

Compte-rendu de la réunion de suivi du CDS du 6 janvier 2005

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

Excusés: M. Louys, F. Ochsenbein

Bibliographie

- **Mouvements de personnel:**
Départ de G. Thureau ; arrivée de M. Brouty (lundi 17 janvier) ; recrutement d'A. Schreyeck en CDD pour 9 mois ; G. Chassagniard est en congés jusqu'au 15 mars puis reprendra à mi-temps ; A. Beyneix partira à la retraite le 30 juin. M. -J. Wagner est à 80% depuis le 1^{er} janvier 2005.
- Réunion à Paris le 14/12/04: réorganisation de l'équipe, vérification de la liste des journaux. Le nombre de journaux dépouillés a diminué. Il y a eu un gros pic dû au projet Hipparcos puis un retard accumulé entre 2003 et 2004.
En 2004, 73 journaux ont été dépouillés dans SIMBAD (le maximum étant de 90), 19 au CDS (MNRAS...) et 54 à l'IAP (A&A, ApJ...). Beaucoup de journaux ont disparu ; des journaux de vulgarisation ont également été éliminés (ex: Orion).
Le partage du travail a été refait en fonction de ce que chaque site reçoit. Attention aux journaux non reçus à Strasbourg dans la perspective d'un rapatriement complet de l'activité à long terme.
- Réunion du 04/01/05 à Strasbourg pour organiser la répartition des différents journaux entre les intervenants locaux.
- Le 25/11/04: réunion sur le nouveau programme SIMBAD avec G. Jasiewicz. Les notes rentrées sur les objets sont maintenant visibles dans la version Web de SIMBAD. La liste des journaux dépouillés est à changer (action M. Wenger).
- Acronymes: B. Vollmer va reprendre partiellement la création d'acronymes.
- Cartes de Kohonen: applications:
 1. il faudrait mettre à jour les cartes actuelles.
 2. ISO: en attente des nouvelles d'A. Salama.
 3. Carte des références pour les objets SIMBAD qui ont beaucoup de références.
- Nécessité de garder les mots clé pour les journaux principaux.
Action à continuer à mener auprès des éditeurs de journaux (cf réponse FG à l'enquête AAS).
- Réunion A&A à Strasbourg: il reste à fixer la date
- Nous recevons les tables des matières de deux nouvelles publications sous forme électronique: PASJ et PASA.

Contacts PASJ initiés pendant la visite de FG au Japon ; contacts PASA repris après le départ de M. Storey (le journal est uniquement électronique).

Contenu

- Radio

Entrée des catalogues radio dans SIMBAD.

- Version B2: un problème de doublons a été détecté avant les vacances. C'est ennuyeux car cela représente plus de 10000 sources. E. Davoust travaille sur ce problème et va envoyer la liste des correspondances à B. Vollmer (il a presque fini le travail de cross-id). La procédure à mettre en œuvre doit être discutée.
- De nombreuses sources n'ont qu'un nom et qu'une seule référence: il faudrait les supprimer et les rentrer proprement ; cela ne concerne que 2-3 catalogues. Reprise de tout cela après l'AAS.
- Un gros travail reste à faire, qui va nécessiter de modifier raccord (il ne s'agit pas de petites mais de grosses modifications dans le programme).
- Research note écrite par B. Vollmer sur la comparaison des catalogues radio avec le NVSS pour détecter les éventuels problèmes systématiques: en cours de relecture par E. Davoust. C'est utile notamment pour les nouveaux surveys. Ex: les australiens. Contacter D. Barnes à ce sujet.
- Specfind est maintenant accessible dans VizieR.
B. Vollmer a extrait les GigaHertz Peak Sources: 250 objets au total, 100 probablement galactiques. L'échantillon connu est inférieur à 100 sources. Il faut faire des observations (Bonn, Green Bank).

