

Compte-rendu de la réunion de suivi CDS du 17 novembre 2003

Présents: M. Allen, L. Arbousse, T. Boch, F. Bonnarel, L. Cambrésy, S. Derrière, P. Dubois, P. Fernique, F. Genova, S. Lesteven, F. Ochsenbein, A. Preite-Martinez, B. Vollmer, M. Wenger.

Bibliographie

- Flot de données : à jour.
- 1000 bibcodes ont été vérifiés.
- Acronymes : le retard n'a pas été rattrapé. Prévoir l'arrivée des 200 acronymes pendant la période de Noël.
En ce qui concerne la participation de S. Lesteven et/ou F. Woelfel et C. Brunet à la création d'acronymes, tout dépendra du nouveau système. La partie qui doit être vérifiée par un astronome reste, mais en ce qui concerne la création elle-même ? Avec le nouveau programme à venir : faut-il former quelqu'un à l'actuel programme ou attendre la mise en place de la nouvelle version ? Celle-ci ne sera pas en place avant le printemps 2005. En attendant, F. Ochsenbein peut apporter une aide de manière ponctuelle. C'est un point à revoir à la mi-janvier.
- Dans le dictionnaire des acronymes : un élément indiquant la ou les colonnes où figurent les identificateurs permettant de faciliter les liens entre SIMBAD et Vizier est en cours d'installation.
- GC for SIMBAD : l'entrée automatisée des informations GSC et TYCHO dans SIMBAD inclut les corrections sur les magnitudes.
- L'outil Sesame interroge SIMBAD, NED et Vizier à la fois. Il serait utile de généraliser l'entrée semi-automatique de données des catalogues dans SIMBAD par copier/coller. Cette fonction serait à intégrer dans le programme de mise à jour.
- Réunion bibliographie à Paris la semaine prochaine pour P. Dubois et L. Cambresy pour rediscuter de certains points. Inquiétude sur le dépouillement des articles en ce qui concerne les systèmes d'absorption. S. Borde a repris les systèmes d'absorption. Elle crée les noms à partir d'articles d'origine, ce qui demande beaucoup de recherche bibliographique et prend énormément de temps.
A propos de ces objets, il y a eu de nombreux échanges téléphoniques et la situation s'est éclaircie. Discussion prévue pendant la réunion bibliographie.
- Réunion bibliographie prévue début décembre à Strasbourg.
- Projet MDA : S. Lesteven a envoyé les résumés de l'A&A et de l'ApJ 1995-1996-2001-2002. Y. Toussaint a déjà fait un travail sur A&A : extraction de mots, d'expressions du Thésaurus. Il y aura un post doc, en décembre, qui travaillera sur ce projet : organiser une réunion. J. Mothe a deux stagiaires (dont N. Hernandez) qui travaillent sur les mots clés d'A&A et d'ApJ. Faut-il une homogénéisation des mots clés ? Discussions téléphoniques.
- Formation pour S. Lesteven, au mois d'octobre, sur les archives ouvertes en bibliographie (ISIDORA), pour chercheurs, juristes et documentalistes. Le CCSDS est une archive ouverte hors INIST. Les documents seront à mettre sur Twiki dès qu'ils seront en ligne. Suggestion : ne pas signer les feuilles de transfert de copyright telles quelles, y indiquer des restrictions.
Remarque sur le Twiki : on peut demander à être prévenu en cas de modifications.

Contenu

Beaucoup d'actions au point mort :

- Pour ISO, la dernière version promise par A. Salama n'a pas été reçue.
- Pour AAVSO, il reste des choses à faire, qui n'ont pas avancé.
- PGC-LEDA : rien n'a été fait sur les vérifications prévues. C'est en attente de commentaires objets. J. Marcout peut peut-être en faire une partie.
- Parmi les choses faites : 1000 étoiles variables ont été vérifiées ainsi que 1000 étoiles carbonées. Les remarques de Mc Connel n'ont pas été intégrées. Il reste un millier de problèmes à régler sur ces étoiles carbonées. F. Woelfel a transféré ce travail à C. Brunet qui avait un catalogue de ces objets en cours d'intégration. Celle-ci s'est consacrée aux =g= qui n'avancent pas suffisamment et a traité les étoiles Ha. Nécessité d'arriver à un équilibre entre les deux. Beaucoup de temps passé sur de longues listes en rapport avec Mac Connel, IBVS.
- Il faut continuer à donner une priorité à la résolution des questions en attente en prenant quelques questions à traiter par semaine (quitte à les passer en =L= et les abandonner).
- Programme de mise à jour : parmi toutes les demandes faites, il reste à F. Bonnarel des problèmes de format d'où un retard sur la nouvelle version à tester. Discussion avec M. Wenger pour la mise en œuvre. Les tests devraient aller plus vite (il y en a moins que pour la première fois). On peut relire en mode visualisation (ligne à ligne).
- Contenu 2MASS :
87 sources dans la première release et pas dans la deuxième. Il faut les regarder une à une : à faire par L. Cambresy dès le mois de janvier.

