

Présents: M. Allen, L. Arbousse, T. Boch, F. Bonnarel, L. Cambresy, S. Derriere, P. Dubois, F. Genova, S. Lesteven, C. Loup, F. Ochsenbein, A. Schaaff, M. Wenger

Excusés: P. Fernique, B. Vollmer

Informations générales

- Le CDS est cité dans environ 10% des articles publiés dans les quatre grands journaux, ce qui est un excellent score ; il faut faire en sorte de maintenir ce rôle pour l'avenir.
- Visite de W. Osborn: identification et cross-id de 3867 étoiles d'amas globulaires ; nombreuses corrections à faire dans SIMBAD.
- Visite de A. Sarkissian et A. Mickaelian les 12-13 et 14 juin. Ils souhaitent mettre dans le VO les données GPO Byurakan.
- Présentation des outils de Pleinpot, Saada et Sitools le 1^{er} juin à l'IAS.
- Réunion Debriefing Victoria le 6 juin à 10h.
- Réunion de présentation du CDS aux nouveaux arrivants: le 13 juin à 14h15.
- Séminaire Simbad4 le 16 juin.
- Journées de la SF2A 2006: du 26 au 30 juin à Paris.
- Réunion DS meeting de VO-TECH (bilan Cycle 3, préparation Cycle 4): du 4 au 7 septembre à Strasbourg.
- Réunion Interopérabilité de l'IVOA du 18 au 22 septembre à Moscou.
- ADASS: du 15 au 18 octobre à Tucson. Nécessité de définir avant l'été ce qui va être présenté à ADASS.
- LISA: exposé sur le travail d'entrée des données dans les services ; posters ontologie, Simbad4.

Bibliographie

- Action cross-id: cross-identifier SIMBAD et 2MASS ; E. Davoust est volontaire pour cela. Il viendra 15 jours en automne pour travailler avec L. Cambresy. Cette action est importante pour préparer l'entrée des magnitudes JHK dans Simbad4.
Dans 2MASS, il y a des sources ponctuelles et des sources étendues qui, elles, devraient être seulement dans VizieR. Il y a 1,5 millions de sources.... Pour le moment, il faut commencer par les sources ponctuelles les plus brillantes.
La correspondance avec l'USNO a été faite: on pourra ajouter aussi les magnitudes optiques. Facteur 2 par rapport au nombre d'objets actuels.
- Le tri des =g= à entrer dans SIMBAD: il faut avoir une discussion sur la stratégie à adopter, entre les chercheurs.
- Interaction VizieR/SIMBAD: il faut prévoir de travailler sur ce point, ce qui permet de limiter ce qui sera mis dans SIMBAD.
- Reconnaissance automatique des noms d'objets (VO-TECH).
C. Bonin y travaille depuis le 2 mai. Il a fait un état de l'art et va utiliser la biblio java *pdf box* pour passer d'un fichier pdf à un fichier texte. Il travaille sur la gestion des textes en colonnes, la gestion des fins de ligne (ligatures), les lettres grecques et

caractères spéciaux,... Dans 3 semaines, l'outil de conversion pdf/texte devrait fonctionner. Globalement, le travail avance bien.

- Bilan de l'entrée des données bibliographiques
Actuellement, nous entrons 12000 références par an.
- Le nombre de tags sur les noms d'objets augmente dans A&A et le taggage démarre dans les journaux de l'AAS. 43% des objets sont mal écrits ou pas reconnus par SIMBAD. Faut-il les reprendre tous ?
- Nos collègues de l'IMCCE souhaitent faire un Simbad du système solaire, avec les astéroïdes, les satellites des planètes et les planètes elles-mêmes. Il vaudrait mieux éviter de complexifier la tâche de A&A et donc rediriger les liens à partir du CDS vers l'IMCCE.
- San Yi est arrivée pour la durée du congé maternité de C. Brunet ; elle rentre des données dans SIMBAD (elle a rentré une dizaine de listes en mai). Refaire une réunion pour organiser la répartition du travail.
- Réunions procédure: la dernière a eu lieu en décembre 2005. Il y a eu un gros travail de préparation de C. Brunet (son travail est en ligne sur le twiki). Il reste des suggestions à garder en mémoire et à reprendre après la mise en service de Simbad4.
- Toute l'équipe est passée sur des PC Linux depuis la dernière réunion.
- Remplacement de la machine bibliographie ; serveur Sun en cours de commande ; serveur spécifique fourni par A&A.

