propositions de UserScenario proposées par le CDS

Voici les US modifiés (pour être plus corrects) avec commentaires

- US 1: detection
 - At some location in space and time I have detections in various wavelengths/ spectral domains. This is a kind of SED.
 - a) I look for CTA datasets at these same space/time locations. If there is nothing there I know the source has flux lower than the sensitivity limit.

b) I modelize the source to fit the known SED. I infer some flux in gamma-ray. I look for datasets at this space/time location where the sensitivity limit is lower than this flux. To validate or invalidate the model.

—> il y avait une confusion entre la notion d'upper limit qui est caractéristique d'une source (son flux ne peut être qu'inférieur à cette limite) et la limite de détection possible dans le dataset (sensitivity limit)

- US2: Starting point is discovered DL5 dataset (a spectrum)
 - a) We discover GTI of the progenitor(s) DL3 datasets
 - b) We look for datasets in gamma ray and other wavelengths (or messengers) with time bounds overlapping the GTI.

—> Un autre US proche a été proposé. lequel ?

- US3: For a given run look for all derived DL3 obtained by DL3 hard analysis. After checking them look, for DL3 data obtained with other analysis modes (soft, etc...) for some of them

—>Reprocesser à partir du DL2 (user scenario initial) ne sera pas rendu public.

Nous nous sommes mis d'accord pour que les équipes LUTH/APC nous fournissent des user scenarii de leur cru

Caractérisation des axes :

On a rediscuté de potentiels champs à mettre dans un Obscore extension pour satisfaire certains use cases.

- energy_min, energy_max en Kev/Mev/Gev: (en plus de em_min, em_max qui resteraient en mètres). Pour certains dans le VO ça se gère par des user defined fonction (de transformation entre les quantités voulues (les *ev) et les quantités stockées (les m). Pour ML+FB ce sont des champs nouveaux dupliquant l'information du standard de base. BK (et autres) ne comprend pas pourquoi on impose que ces em_min, em_max en m. Il suffirait de préciser la quantité/unité dans em_unit.
- t_support qui serait la concaténation des GTI (format interval ou moc, on s'en fiche à ce stade). Utile pour une recherche fine des contreparties CTA avec des données très datées provenant d'ailleurs. —> ne semble pas vraiment nécessaire . Une fois qu'on aura les candidats raisonnables avec le t_min, t_max on recherchera les GT dans les données pour valider ou pas le choix. t_exposure est la durée du support, ce n'est pas la détermination du support temporel.
- o_upperlimit ou o_min ou o_sensitivity_limit. Il y avait une confusion. L'upperlimit s'attache à une source. Son flux dans une bande est supposé être inférieur à cette valeur. Une notion caractérisant le dataset serait le o_sensitivity, la sensibilité du résultat de l'observation. Aucun flux inférieur à cette valeur n'est détectable.

Cette dernière discussion nous a conduit à nous demander si on discutait de catalogue ou de data. En fait de sources ou de datasets. Mais un catalogue comme ensemble de mesures sur des sources est lui-même un dataset qui peut être découvert par ObsTAP/ObsCore comme une image ou un cube. L'ESO l'a fait. On a fait une petite démo depuis Aladin.

En fait pour l'équipe LUTH il s'agit aussi de revisiter les champs ObsCore de base non aboutis.

Deuxième partie : provenance

- Mathieu présente à nouveau le concept de last_step_provenance rebaptisé "one_step_provenance" parce que cette vue du modèle n'est pas réservée à la dernière étape. Mais la vue a évolué quelque peu récemment au Luth. L'équipe nous propose une vision de l'activité centrale du "one_step" reliée à diverses entités. Bien sûr il y a les entity "used" et les entity "generated" mais aussi le software qui est vu comme une entité (reliée également à l'activity description" de même que l'environnement logiciel et même la détaillée provenance vue comme un log précis des différentes sous étapes de l'activité centrale. Le lien pour cette entité là serait aussi de type wasgeneratedby.
- FB et ML ont relu des parties de la discussion de Juin. Notamment sur les modes de l'observation : "tracking" (pointed, deriving), "scan" (...), pointing (on-off, wobble). Ils proposent de faire de chaque observation une activité (activityDescription de type observation) avec ses paramètres "tracking mode, "scan mode, pointing mode. Cette activité est au sommet de la provenance du dataset. Elle peut utiliser (rôle observationConfig) des facility et des instruments vus comme des entity et dont le nom permettra de pointer vers le service de facility/instrument en cours d'élaboration dans le VO. On pourrait simplifier en voyant cette observationConfig comme un simple jeu supplémentaire de paramètres de l'observation. En terme d'implémentation cette partie observationnelle de la provenance pourrait être intégrée à un service provenance. On imagine aussi l'intégrer à une table ObsCore extension, mais cela se révèle trop compliqué, trop sujet à la variabilité.
- En fait l'ensemble de la description de l'observation et du processing/analyse peut se résumer dans le concept de release, associée à un workflow. Ce terme de release, que nous hésitons depuis juin à savoir où placer quelque part dans ObsCore serait un bon point d'entrée à une interrogation d'un service provenance pour récupérer cette description du workflow.
- comment relier les descriptions ObsCore à la provenance du dataset ? → DataLink, ProvSAP, ProvTAP.