

Compte-Rendu de la réunion de suivi CDS du vendredi 19 décembre 2003

Présents : M. Allen, L. Arbousse, F. Bonnarel, C. Bot, L. Cambresy, S. Derrière, P. Dubois, P. Fernique, F. Genova, F. Ochsenbein, A. Schaaff, B. Vollmer, M. Wenger.

Excusés : T. Boch, S. Lesteven

Bibliographie

- La réunion bibliographie prévue début décembre n'a pas pu être organisée.
- Statistiques sur l'activité bibliographie :
 - o Les "=g=" : sont à peu près stables. Il faut s'attendre à une forte remontée en janvier suite aux créations d'acronymes programmées en fin d'année.
 - o Les "=C=" : sont en augmentation mais certains devraient être pris en compte par l'entrée systématique des catalogues radio.
 - o Les "=f=" : il y en a 530 soit entre 100 et 150 de plus par an. Ils attendent relecture et sont chez P. Bauer qui en fait régulièrement.
Intervention de L. Cambresy à ce sujet.
- Réunion MDA du 9 décembre 2003
 - Travail sur les UCD : extraction des informations des ReadMe ; rechercher les UCD, les unités et les noms de colonnes. Valider la liste des UCD en les retrouvant à partir des informations contenues dans les ReadMe.
 - Travail sur le Thésaurus :
J. Mothe a un stagiaire chargé d'extraire des mots et des groupes de mots en utilisant Tétralogie et Syntex dans le but d'évaluer la couverture du Thésaurus et de le mettre à jour.
Le but du travail est d'utiliser les UCD et le Thésaurus en vue de construire une ontologie. Les logiciels d'analyse sont différents.
La prochaine réunion aura lieu le 12 janvier 2004 (à confirmer) ; vérifier la présence d'A. Preite-Martinez à Strasbourg le jour choisi pour la réunion.
→ S. Derrière a relancé M.-L. Dubernet pour les UCD de physique atomique et moléculaire. A. Preite-Martinez a extrait les UCD présents actuellement dans ce domaine et ils ont été envoyés à M.-L. Dubernet.
- Réunion bibliographie à Paris le 25 novembre dernier (L. Cambresy, P. Dubois).
Systèmes absorbants : la méthode pour rentrer les acronymes n'est toujours pas définie. Un nouveau catalogue de compilation pourrait régler le problème.
En attendant une décision, il ne faut pas créer d'acronymes pour l'instant.
S. Borde n'a pas encore envoyé les derniers acronymes préparés dans ce domaine.
- New Astronomy : passe sur Science Direct (action F. Ochsenbein)
- AAS : Téléconférence avec R. Kennicutt et G. Schwartz. Objet de la discussion : liens sur les noms d'objets dans les journaux de l'AAS. La discussion sera reprise lors de la réunion de l'AAS à Atlanta.
- Mise à jour : il n'y a pas encore de réactions de tout le monde. S. Laloë est satisfaite de la dernière version. Il faut tester rapidement, en particulier F.

Ochsenbein et P. Dubois qui s'occuperont du programme de mise à jour à Atlanta. Il reste une erreur à corriger sur la fusion des objets (détectée par G. Jasniewicz). F. Bonnarel propose de mettre le nouveau programme en opération uniquement pour S. Laloë dans un premier temps afin d'avoir un retour sur l'utilisation réelle. F. Bonnarel reste à Strasbourg et pourra intervenir en cas de besoin (le programme sera installé dès lundi matin et P. Dubois fera des tests entre Noël et Nouvel An).

Contenu

- Pas de changement par rapport au compte-rendu du 17 novembre, sauf sur les catalogues Radio (voir ci-dessous).
- ISO : A. Salama a promis que les cross-identifications manquantes seront envoyées cette semaine.
- E. Davoust avait proposé de rentrer les QSO du SDSS. Le catalogue est paru dans le dernier AJ de décembre : lui reparler de l'inclusion de la cross-identification du catalogue. Il contactera F. Woelfel à ce sujet.
- L. Cambresy reprend les 87 sources de 2MASS dès janvier puis passera au catalogue ReadMe.