Contenu

- Catalogues Radio
F. Woelfel est entrain de rentrer les derniers objets qui posaient problème dans SIMBAD. Il y avait 16550 objets dont la cross-identification SPECIFIND donne autre chose que la cross-identification existante dans SIMBAD (des identificateurs SIMBAD qui ne devraient pas y être selon SPECIFIND).
Pour rentrer ces objets à problème, il a été décidé de rajouter les identificateurs trouvés par SPECIFIND à l'objet SIMBAD et de mettre une note dans SIMBAD pour dire quelles sources n'ont pas été cross-identifiées par SPECIFIND + un lien sur l'objet dans le catalogue SPECIFIND dans Vizier (B. Vollmer a vérifié que cela ne pointe pas dans le vide).
Statistique globale de l'opération radio dans SIMBAD:
217000 objets ont été créés ou modifiés (environ 136000 nouveaux objets et environ 81000 objets modifiés).
Il reste à faire repasser RACCORD avec l'ensemble des sources radio de SPECIFIND afin de comparer les coordonnées SPECIFIND avec celles dans SIMBAD.
- Projet d'observation des « gigahertz peaked sources (GPS) » avec le télescope de 100m d'Effelsberg
B. Vollmer avait identifié 200 candidats GPS dans le catalogue SPECIFIND. Ces sources ont été observées quasi simultanément à 3 fréquences par le 100m d'Effelsberg en collaboration avec T. Krichbaum du MPIfR. Les données ont été réduites par une étudiante (L. Letemple) et B. Vollmer vient d'en finir la calibration. Le taux de détection est bon (70% à 3 fréquences, 95% à 2 fréquences). Il reste à mettre les spectres radio de SPECIFIND et d'Effelsberg ensemble.

Logiciel

- Simbad4: état d'avancement de Simbad au 29 Mai 2006.

Simbad4 peut être considéré comme opérationnel. Globalement, tout fonctionne. A. Oberto a corrigé un certain nombre de problèmes liés à la mise à jour (tests effectués lundi dernier). Il reste cependant quelques points à terminer avant le démarrage:

1. Utilisation des scripts et personnalisation des formats
Ceci est indispensable pour remplacer entièrement les modes ligne à ligne, mail et requêtes web par listes qui n'existeront plus (remplace telnet et mail qui ont encore des utilisateurs).
==> fait à 90%, le reste est plus de l'emballage
2. Librairie AstroPos (F. Ochsenbein) à compléter pour le calcul des coordonnées en tenant compte des mouvements propres, vitesses radiales et parallaxes
==> en cours
3. Reprendre un minimum de user's guide pour décrire les nouvelles fonctionnalités et un chapitre consacré au passage Simbad3 à Simbad4 là où il y a des différences
==> en cours à rythme lent pour l'instant
Pour les magnitudes et les flux, il faudrait prévoir une réunion pour décider d'une liste.

Discussion à envisager (F. Ochsenbein, M. Wenger et S. Derriere) sur l'extension des types d'objets, l'étendue des différents champs.

Faut-il permettre de donner le même identificateur pour plusieurs objets ? Il est possible de l'interdire pour le moment au niveau de la mise à jour: discussion sur ce point pour savoir ce qui est autorisé ou non. Il faudrait une liste des objets différents ayant actuellement un même identificateur pour la faire circuler afin que tout le monde puisse la voir.

