

Compte-Rendu de la réunion biblio Jeudi 19 mai 2016 14h15

Présents : Caroline Bot, Marianne Brouty, Mihaela Buga, Aline Eisele, Sylvain Guéhénneux, Gilles Landais, Soizick Lesteven, Cécile Loup, Anaïs Oberto, Emmanuelle Perret, Evelyne Son, Patricia Vannier,

Excusés : Catherine Brunet, Magali Neuville, Pierre Ocvirk, Fabienne Woelfel

Ordre du Jour :

- L'augmentation du nombre de petites tables préparées en amont du CDS dans les journaux de l'AAS (les "Machine Readable Tables MRT") et les conséquences pour VizieR et SIMBAD
- Les =s= : quelle est la politique pour les sources citées dans ces tables ? un statut =s2= signifie-t-il en pratique que les objets ne seront jamais entrés dans SIMBAD (comme la plupart des =g2=) ?
- Le point sur les périodiques et leur visibilité (il y a ceux qui sont difficilement accessibles en ligne, ceux qui n'existent que sous forme papier, ...)
- L'affichage web des listes d'objets dans SIMBAD (certaines colonnes très larges gênent la lisibilité)
- Qu'en est-il des mesures de périodes ? les rentre t on toujours ou non ?
- gsc4sim : définir proprement une procédure pour les demandes de nouveaux catalogues
- et bien sûr le tour de table habituel ...

Les points ne seront pas traités dans l'ordre.

1. L'augmentation du nombre de petites tables préparées en amont du CDS dans les journaux de l'AAS (les "Machine Readable Tables MRT") et les conséquences pour VizieR et SIMBAD

Ces tables MRT, préalablement préparées par l'AAS, sont accessibles via les articles des journaux de l'AAS. Ces tables, généralement longues, ne sont pas traitées par les DJINistes mais passent par VizieR puis Cosim. Or, de plus en plus de petites tables sont publiées en MRT.

Si Emmanuelle ne voit pas que la MRT est en fait une petite table, elle part dans le circuit « habituel » et ne sera traitée que quelques mois plus tard. (Le traitement d'une « petite » table pour VizieR prend autant de temps que celui d'une grande table).

Il faudrait faire remonter la petite table à Magali dans ce cas.

Il se pose la question du traitement de ces petites tables MRT.

Problèmes rencontrés :

- Comment différencier ces tables dans les articles ?

Des pistes sont envisagées :

Il y a une différence de définition des tables en MRT : celles qui ne sont que partiellement visibles dans l'article sont notées « entirely available online », ce qui permet aux Djiniistes de savoir s'il y a du contenu supplémentaire à signaler ou non (en rentrant les expressions régulières dans DJIN).

Mais comment connaître le nombre de lignes ?

- Recontacter l'AAS pour qu'ils mettent le nombre de ligne des tables en caption ?
- Anaïs pense que DJIN pourrait calculer le nombre de lignes.
- Gille pense aussi qu'une automatisation serait possible, Thomas Delacour travaille sur un programme qui récupère les informations en ligne.
- Anaïs propose de charger les données attachées à l'article en local et que DJIN analyse ces données (nombre de lignes des tables).

- Doit-on rentrer ou non une petite table partiellement dans l'article ?
Il faudra en rediscuter, notamment avec Pierre O.

Il faudrait demander à Thomas d'installer la nouvelle version d'Acrobat reader sur les machines de tout le monde qui est plus pratique que « evince » et permet notamment l'affichage de tous les liens hypertextes . A demander à Mark.

2. Les =s= : quelle est la politique pour les sources citées dans ces tables ? un statut =s2= signifie-t-il en pratique que les objets ne seront jamais entrés dans SIMBAD (comme la plupart des =g2=) ?