Radio

Le premier fichier pour SIMBAD a été préparé. Etude des spectres « déchets » : souvent dus à une barre d'erreur interne à une des mesures dont la signification n'est pas toujours claire. Ces problèmes sont surtout rencontrés avec le WENSS et le NVSS qui ont de petites barres d'erreur. On passe de 30000 à 94000 objets acceptables pour SIMBAD en augmentant la barre d'erreur. Un format de sortie a été préparé pour permettre une visualisation rapide et aisée des points pour lesquels aucun spectre ne peut être trouvé (déchets). Vérification par sondage (réunion avec P. Dubois et F. Ochsenbein).

- Vérifier les intervalles de fréquence afin d'éviter une situation où deux points voisins ne seraient pas indépendants en raison de leurs erreurs.
- Création d'un format de sortie SIMBAD.
- Il y a, selon les statistiques effectuées, 94413 spectres indépendants. Afin d'obtenir ce nombre de spectres, il a fallu augmenter l'erreur minimum de flux à 20% pour le spectre trouvant l'algorithme (avec, en plus, application d'un facteur 1.5 à l'erreur de flux).
- Premier essai avec raccord : un gros catalogue mal codé bloquerait ou retarderait le travail ; P. Dubois et F. Woelfel le reprendront après l'AAS. De plus aucune mauvaise cross-identification n'a été détectée, ce qui est un peu étonnant.
- Une publication donne la cross-identification WENSS et NVSS : 185000 sources correspondantes ont été trouvées. B. Vollmer en a pour le moment retrouvé la moitié, ce qui est satisfaisant : l'article donne le nombre maximum qui peut être trouvé mais pour notre validation il faut qu'il y ait un troisième point à une autre fréquence.
- Nécessité d'un métacatalogue à inclure dans Vizier mais sous quelle forme et comment?
On pourrait stocker un pointeur d'adresses mais la solution d'un métacatalogue serait plus rapide. L'objectif est de rentrer le plus de choses

possible dans SIMBAD pour février puis prévoir une réunion avec P. Dubois et F. Ochsenbein. Il faudrait faire un article sur ce sujet.

- Les essais à faire sont les suivants :
 - Pulsars : catalogue de Lorimer et al. 1995
 - northern galaxies : Israel & van der Hulst 1983
 - southern galaxies : PKS catalogue
 - HII regions : Paladini et al. 2003
 - SNRs : a catalogue of Galactic Supernova Remnants par D.A. Green
 - Orion region : A. Richards
 - 6cm-20cm : Lawrence et al. 1986 (suggéré par E. Davoust)

Il faut penser à une manière automatique de faire ces essais car ils prennent trop de temps.

- Travail prévu pour janvier :
 - Ecrire une documentation détaillée
 - Inclure les sources dans SIMBAD

Catalogues

- Test de transfert de Supercosmos (qui se fera par le réseau) : F. Ochsenbein est en contact avec N. Hambly et M. Reed. Confirmation que chaque process ftp est auto-limité. Il pourrait y avoir par ailleurs un problème dû au firewall : voir avec J. Y. Hangouet. Quand le réseau ULP sera amélioré (à partir de lundi), ce sera nettement mieux. Wget marche très bien (système de reprise) mais seulement pour les petits fichiers (en deçà de 2 Go).
- Tests de connectivité : le CDS a les résultats les moins bons par rapport aux autres Instituts du projet AVO (notamment les britanniques). La question est de savoir quand sont effectués les essais et comment sont obtenus les chiffres ? Il va falloir reprendre les tests quand l'augmentation de capacité d'OSIRIS sera installée.
Le projet AVO teste la connectivité des instituts participants pour préparer l'étape suivante : installation de systèmes avancés de traitement de données ; il faut alors amener les données d'un point à un autre pour leur faire subir ces traitements. Le projet AVO souhaite prendre contact avec GEANT sur ce thème. F. Genova prendra contact avec RENATER.
- Il faudrait faire des tests avec de gros fichiers. Un log qui donne le temps de transfert des fichiers FITS a été rajouté dans Aladin, sans coût supplémentaire. Ce système présente l'avantage de faire un bilan rapidement (il serait intéressant de l'intégrer dans les tests européens !).
- Supercosmos : il va falloir mettre le catalogue en forme dès janvier.
- Sloan : 1.1T (1200 colonnes). Achat d'un RAID IDE qui arrivera mi-janvier. Il y aura un feedback de Cambridge et Edinbourg qui devraient avoir fait l'installation avant nous. Pour information Edinbourg a installé plusieurs gros serveurs Microsoft.
- Catalogues radio : les catalogues proposés par H. Andernach via A. Richards (catalogues anciens souvent cités) sont à peu près prêts. NB : Le service catalogue du CDS est plus complet en terme de catalogues radio que le service russe de catalogues radio mis en place il y a quelques années avec l'aide de H. Andernach.