- ISO

Les dernières mesures sont rentrées dans SIMBAD. Les mesures ne sont pas dans SIMBAD mais récupérées dans VizieR. Il reste un bug sur les objets qui ont beaucoup de mesures ISO (+250 mesures). Correction du bug à faire par F. Ochsenbein (problème SQL Sybase: requête Sybase). C'est le premier lien direct SIMBAD/VizieR.

- DENIS

Tous les strips ont été vérifiés: 3732 strips qui ont été validés sont à récupérer. Tous les strips ont été rescannés un à un. Ils donneront une liste d'images à ne pas rendre publique. Les observations multiples ne seront pas gérées lors de la mise en ligne du catalogue au CDS.

- VO Paris

Séminaire F. Bonnarel le 17 février à Paris: images dans le VO.

Catalogues

- Cluster: le cluster est en opération mais il reste encore des réglages à faire. Il a tourné sans problème pendant les fêtes. Il y a 9 machines en tout + une machine de secours. Idée: faire un pool commun avec Aladin.
Tous les calculs de recalibration d'Aladin pourront passer sur le cluster. Certaines machines peuvent fonctionner pour les calculs mais pas forcément pour les catalogues.

Monitoring de l'état du système: faut-il un stage pour mettre en place un outil ? Projet tutoré de test des outils existants à mettre en place. Le découpage des requêtes n'aboutit pas forcément à un gain de temps.

- Très grand succès du cdsclient dont l'utilisation augmente fortement.
- Le miroir de VizieR au Canada est en train d'être remplacé.
- UCD1+: ils ne sont pas encore acceptés formellement. FO a une version en test où il peut mettre les deux dans VizieR.
Il vaut mieux, pour l'instant, faire un ajout plutôt qu'un remplacement.
Il y a une discussion à lancer sur ce que l'on fait dans le standard VOTable pour gérer le numéro de version des UCD.
- UCD1+ : commentaires
SD a repris le document et la liste reçus après le « Request For Comments » en tenant compte de la plupart des remarques. Les commentaires sur les remarques doivent être envoyés à la liste.
Il y a 2-3 points où un vote sera nécessaire: A. Micol veut que tous les UCD se terminent par un « ; ». R. Williams ne veut plus être chairman du groupe UCD: qui prendra sa suite ?
- Soft VizieR
Interrogation par liste accélérée: pas encore inclus dans la démonstration ; point à éclaircir et à discuter avec FO avant son départ pour l'AAS.
- Depuis ce lundi 3 janvier, L. Cambresy a le droit de rendre des catalogues publics par ftp, il reste à mettre en place la procédure pour VizieR. LC en attente des nouveaux UCD peut commencer à relire (d'abord sous le contrôle de FO) et va mettre en route la procédure avec P. Vannier. Il faut accélérer l'entrée dans VizieR des nouveaux catalogues.
- Côté utilisation, il y a 25% de plus par rapport au début de l'année 2004. L'activité du site a augmenté très sensiblement. 2MASS est le catalogue le plus interrogé.
- VizieR peut être interrogé directement par Web Services. Pour le moment le WS est limité, en projet accès WS complet. Il y a des gens intéressés dont M. Hill, A. Alan, P. Osuna.
Nécessité d'établir un calendrier prévisionnel dès que possible.

Logiciel

- Simbo donne accès à une liste d'informations très restreintes sur les objets. Mettre un lien direct sur SIMBAD.
- Logiciel SIMBAD: état d'avancement de SIMBAD début janvier 2005: le rapport respecte le découpage suivant:
 1. Le serveur de données, c'est à dire le schéma et le serveur de la base de données
 2. Le serveur SIMBAD (intermédiaire obligé entre les clients et la base)
 3. Les clients
 4. Les utilitaires divers

5. Le développement et la documentation
6. Les extensions possibles
7. Le déploiement final
8. Les réunions à prévoir
9. Un planning optimiste

1- Le serveur de données

Le schéma de la base est dans une version complète (suffisante), mais pas nécessairement définitive. Il va cependant être sage de reporter à une version ultérieure tout ce qui ne serait pas indispensable au début.