4. Faire encore quelques tests de « SimbadStress », un programme qui impose des conditions extrêmes aux serveurs et révèle des bugs inaccessibles autrement. Raccord est aussi très utile pour cela.
==> en cours
5. Web services et autres accès par programmes
Les demandes arrivent déjà, alors autant les prévenir. Deux aspects:
 - le développement de Web services propres (en cours par Anaïs)
 - l'interfaçage de Simbad4 avec le WS sesame (name resolver)Documenter les paramètres d'Url (et tester) pour accéder directement par Url depuis une application quelconque.
6. Démarrage des machines et synchronisations
Pour l'instant, un démarrage de l'ensemble des machines (après une coupure de courant par exemple) ne rend pas l'ensemble des serveurs opérationnels. Il reste au moins un problème entre Apache et Tomcat.
Par ailleurs, un serveur qui tombe nécessite de relancer l'ensemble des applis (5 en tout) à la main. Ceci peut sans doute être amélioré et contribuerait à la facilité de maintenance. Il est nécessaire de résoudre ce problème avant le démarrage.

Gestion de la copie miroir à Harvard: la mise en œuvre du clone n'a pas encore été testée. Il est prévu de faire les modifications en rejouant les commandes de mise à jour. N.B: Simbad4 est déjà installé sur plusieurs machines. On mettrait VizieR et SIMBAD sur la même machine et on partagerait par la suite le même moteur postgres? Une discussion s'impose (éventuellement via F. Ochsenbein en marge de LISA V).

7. Formation à la création d'acronymes à prévoir pour B. Vollmer et F. Genova.
8. adsbib (liste des nouvelles références, récupérée par ADS tous les 10 jours) est à reprendre.

Quelques points de moindre urgence:

- Passage aux UCD1+

- Optimisation de certains types de requêtes (surtout longues) à revoir:

Ex: Vega, Sirius, M31: le temps d'interrogation moyen est de 2 secondes et demi. Pour l'instant, quand des requêtes sont faites, elles récupèrent toujours la totalité des informations associées à un objet: nécessité de limiter à certaines informations utiles (à décider).

Si une mise à jour a lieu en même temps, l'interrogation ne s'en trouve pas bloquée.

- Outils divers:

Création des noms de journaux: M. Wenger s'en occupe.

Création des statistiques: la sortie parfile de la nouvelle base sera à peu près identique. Il y a des soucis avec les noms d'objets, pour l'instant.

Création de bases tests: on ne peut pas le faire rapidement (3 dans une journée au plus !) or F. Woelfel en a besoin. La question reste posée ; peut-être y-a-t-il une solution dans Postgres 8.

Liste partielle des numéros qui se suivent à développer, difficile à mettre en place. Problème avec les identificateurs entre crochets, case insensitive. Mode wild card « bricolé » pour que les crochets soient pris en charge proprement et qu'on ne soit plus obligés de les taper.

- Faut-il refaire une journée de tests ? Il existe une base de tests qui est disponible en permanence, elle date d'il y a une semaine. Rediscuter du calendrier et voir avec A. Oberto ce qu'elle en pense.

- Finaliser les procédures de correction batch de bibcodes et d'identificateurs.

- L'exploitation des liens hiérarchiques et associatifs reste à voir.

Démarrage: soit fin juin, soit début septembre. Il faut garantir une surveillance pour maintenance continue durant l'été si on démarre en juin.

Catalogues/VizieR

- G. Landais est arrivé début janvier :

Son premier travail consistait à regarder comment fonctionne Posgres et comment se comporte l'interface d'interrogation (travail terminé).

Son 2^e travail : un pipeline pour alimenter un miroir Posgres (début).

La 3^e partie de son travail est la transformation de données ; retraiter les ReadMe et les cataloguer dans Posgres.

- Il n'y a pas d'urgence pour le passage Sybase/Posgres pour la base de Strasbourg, mais ADS souhaite passer en Posgres. Faire des études plus détaillées de performances afin qu'Alberto Acommasi (ADS) ait une base dès que possible. A. Acommasi avait une licence gratuite jusqu'à présent mais limitée à 5 gigas de données, ce qui est trop peu. Nécessité de commencer à discuter avec lui pour les premières installations.

On a identifié beaucoup de différences Posgres-Sybase ; il faut explorer les différences (sauf sur les commandes de base, peut-être y-a-t-il un problème même avec instruction LIKE).

Calendrier prévisionnel pour la fin de la 2^e partie? Après les vacances...