Catalogues

- Le problème de reprise de certains catalogues en attente avec l'aide de L. Cambresy ne sera pas réglé avant janvier.
- Les gens de l'Opacity Project veulent installer la base Tip Top. Visite les 11 et 12 décembre de Mendoza et Zeippen.
- Contenu : les archives des missions.
 - Les associations HST sont maintenant accessibles (co-addition d'images). Les images sont d'une meilleure qualité : 1600/1600x32.
 - Le CGPS est aussi accessible ; on le voit à partir d'Aladin. Il faudrait au moins avoir un preview pour la démonstration.
- En ce qui concerne les grands catalogues :
 - UCAC 2 a été ajouté. Il permet de faire des réductions astrométriques très précises. Ne couvre que les $\frac{3}{4}$ du ciel pour le moment.Reprendre contact avec :
 - DPOSS 2
 - SDSS
 - Supercosmos
 - Registry VizieR OAI : S. Derrière doit discuter avec F. Ochsenbein. Le format par défaut est OAI-Dublin Core. On peut arranger les métadonnées sous une autre forme. Pour chaque élément, on renvoie une info en XML avec le format de métadonnées

demandé. Difficultés pour envoyer le format RSH (la dernière version ne marche pas encore tout à fait).

Comment faire comprendre à OAI qu'on veut autre chose que le Dublin Core ?

Chaque élément du Registry a un identificateur qui reste le même quelque soit le format de données. Appellation unique : oai.

Radio

- Une opération d'extraction des spectres à partir des données contenues dans Vizier est en cours. Elle concerne pour le moment 16 catalogues contenant 3.4 millions de sources. Durée de traitement : 30 minutes. Des spectres confirmant l'identification ont pu être construits pour 70000 sources, ce qui confirme 700000 cross-identifications qui devraient pouvoir être entrées dans SIMBAD. Le nombre de sources pour lesquelles l'opération peut être effectuée (il faut 3 points à différentes fréquences sur le spectre) est limité par la sensibilité à 6 centimètres. Le NVSS par exemple donne 2000000 sources à 1.4GHz.

NED emploie une méthode différente (par proximité en tenant compte du type d'objets pour les radio galaxies, ou des citations des objets dans les publications).

La pente des spectres est très bien répartie, et il est intéressant de rechercher les objets particuliers dans les ailes de la distribution des pentes (par exemple, recherche des spectres inversés). Il est difficile de prévoir l'extrapolation à Planck, 6cm étant assez éloigné de la première bande de Planck. Vérifier si l'opération de cross-identification en cours intéresse le projet Planck.

Statistiques : on retrouve 70% et 80% respectivement des sources 4C et B2 parmi les sources " avec spectre ".

Plusieurs questions ou remarques :

- 3C et 3CR ne sont pas encore dans Vizier (ils le seront début décembre).
- Il y a des objets en doublons dans le NVSS et le WENSS. Il faudra définir la procédure pour construire le lien de SIMBAD vers Vizier, les deux objets du catalogue pouvant avoir des mesures différentes. Le spectre permet de déterminer la bonne valeur, mais comme le lien SIMBAD-Vizier est un lien vers les données originales, il faudra donner accès à toutes les mesures (même problème dans le premier catalogue GSC où les étoiles brillantes apparaissent en plusieurs exemplaires).
A discuter avec F. Ochsenbein.
- Il faut convertir les angles de position en coordonnées galactiques en angles de position en RA-DEC. A inclure dans Vizier au même titre que les transformations de coordonnées.
- Question scientifique : quelle est l'origine des spectres inversés (YSO, AGN, régions HII optiquement épaisses)? A discuter avec A. Richards.