Les derniers rajouts, opérationnels, concernent les mesures, les notes, ainsi que la possibilité d'avoir plusieurs types d'objets pour un objet astronomique.

Des bases de tests existent avec l'ensemble des données. Un chargement complet est en cours, avec quelques bugs à résoudre dus à la variété des données.

2- Le serveur SIMBAD

La version actuelle peut être considérée comme opérationnelle. Les développements restant concernent la complétude des fonctions et d'éventuelles évolutions.

Classes de données astro: globalement terminées.

Identificateurs:

Les descriptions sont complétées environ tous les mois pour éviter d'avoir trop à rattraper. Il n'y a pratiquement plus d'erreurs.

Le préprocesseur "sesame-simbad" est implanté et les règles en cours de révision. Les deux points restant à faire concernent la définition binaire des identificateurs (descriptions à compléter) et l'interface graphique d'aide à la création d'acronymes ("acronymBuilder"). Il reste de nombreux exemples de codages de noms d'objets qui ne marchent pas. A vérifier avec F. Genova et S. Borde.

NB: le processus de vérification des noms d'objets dans les textes s'appelle aussi Sesame. Réfléchir à une homogénéisation.

Mesures:

Les descriptions sont terminées et opérationnelles. Contrairement à Simbad3, le mécanisme de « normalisation » des mesures est indépendant de celui des identificateurs.

API SIMBAD:

Elle est complétée régulièrement par de nouvelles fonctions d'interrogation et de mise à jour.

Prochaines évolutions: les requêtes par mesures, et les échantillonnages.

Cette partie de l'API nécessitera un « mini-langage » de requêtes en cours de réflexion. Ceci n'a pas progressé depuis octobre.

3- Les clients

Les clients suivants sont en chantier:

Client Java de mise à jour

A. Oberto continue son développement en liaison avec les personnes concernées. Les discussions avec G. Jasniewicz fin novembre ont été très utiles dans ce contexte.

Client WEB d'interrogation

La reprise du développement fait par J. Deprez au printemps 2004 est à l'ordre du jour des semaines à venir.

Client Java d'interrogation

Toujours en sommeil pour l'instant. Il ne sera pas opérationnel au démarrage de SIMBAD4.

Client batch et langage de script

Rien de vraiment neuf.

Package simcli en langage C.

La partie gestion des protocoles est terminée. Il reste à développer la transmission des données dans le format attendu par simcli, donc en reproduisant le format tel que SIMBAD3 le fournit.

Web services

Rien de neuf. Il va être nécessaire de définir les Web Services souhaitables dès le début puis de les développer...

4- Les utilitaires divers

raccord

Dans un premier temps, raccord continuera à s'appuyer sur simcli branché sur simbad4 avec éventuellement quelques ajustements sur les programmes de mise à jour.

Dans un deuxième temps, un raccord2 sera développé. Certaines discussions ont commencé et ont permis de dégager des idées d'évolution.

log

La production de log est opérationnelle et fonctionne. Elle doit encore être étendue à l'ensemble des informations qu'il est souhaitable de logger.

Un programme moniteur avec interface graphique est opérationnel.

Tests de données

Divers petits programmes (graphiques) permettant de tester tous types de données ont déjà été développés : Identificateurs, bibcodes, types spectral et type morphologique. D'autres suivront au fur et à mesure des besoins.

Toutes ces petites applications seront regroupées dans une application "boîte à outils" disponible en local.

Divers contrôles de cohérence de données dans la base sont en phase de réflexion.

Contrôles

Les différents serveurs pourront être contrôlés par des moniteurs. Le moniteur de logs est déjà opérationnel.

Un moniteur général permettant la visualisation de l'ensemble des applications est en cours de réflexion.