- Il pourrait y avoir des échanges techniques intéressants avec Simbad4 qui utilise aussi Posgres. Il est prévu que G. Landais fasse un topo dont la date reste à fixer (dans les 15 jours).

- Matériel : machines commandées et arrivées

8 machines calcul dont 2 « faites sur mesure » pour VizieR

2 machines VizieR

C. Pestel travaille sur les aspects système-réseau pour la mise en place du cluster de PC ; il est à mi-stage (terme: fin août) ; il a commencé des installations en salle machine et nous serons bientôt fixés sur les performances du nouveau cluster.

Il reste des options à trancher pour choisir l'installation la plus judicieuse.

A terme, il faudra remplacer le cluster actuel par les nouvelles machines. Un axe important: la facilité de maintenance. Utilisation d'ali développé par T. Bucher.

L'environnement qui sera mis en place intègre tous les aspects installation et désinstallation des machines ; comment substituer une machine à la machine principale en cas de défaillance ? L'utilisation d'une machine-relais est en cours d'étude.

Les 8 autres machines sont prévues pour faire du calcul donc pour Aladin.

- Travail de tous les jours : entrée des données par M. Brouty et P. Vannier. M. Brouty gère bien son temps partagé entre SIMBAD et VizieR. L. Cambresy vérifie les ReadMe.

Tous les retards sont en partie résorbés (en particulier sur les AJ). La situation est nettement meilleure qu'il y a 1 an.

Aladin

- Interface: release majeure en février, en particulier le bouton « All VO ».

La liste des modifications principales données par P. Fernique fin janvier est la suivante:

- o Les nouveautés visibles :

- Une installation d'Aladin des plus simples sous Windows (tout en un "Aladin.exe" - rien à installer - déjà un grand succès)

- Un petit relookage (visibilité logo des fournisseurs de données) [PF]

- Deux nouveaux boutons dans le formulaire des serveurs :

- . All VO: permet d'interroger simultanément l'ensemble des serveurs (images/catalogue/spectre) enregistrés dans le "registry VO" [PF]

- . SkyBot: permet d'afficher les astéroïdes et comètes
[coopération avec l'IMCCE de Paris + TB + PF]
- Un nouvel outil "plug-in" : VOspec (développé par l'ESAC/Madrid)
de visualisation des spectres [TB]
- Des "users preferences" (répertoire par défaut,...) [PF]
- o Les nouveautés moins visibles :
 - La manipulation complexe des FoV d'instruments (rotation souris, déplacement souris, affichage partiel) [PF]
 - Le chargement de description de FoV via un fichier XML [TB + FB]
 - Le support d'une interface Java baptisée VOApp pour le contrôle d'Aladin par une autre application java [PF]
 - Une nouvelle commande script "set" pour la modification de certaines propriétés des plans + extension de la commande "status" [PF]
 - Le support du navigateur Konqueror [PF]
 - La possibilité de calibrer astrométriquement des images JPEG [PF]
 - Correction de bug pour la recalibration des catalogues [PF + FB]
- o Les modifications internes :
 - Une gestion plus souple du GLU lors de l'utilisation d'Aladin en mode Standalone (indépendance vis-à-vis du site Strasb.) [PF]
 - L'installation simplifiée d'Aladin applet sur un serveur Web même s'il ne dispose pas du GLU (déjà mis en place à l'ESO, ESAC et STScI) [PF]
 - Un parsing XML/TSV plus homogène et plus "malin" pour la reconnaissance des coordonnées [PF]

Le FAQ a été mis à jour, relu et réorganisé en chapitres par D. Durand.

Plusieurs releases mineures depuis.

Un système de trois versions a été mis en place: exploitation, prototype (orienté VO-TECH) et beta version qui sont toutes trois accessibles sur le site.

Le prototype comporte toutes les nouveautés de la version beta ainsi que l'accès à VOspace + Plastic. Cette version nécessite des bibliothèques additionnelles.

Plastic sera bientôt intégré dans la version officielle. Depuis quelques semaines, la version beta supporte les extensions FITS et FITS binaires.