SIMBAD

- Définir le travail d'A. Oberto qui commence à travailler pour le CDS (et non plus pour le projet IDHA) le 1^{er} décembre.
- Equilibre à trouver entre les développements/corrections résiduels pour SIMBAD 3 et le développement de SIMBAD 4. Réunions à organiser avec, notamment, P. Dubois et S. Lesteven sur la « SIMBAD dream list ». Regrouper les idées de fonctions à faire évoluer sur SIMBAD.

- Une présentation va être faite en amphi pour le personnel CDS sur ce qui existe déjà ainsi que sur l'architecture. Il est encore trop tôt pour faire une réunion ouverte aux autres chercheurs de l'Observatoire.

Logiciels

- 2 stages sont en cours :
 - «développement de l'interface graphique de SIMBAD» B. Baranne.
 - « mise en place d'un cluster de calcul pour Aladin » T. Bucher. Le stage avance très vite : il y a déjà des résultats en terme d'installation de parties du code. Deux utilisations : l'interface d'A. Oberto a été modifiée par A. Oberto / T. Bucher et la nouvelle version d'Aladin utilise aussi depuis vendredi dernier le cluster.
Pas de tâche parallélisée pour le moment, seulement une distribution des tâches. Le stagiaire est là jusqu'en février. Priorité est donnée à un travail sur le cluster de calcul de VizieR. Il faudrait fixer une date de réunion avec l'équipe VizieR.
- Cluster pour catalogues : il s'est avéré que le cluster installé avec Clic n'était pas efficace car beaucoup de process sont lancés alors qu'ils font peu de choses. Pour le moment et afin de résoudre les problèmes de performances de la machine VizieR qui est trop chargée, les grands catalogues sont sur un PC. Une étude est en cours pour déterminer pourquoi les performances se limitent à 10 requêtes par seconde. Les tests sont faits sur des soumissions individuelles, le démarrage des process, etc... Au vu de la forte augmentation des requêtes, il faudrait trouver une méthode de dispatching efficace : un seul PC est un peu juste ; il faut pouvoir passer d'une machine à l'autre très facilement. D'un point de vue stratégique, toutes les forces doivent se consacrer à ce problème de catalogue. Le problème se pose de façon générale pour les services de la maison.
- En ce qui concerne les fonctions de cross match, T. Boch a avancé dans le cadre de la préparation de la démonstration (en backup du développement d'Astrogrid). Astrogrid veut le coupler avec l'extraction de catalogue à partir d'images. Il faut poursuivre cette étude qui doit déboucher dès que possible sur un service CDS.
S. Derrière prépare un outil qui prend 2 VOTable en compte (2 listes d'objets). Il renvoie les listes cross-identifiées en VOTable. L'outil est basé sur une interface Java. On pourra installer facilement un Web Service par-dessus. La mise en place d'un service de cross-identification est prévue depuis longtemps. Il y a eu trop de problèmes pour reprendre le logiciel de P. Ortiz : toute la partie description en Perl et toute la gestion des en-têtes sont extrêmement compliquées.
Il n'y a pas pour le moment de cross-match entre grand et petit catalogues. La démonstration consiste à faire une jointure par position.
Dans un premier temps, il faut envoyer toutes les contreparties jusqu'à une certaine distance puis, dans un deuxième temps, il faudra ajouter des fonctions supplémentaires d'identification (en utilisant les UCD,...). A moyen terme : remonter le code C en Java comme le RGB ?
Ce service devrait pouvoir fonctionner indépendamment d'Aladin. Le fait de pouvoir l'appeler à distance (et non seulement sur la machine locale comme dans Aladin) serait une bonne philosophie. Il faudrait avoir quelque chose de prêt à interfacier avec ACE pour la démonstration. La meilleure solution est le transfert des VOTable via Aladin.