Statistiques

Des requêtes SQL dédiées à la production de statistiques sont en cours de développement, ainsi qu'une application de collecte quotidienne de statistiques.

5- Développement et documentation

La gestion des sources par CVS s'avère très agréable à utiliser dès que plusieurs développeurs participent au projet.

La documentation non javadoc (architecture, procédures, etc ...) a été réorganisée en document maître et documents secondaires, facilitant leur gestion.

6- Les extensions

Rien de neuf.

Il sera utile d'associer à SIMBAD4 le dictionnaire de nomenclature, soit sous forme d'interface d'accès à ses données, soit de façon plus complète dans la base de données elle-même.

Quoi faire de la bibliographie CDS, au delà de l'information strictement présente dans SIMBAD ?

7- Le déploiement final

Le premier PC pour SIMBAD est arrivé en novembre. Avec un deuxième PC bipro disponible à l'observatoire, il a été possible de faire des tests poussés dans la configuration suivante:

1. PC base de données: bipro/3Ghz/4Go/80Go raid miroir

Exécute postgresQL avec une base de données complète comme seule appli.

2. PC serveur SIMBAD: bipro/3Ghz/2Go/disque IDE

Exécute le serveur SIMBAD (appli JAVA)

3. PC client: monopro/2Ghz/1.5Go

Exécute une application cliente qui envoie des salves de requêtes en « n » threads en parallèle.

Les tests ont montrés que:

1. Postgres génère un process pour chaque connection. Celui-ci occupe environ 11 Mo.

4 Go représente une bonne taille mémoire permettant plus de 300 connections simultanées. La machine a été loin de la saturation, même dans les tests extrêmes.

2. JAVA fonctionne très bien en environnement multiprocesseur. Sans aucune configuration particulière, cet environnement est complètement exploité par le multi-threading. Le serveur fonctionne très bien dans 2 Go de mémoire. La demande CPU est évidemment forte.

3. la communication entre les machines (100Mb/s) n'est pas un handicap: Mettre le serveur base de données ET le serveur SIMBAD sur la même machine améliore d'environ 15% les performances pour des requêtes uniques. Mais dès que l'on exécute des requêtes simultanées, les performances se dégradent. Elles restent cependant acceptables pour un fonctionnement dégradé en cas de panne d'un serveur. La connexion gigabit entre les machines améliorera un peu les performances.

Quelques résultats sur 3 types de requêtes:

ID : requête d'un objet en retournant l'ensemble de ses données (lecture de 5 tables)

30m: requête dans un cercle de 30m (env. 70 objets)

1d : requête dans un cercle de 1d (env. 200 objets/requête)

En considérant le point de vue du serveur (nombre de requêtes qu'il peut traiter/sec) et le point de vue de l'utilisateur (temps de réponse)

threads	ID		30m		1d	
	serveur req/sec	util. sec	serveur req/sec	util. sec	serveur req/sec	util. sec
1	17	0.06	1.4	0.7	0.5	1.8
2			2.5	1.53	0.5	3.9
5			2.6	1.78	0.8	5.86
10	90	0.09	3.1	2.98	0.7	11.9
20	70	0.26	3.1	5.53	0.8	21.8
30	121	0.21				
60	160	0.28				

Ce test était sans les mesures et les notes.

Essayer de faire les mêmes tests avec le système actuel.

Autre point positif: sur les quelques centaines de milliers de requêtes faites, il n'y a eu aucun plantage de logiciel, sauf dans la gestion du pool de connections lorsqu'on approchait des limites avec 60 threads de requêtes simultanées.

8- Réunions à prévoir

Restent à faire:

1. bibcodes & bibliographie
2. démo/discussion générale avec l'ensemble des acteurs de la mise à jour.

9- Calendrier toujours optimiste

- Fait en nov/déc:

intégration des mesures dans la base + les notes
installation de la configuration de test sur deux machines
tests intensifs, surtout tests de charge pour valider les options hardware
compléter les fonctions du serveur.