Logiciels

- Logiciel SIMBAD :
 - o Depuis début décembre, A. Oberto a commencé à travailler sur le nouveau logiciel SIMBAD. Elle s'est installée sur son PC définitif. Elle est chargée de mettre en place des mécanismes de tests généraux qui sont repris chaque fois que l'on met en place des logiciels.
 - o F. Ochsenbein, P. Dubois, M. Wenger et S. Lesteven ont passé 3 après-midi à reprendre la "Dream List" pour rajouter des fonctionnalités. Cela introduit des modifications du schéma. Le tout suit son cours.
- Clusters de PC
 - o En ce qui concerne le cluster pour catalogues, 3 machines ont été réinstallées et intégrées dans l'environnement de T. Bucher. Des tests des développements de ce dernier sont réalisés par A. Schaaff et des discussions ont lieu afin de déterminer ce qu'il y a à améliorer pour obtenir une meilleure performance. Selon F. Bonnarel et P. Fernique, l'architecture est suffisamment souple. Un document pour les utilisateurs est en cours de rédaction.
 - o Aladin : la version pour la démonstration de janvier fait des appels au dispatcher pour les images 2MASS. Le faire savoir au moment de la démonstration. Exercice : lancer Aladin Java à distance ; pour faire des calculs particuliers, renvoyer juste l'image ; déléguer le calcul au cluster Aladin. On pourrait par exemple rajouter un bouton RGB dans le preview. Cela ne change pas le serveur principal.
 - o Lorsque l'on utilise un cluster, le gain n'est pas énorme : temps d'attente sur la base de données, etc...
- Stages : il faudra parler des sujets de stages lors de la prochaine réunion de suivi qui se tiendra en janvier.

Aladin

- Nouveauté depuis 3 heures : Supercosmos H α est en ligne. Tout fonctionne mais l'on n'a qu'une seule partie du ciel : il faut intégrer des descriptions de bases de données externes dans le serveur Aladin (au niveau du GLU).
- Le nouveau bouton Supercosmos augmente la visibilité ; l'impact est meilleur.
- La documentation que fait O. Bienaymé reprend celle de D. Egret. Il va la peaufiner et intégrera les nouveaux points mi-janvier. Les nouveaux développements de la version 2 se feront ensuite : calendrier ?...
- Mise en place du 1^{er} Tutorial terminée par C. Bot. En attente du feu vert pour le Tutorial astrométrie : B. Vollmer l'a envoyé à O. Bienaymé pour le tester. Il faudrait améliorer quelques points. Le tutorial est long et complexe, et, par exemple, on peut perdre le filtre sur lequel on a travaillé longtemps.
- RAID : problème sur le 1^{er} RAID Aladin

6 semaines après le début, impossible de savoir d'où vient le problème. La maintenance ne se déplace pas et fait faire les tests à F. Bonnarel ! C'est sans doute un problème de connexion (faux contact sur un fond de panier). Le fournisseur va donner une nouvelle version du firmware avec des logs plus complets.

Il faudrait faire une copie de sauvegarde de l'essentiel.

- Les intégrations de J. Marcout sont perturbées par les coupures mais continuent du côté Short Red DSS2. Reste couleur POSSI Blue : vieilles plaques avec qualité DSS2.
- Sloan : on n'a pas pu faire le file system sur Aladin avec le nouveau RAID, on va le faire sous Linux. Cela prendra environ 15 jours.
- 2MASS et DENIS : ils sont en stand by. Il faudrait au moins rentrer ce dont on a besoin pour la démonstration en ce qui concerne 2MASS.
- De nouvelles données GOODS ont été rentrées, avec de nouvelles données spatiales, pour la démonstration de janvier.
- IDHA/MDA : réunion de clôture du travail d'A. Oberto. Faire le point sur le software. Le logiciel est repris par les thésards : Fonctions : F. Flitti ; Interface Web : M. Petremand. Stage proposé au DEA astronomie sur les méthodes markoviennes/méthodes multiéchelles Stark/Bijaoui. C.Collet a participé à une réunion sur la spectro intégrale de champ. On devrait avoir des données hyperspectrales simulées de Guiderdoni début janvier. Travaux sur les cubes radio : à voir avec B. Vollmer.