- VOPlot : la dernière version prend en compte 2 tables séparées. Une ressource peut également comprendre plusieurs tables. La méthode de travail : avoir un cahier des charges très précis.
La version courante est intégrée dans Aladin. Capacités de VOPlot en terme de nombre de lignes : c'est lent après quelques milliers de sources; lenteur au niveau des stockages de données.
Mêmes limites pour Mirage. Il faudrait quantifier les limites de manière plus précise.
Action : insister sur la limite en nombre de sources ; insister sur la réalisation de tests scientifiques à Pune.
- Logiciel de la démonstration de juillet : il y a eu un écrasement par mégarde lors de l'installation de la nouvelle version d'Aladin. F. Genova va envoyer un message pour dire que les problèmes sont réglés maintenant. 4 serveurs images sont maintenus, celui de janvier 2003 va être arrêté. Ceci illustre les difficultés qu'il y a à maintenir un prototype sur le long terme.
- En ce qui concerne la démonstration de janvier 2004 : mettre les données dans le serveur image ASAP. Il faut plus de flexibilité dans le serveur image pour pouvoir changer la taille de l'image. Le problème : une taille maximum est codée en dur pour chaque serveur. La taille maximum est la taille des morceaux stockés. On pourrait facilement obtenir une image plus petite.
Pour la démonstration de janvier, on peut avoir une taille variable, plus petite que la limite. Il faut être prudents avec la taille maximum car les images sont très grandes.
SIA serveur ISO : il y a une version de test en ligne. De façon générale pour la démonstration : problème de définition des différentes tâches de manière plus précise ; il faut se mettre d'accord sur les différentes interfaces.

Aladin

- Données : J. Marcout continue l'intégration. Des données manquantes ont été identifiées : des astrométries du côté de la MAMA et des informations du côté DSS 2. J. Marcout prendra sa retraite en juillet.
- 2MASS : les données sont sur le disque mais ne sont pas encore installées.
Les headers FITS n'ont pas changé. L'intégration sera faite avant janvier 2004.
- DENIS : il faudra le mettre en test. Les images ont l'air bien et sont accessibles sur le disque RAID à Meudon.
- SLOAN : le transfert des données SLOAN se fera en décembre en utilisant le nouveau lien Gb d'Osiris en test. Il reste à installer le RAID. Le tout prendra environ 15 jours.
- IDHA VOTable utilisé dans Aladin 2.0.
Pour la nouvelle démonstration, F. Bonnarel, M. Louys ainsi que T. Boch sont en train d'étudier des extensions vers les données spectrales :
spectres 1D (données FORS2) ; cube de données : CGPS. Descendre plus bas vers les différents plans, les données résumées (à voir avec F. Ochsenbein). Voir ce qui est disponible comme image.
Cela va donner lieu à un fichier VOTable test, dont l'intégration dans le prototype AVO sera à faire par T. Boch.
- Soucis sur le RAID : celui-ci marche sur un seul contrôleur. Problème de pérennité du système RAID. Le contrôleur défectueux doit être remplacé. On a failli perdre 3 RAID 7 si on avait suivi les conseils de la maintenance. Intérêt d'avoir 2 contrôleurs !

- Aladin 2 est en ligne. Beaucoup de fonctionnalités issues de l'AVO étaient peu testées en mode Applet ou sur de vieilles machines. Cette version marche moins bien sur Mac. Le FAQ a été mis à jour et des modifications ont été apportées à la page d'accueil.

Pour le manuel : à voir. Les use case : C. Bot et B. Vollmer sont d'accord pour gérer cette page.

Concernant le copyright, une forte pression est faite pour que la provenance des images et des données soit très clairement indiquée. Le copyright doit être bien visible, en bas, dans la ligne de description. Reste à le remonter via les logs de mission.

- VOTable : le document de définition a été re-rédigé ; en attente des commentaires.
- UCD : travail en cours. Dans les catalogues ainsi que dans les colonnes, les relations entre UCD1 et UCD2 sont en cours de vérification. Il y a parfois des relations 1→N. 1/3 du travail a déjà été fait et le reste devrait être fini lundi prochain le 24 novembre. Des progrès ont été faits sur les différentes possibilités de définir le système. Il reste de la pédagogie à faire vers les utilisateurs.
- SPIE : la réunion se tiendra en juin à Glasgow. On a jusqu'au 24 novembre, date limite, pour soumettre les résumés.
- Publication du travail de B. Vollmer : explorer la possibilité de publier dans A&A.
- AAS : il y a un poster à faire. La réunion a lieu à Atlanta. Ceux qui désirent y aller doivent se mettre en relation avec F. Genova sans plus tarder.
- Planck : réunion WG8 la semaine prochaine à Londres pour préparer la réunion qui se tiendra du 26 au 30 janvier 2004 à Paris (suivi par L. Cambresy).
- A. Richards viendra 2x1 mois l'année prochaine.
- Création officielle de OV France la semaine prochaine.