Cécile : Comme les =g2=, les =s2= sont à faire en priorité 2 car ils ont peu d'impact scientifique. Etant donné la charge de travail cela signifie effectivement que les =s2= ne sont pas fait (seuls les =s2+= ont une chance d'être faits) et tans pis si certaines références en =s2= ne sont pas complètement traitées.

Donc, on ne rentre plus TOUT dans SIMBAD.

Il n'y a plus d'exhaustivité et SIMBAD devient une base d'objets d'intérêt.

En cas de recitation d'une référence en =s2= ou =g2=, s'il y a beaucoup d'objets recités on remet un =s= pour que le cas soit réanalysé. S'il n'y a qu'un ou quelques objets recités on demande à l'astronome référent.

Au final il y a très peu de =s2=. L'intérêt des objets est inversement proportionnel à la taille de la table : les petites tables contiennent souvent des objets importants.

Ils ont conscience que lorsqu'ils classent une table en =s2= les objets ne seront pas rattachés dans Simbad. Le même objet cité dans le texte sera rentré dans SIMBAD.

Autant que possible les objets d'une table en =s= doivent être rentrés dans Simbad par les Djinistes, sinon tant pis.

Evelyne évoque les cas d'objets cités tels que « in the direction of the Galactic Center », ou les énumérations d'objets non étudiés « except HD 123, HD 245 »).

Cécile est d'accord pour dire qu'il ne faut pas les rentrer, quand on les rencontre, mais ce n'est pas grave de les rattacher à la référence s'ils ne sont pas évidents.

Pour ces cas, Cécile fait confiance à l'expertise des Djinistes pour prendre la bonne décision, et le cas échéant contacter l'astronome référent. En cas de doute ultime on peut demander à Cécile.

3. Qu'en est-il des mesures de périodes ? les rentre t on toujours ou non ?

En fait, la question est plutôt : rentre-t-on toujours les « vartype » de la même manière ?

Avant on se référait aux mesures du GCVS (qui est une compilation) pour rentrer les vartypes, mais cette compilation est dépassée, sans refs d'origine... Désormais on rentre les vartypes tels qu'ils sont donnés par les auteurs, même si du coup ce ne sera pas homogène dans SIMBAD.

Cécile traduit systématiquement le type de variable en vartype dans le commentaire de travail sinon voir avec elle.

Avec un type de données hétérogène, la recherche par vartype n'est plus possible. Les astronomes en sont conscients.

Se pose alors la question de l'utilité de la mesure. Elle reste utile au cas par cas pour préciser sur quel type d'objet porte la mesure de période ou de variabilité.

Marianne pointe le fait que par conséquent la recherche par vartype n'est pas possible, Cécile en est consciente mais ce n'est pas très grave car ce n'est pas un critère de tri pour les utilisateurs. Cécile rappelle que Simbad n'a pas vocation à être un catalogue homogène, mais une compilation de ce qu'on trouve dans la littérature.

Point sur la Mesure de vitesse LSR :

Les Vlsr vont être rentrés en données fondamentales et en data mais seulement via les tables.

Ceci car il n'y aura pas de qualité mais obligatoirement une incertitude de mesure déterminée par les astronomes.

Attention : HRV n'est pas Vlsr. Cécile a récupéré toutes les références en attente des Vlsr et Laurent est en train de vérifier les tables Vizier vs SIMBAD pour corriger les HRV qui seraient en fait des Vlsr.

Cécile et Caroline envisagent de rediscuter sur la manière d'entrer les erreurs et incertitudes de mesure des VLRS, en particulier pour le milieu interstellaire.

Pour les Djinistes : petites tables avec VLRS à signaler en =s=.

N.B. : Depuis la réunion, mail de Cécile pour suspendre l'entrée des Vlsr car il manque une dimension importante dans la mesure...

3. Le point sur les périodiques et leur visibilité (il y a ceux qui sont difficilement accessibles en ligne, ceux qui n'existent que sous forme papier, ...)