Projet AVO

- La réunion AVO de la semaine dernière
Sujet principal : l'intégration des divers outils pour la démonstration. Il n'y aura pas d'interface avec les développements éventuels d'AstroGrid, qui se contentera de produire un catalogue pour la démonstration "live". En parallèle éventuellement le Myspace d'AstroGrid. La démonstration sera basée sur Aladin et inclura ACE et la fonction de cross-identification. Entrées/Sorties VOTable. Voir ci-dessus l'entrée des données dans le serveur pour la démonstration.

Standards

- Lors de la réunion Codata sur les archives, problème posé concernant les métadonnées (suite à un exposé d'A. Szalay et J. Grey) : avoir des fonctions disponibles pas seulement dans le monde Microsoft ! Or, toutes les discussions VOQuery vont de plus en plus dans cette direction. De plus, dès le début, le cone search avait été mis en place sans aucune discussion. Comment intervenir dans la discussion ? Il faudrait arriver à faire des propositions concrètes...
- Une certaine frustration sur le processus de décision sur les standards IVOA :
 - o SIA est le standard depuis le début et reste inchangé depuis octobre 2002.
 - o SSA ressemble au SIA mais en plus pauvre plat, peu hiérarchisé, avec encore moins de champs. Le draft n'a pas été diffusé sur la liste DAL.

- Registry, au contraire, s'est amélioré.

Le problème est particulièrement aigu pour le Data Access Layer. Quelles que soient les bonnes volontés en la matière, rien ne bouge. La réaction de R. Williams a été de dire qu'il ne fallait surtout pas y toucher. Or, les besoins sont là et une évolution sera malgré tout nécessaire.

Prévoir une téléconférence avec R. Williams.

- UCD : A. Preite-Martinez a repris tous les UCD1 dans Vizier, en regardant comment les réécrire sous UCD1+ (ou UCD2).

Le principe est que le premier mot utilisé doit porter l'essentiel du sens.

Exemple : s'il s'agit d'une erreur sur une magnitude, on met simplement le mot "erreur". Dans les UCD1 il fallait créer des nouveaux mots à chaque fois car ils étaient monoblocs.

Le vocabulaire est ainsi simplifié, l'agencement des mots plus souple. Il n'y a pas d'instance centrale pour définir toutes les combinaisons possibles.

A. Preite-Martinez a fait la traduction des UCD1 dans le nouveau système.

Discussion téléphonique avec R. Williams pour faire le point sur les UCD.

Plan défini : S. Derrière fait un document de synthèse sur ce qui est proposé. Quels sont les changements proposés par rapport aux UCD1 et les bénéfices que cela implique pour les gens utilisant déjà les UCD1 ; idées sur les matching functions qui comparent 2 UCD ; exemples d'application pratique.

Ce document sera envoyé à R. Williams et T. McGlynn et doit être très consensuel.

S'ils sont convaincus, tout le monde suivra. Le problème lors de la réunion Interopérabilité a été le manque de temps pour montrer des exemples et donner des explications.

- Registry : la situation s'est améliorée avec le passage à une implémentation pratique. S. Derrière a mis en place un accès OAI à Vizier. L'accès a été utilisé mais nous ne savons pas qui a testé. Il a eu une discussion avec R. Plante (choix du name space,...).

Pas de contact pour le moment avec AstroGrid qui a mis en place un registry.

- F. Genova propose d'intervenir sur ce problème de la « prise de décision » à la prochaine réunion IVOA. La preuve par la pratique ne suffit pas – voir l'absence d'évolution du SIAP, les discussions menées par le groupe Data Model, non conformes à ce qui avait été décidé en réunion,... Il est très difficile d'intervenir de façon efficace, et ce n'est pas seulement dû à l'obstacle de la langue.