- Système suivi de bugs opérationnel et utilisé pour le moment en interne.
 - Plusieurs bugs d'astrométrie ont été corrigés. Dans la version beta, il y a un système qui permet de gérer les solutions polynomiales du WCS mais pas la dernière.
 - F. Bonnarel travaille sur le script pour entrer les plaques de Schmidt dans Aladin ++.
 - Hardware: commande de la nouvelle machine en parallèle avec celle de la machine bibliographie ; le nouveau serveur Sun aura à peu près 10 fois la capacité actuelle.
 - RAID: cette semaine, installation d'un nouveau RAID, reprise du RAID sur la nouvelle machine et installation du tampon intermédiaire indépendant des RAIDS.
- La fragilité SCSI sur le RAID n'a jamais été réglée (température et chocs mécaniques). Il faut noter aussi une panne disque en plus des 2 pannes de RAID.

VO / MDA / VO TECH

- Aspects image:
Standard: à voir la semaine prochaine à la réunion de debriefing de Victoria.

- Collaboration avec le LSIIT: départs prévus de J.-J. Claudon et M. Petremand qui passe sa thèse à l'automne. Il finit de rédiger et a rencontré de nombreux soucis avec les données simulées Gallics...les LSB sont plus intéressants.
Date à trouver pour une nouvelle réunion avec W. Van Dril (la précédente a eu lieu début mars). Un nouveau stage va commencer le 19 juin pour une durée de 2 mois (Wolff) dirigé par F. Bonnarel et M. Louys. B. Vollmer prévoyait de proposer un sujet de thèse selon les résultats.
Travail de J.-J. Claudon: les outils installés peuvent-ils être utilisés par les astronomes ? Ce n'est pas encore convaincant. Il reste à faire des tests pendant les 3 semaines précédant son départ (fixé fin juin). Il y a encore des travaux en cours avec E. Slezak qui souhaite parler d'AIDA et des workflow à l'IAU.
Les 10 et 11 avril: journées MDA. Il reste à mener une réflexion sur une maintenance de type logiciel libre.
A. Bijaoui: séduit par l'infrastructure pour pérenniser les logiciels + construction de workflow. En ce qui concerne la récupération des résultats dans Aladin, il y a des soucis liés à l'ensemble des différentes étapes (outils matlab, compilation). Il est prévu que AIDA fasse partie du nouveau cluster de calcul mais il y a des problèmes de maintenance (changement version matlab,...).
Conclusion: il est difficile de réutiliser les modules de recherche développés sous Matlab dans un autre environnement. A. Bijaoui a essentiellement des modules Fortran, ce qui est différent.
La moulinette LSB doit être intégrée avant le départ de J.-J. Claudon.
Réunion de discussion CDS/LSIIT prévue le 2 juin.
- Ontologies: une réunion a eu lieu le 20 avril, suite à la réunion plénière MDA, avec l'IRIT et le LORIA.
Travail sur le Thesaurus: aucun fichier n'a été obtenu malgré les promesses faites en avril ; relance à faire.
Discussion avec le LORIA sur le plan de travail et sur des applications ; ceci est repris en partie par Alexandre Richard arrivé au mois d'avril et qui est employé sur VO-TECH.
- Logiciel d'extraction de données radio du Registry
Il y a deux logiciels développés par B. Gassmann:
 - o Un outil général permettant l'extraction des sources d'après des UCDs dans le Registry
 - o Un outil spécifique pour « nourrir » SPECIFIND permettant l'uniformisation des données extraites ; c'est un outil semi-automatique
 L'outil général est prêt ; l'outil spécifique est en phase de test, en reproduisant les résultats obtenus suite à une uniformisation à la main sur 20 catalogues.
B. Vollmer avait espéré pouvoir montrer à Prague des résultats avec 100 catalogues au lieu de 20, mais le calendrier ne le lui a pas permis. Il va donc soumettre à l'UAI deux papiers sur SPECIFIND ainsi que les deux outils pour dire que c'est « prêt à l'emploi ». Il présentera également les résultats des observations de Effelsberg.
- Pas de projet assez mur pour faire une proposition à l'appel d'offre ANR Masses de Données.