Liste des périodiques pour lesquels sont rencontrées des difficultés de traitement, ainsi que leurs priorités :

Aline :

-Acta Astronomica (AcA) problème d'accessibilité au site internet (normalement en année N+1, mais pas homogène) , priorité 2

-Chinese Journal of Astronomy and Astrophysics (ChA&A, anciennement AcASn) intraitable avec DJIN , en priorité 2

-Astrophysics (Ap) intraitable avec DJIN, en priorité 2

- PASJ : disponibilité des listes biblio aléatoires depuis que ce journal a été repris par le publisher du MNRAS, priorité 1

Magali :

- BaltA ? En priorité 2

- GCNR non reçus depuis février , en priorité 3

De manière générale, il serait bon de revenir sur la politique de traitement des petits journaux.

Certains sont traités « historiquement », ont-ils encore une légitimité scientifique ? Certains paraissent légitimes car recités, cela peut être un critère de choix ?

Il n'y a pas de contre-indication à utiliser le preprint pour la mise à jour dans Simbad (pour les AcA notamment).

Il faudrait demander les fichiers aux éditeurs, notamment pour le BaltA, en leur signalant l'enjeu : plus de traitement dans SIMBAD. Voir la liste avec Mark pour lancer des discussions.

GCNR : Soizick a récupéré récemment des GCN, elle peut peut-être récupérer les GCNR de la

même manière.

Pour le ChA&A et Ap, Aline va les renvoyer à Anaïs qui va faire des tests dans DJIN.

Pas de remise en question scientifique du traitement de ces journaux, notamment des chinois, qui peuvent monter en puissance. Il n'est pas possible à un astronome de présélectionner les articles à traiter sur la base des titres donc il faut continuer de les traiter en trouvant des solutions techniques.

4. L'affichage web des listes d'objets dans SIMBAD (certaines colonnes Trop larges, d'autres compactées sur 2 lignes → gêne la visibilité)

Tous les critères sont difficiles à prendre en compte mais Anaïs va voir ce qu'elle peut faire pour éviter les noms d'objets et les coordonnées sur 2 lignes.

6. gsc4sim : définir proprement une procédure pour les demandes de nouveaux catalogues

Il est décidé de faire un groupe de travail avec Evelyne, Gilles, Mihaela et Cécile.

Gilles ne pourra pas s'y consacrer dans l'immédiat mais au moins établir un cahier des charges et le projet pourra être confié à un informaticien en CDD par exemple.

Gilles propose de refaire entièrement le programme gsc4sim plutôt que de le corriger.

En parallèle on discutera de la création de petits programmes permettant d'extraire les commandes pour la mise à jour de SIMBAD à partir des tables de Vizier (Marianne demande en particulier d'avoir ce service disponible pour le catalogue LAMOST).

Evelyne demande s'il est possible d'ajouter systématiquement aux catalogues de Vizier les infos de longueur d'onde et qualité des coordonnées dans le commentaire de travail de la référence d'origine des tables. Cécile dit que ce n'est pas possible pour tous. Elle essaie de décrire au maximum les données dans le commentaire de travail de la référence d'origine. Elle ne le fait pas systématiquement pour les =g2= par manque de temps. Cette info n'est pas non plus disponible pour les catalogues « yCat ». En cas de besoin on peut lui demander et elle ajoutera les infos au cas par cas.

Gilles aimerait créer plus de liens entre SIMBAD et Vizier, notamment quand des données de Simbad viennent de Vizier, pour permettre de retrouver d'où viennent les informations. A rediscuter plus en détails. Cécile demande à ce que TOUS les « yCat » soient référencés dans Simbad.

La création d'un lien redirigeant vers NED quand on n'a pas l'objet au CDS est également évoquée.

7. Tour de table :

Evelyne re-pose la question d'un « bibcode » pour préciser qu'une donnée a été « estimée par l'équipe CDS ».

Cécile aimerait également qu'on puisse ajouter cette info pour NED et d'autres sources (à définir), à utiliser seulement en cas de nécessité absolue.

Anaïs peut autoriser une liste de mots à la place d'un bibcode.

Marianne : Fabienne a commencé à travailler sur le dictionnaire, elle remercie Anaïs pour son aide (et aussi pour avoir créé des bibcodes cliquables dans les commentaires).

Anaïs fait un point sur les nouvelles fonctionnalités proposées :

- demande de retours sur la nouvelle version de DJIN.

- création d'une commande qui permet de télécharger les pdf du MNRAS en local depuis la mise à jour de Simbad, fascicule par fascicule.
- ajout d'une fenêtre pop-up qui affiche le titre quand on passe la souris sur un bibcode.
- affichage du nombre d'objets avant le lancement d'une recherche par références, wildcards ou position.
- Dans le formulaire de recherche par coordonnées, elle va faire apparaître un Aladin lite qui donne visuellement le rayon de recherche indiqué.

Anaïs aimerait généraliser l'utilisation de Redmine (site de gestion des échanges entre suggestions et réponses, consultable par tous). Elle nous enverra les liens par mail.

Elle a monté beaucoup de dossiers avec entre autres des idées de développement pour DJIN, COSIM, Simbad web... Elle aimerait que tout le monde l'utilise (possibilité de voir les historiques de discussion par exemple, ce qui est impossible par e-mail).

Il s'agirait plutôt de remonter les idées d'amélioration ou des suggestions sur du long terme plutôt que des petits bugs qui peuvent être traités rapidement.

Elle propose de mettre un lien privé sur la page d'accueil du CDS pour y accéder plus facilement.

Anaïs recense des formations demandées (depuis longtemps) : VI, python...

Elle est prête à faire une formation VIM (VI amélioré) si besoin.

La demande de Philippe d'une formation SQL a été avantageusement remplacée par la possibilité de faire des recherches en full texte dans le commentaire de travail sur la page web.

Emmanuelle rappelle une demande de formation sur les fichiers FITS pour l'équipe Vizier (manipulation des fichiers, extrait des informations/descriptions).

Bibcodes et liens ADS/Vizier/SIMBAD :

Dans le cas où les bibcodes cités dans Vizier n'existent pas chez nous, ils atterrissent sur une page de « WARNING : not available at CDS ».

Il faudrait plutôt renvoyer vers ADS avec une indication du style « try ADS ». Mais attention aux liens morts.

En cas de requête d'un objet qui n'est pas dans Simbad, Anaïs propose de créer un lien redirigeant vers Vizier ou NED. Avec un commentaire explicatif « this object has no bibliography in SIMBAD, see in Vizier ». Car pour beaucoup d'utilisateurs, le fait que SIMBAD ne soit pas une base de données exhaustive est encore flou. Certains pensent que tous les objets de Vizier sont dans SIMBAD.

Dans le cas des bibcodes en yCat (pour les grands catalogues Vizier), il faut systématiquement créer la référence associée dans SIMBAD.

Caroline : cours sur les milieux interstellaire, un StuDs est en cours pour trouver une date avec les documentalistes. La formation sera ensuite ouverte à ceux qui le souhaitent.

Evelyne : Un ménage dans les DLA/Absorbers et masers serait utile, ces objets font beaucoup de bruit dans Aladin et SIMBAD.

C'est typiquement le genre de suggestions qu'on pourrait faire dans Redmine avec un dossier pour les opérations spéciales.

Mihaela : Avancées sur une interface graphique de Cosim ? Anaïs l'avait mis de côté mais propose de faire tester à Mihaela et Emmanuelle.

Soizick : LISA VIII en 2017 (Mai-Juin) à Strasbourg confirmé.

Thomas Delacour a bientôt fini la première des trois parties de son travail (l'entrée des références biblio dans SIMBAD) à partir des données des éditeurs . Les résultats sont concluants. La fusion Simbad/Biblio est la prochaine étape.