- jan/fév:

installation de l'interface web. Compléter pour la rendre opérationnelle.
tests sur l'interface de mise à jour
mise à jour opérationnelle, y compris la fonction batch.

- mars/avril

implantations de web services, sortie VOTable actualisée (UCD 1+)
Raccord et contrôles quotidiens de données

- mai/juin

tests intensifs, feedback des testeurs, corrections de bugs, puis
basculement...

- Sortie SIMBAD en VOTable: le script rapide développé par P. Fernique a permis l'inclusion de SIMBAD dans Datascope.

Aladin

- Serveur image
Une réunion a eu lieu le 30 novembre dernier pour discuter des problèmes de contenu. Le compte-rendu est à finaliser.
 - Intégration des plaques de Schmidt dans le nouveau serveur – prioritaire
 - Nécessité de rentrer le SDSS 3 très rapidement: achat d'un système RAID supplémentaire.
 - Données IRAS-IRIS: il y aura une version test ce soir.
 - Compléter les données MSX
 - Données INT pour XMMLes mosaïques 2MASS devraient être traitées à terme.
En décembre, reprise de la classe astrométrie (pas mal de problèmes pour le ciel complet). La première version est déjà disponible dans le prototype AVO (depuis avant les vacances). La deuxième version reste à venir et prend en compte les solutions polynomiales des plaques de Schmidt – prérequis pour la conversion du système de gestion.
- Aladin Java
 - La version beta qui intègre à peu près tout ce qui a été demandé est en test. Pré-lancement: article et présentation de F. Ochsenbein à ADASS. P. Fernique va fournir une démonstration de bout en bout avec 3 niveaux. Aladin V3 est la

base du prototype AVO montré fin janvier à Madrid. Problème: il faut que la version soit suffisamment stable d'ici là.

- La mise en œuvre de cette version très différente a repoussé en particulier l'intégration des données système solaire. P. Fernique s'y mettra après la démonstration de janvier.
- Intégration des résultats du stage de R. Wetzl (décomposition d'images par calcul distant). Le CADC a fait des cartes de champs tout le mois de décembre, le service a tenu la charge.
- Liste des nouveaux développements dans la version 3: l'interface est plus difficile d'accès ; beaucoup d'idées nouvelles vont venir ; cubes en tant que séquences animées ; fonctions « 2 en 1 ». Il va falloir refaire les tutoriels de base avec la nouvelle version.

Liste des nouveautés introduites :

- Rééchantillonnage d'une image par rapport à la solution astrométrique d'une autre image (2 méthodes plus proche pixel + bilinéaire) (PF)
- Zoom synchronisé sur toutes les vues sélectionnées (PF)
- Possibilité de "locker" des vues (PF)
- Nouvelles commandes scripts gérant les nouvelles fonctionnalités (PF+TB)
- Help interactif pour les commandes scripts (PF)
- Nouvelle table de fausses couleurs (C. Pichon+R. Wetzl+P. Fernique)
- Impression de mosaïques pour le multivue (PF)
- Aide à la détermination de la largeur à mi-hauteur sur un CutGraph (PF)
- Amélioration des calculs de calibration astrométrique (FB)
- Amélioration de la gestion mémoire pour la sélection des objets catalogues (PF)
- Correction plusieurs bugs (Scrollbar multiview, Filtres, Affichage ToolBox, RGB inversé, blocage du recalcul des projections, zoom pointé, curseur souris erroné, gestion du cache d'accès aux vrais pixels...)

Encore à faire avant lancement officiel :

- Documentation (FAQ (PF) + manuel (OB))
- Adapter la recalibration astrométrique (image ou catalogue) à la gestion multivue (PF)
- Adapter les FoV au Multivue (PF)
- Introduire les menus additionnels CDS pour l'aide à l'insertion dans SIMBAD (PF+discussion)
- Rendre plus visible les aides à la rédaction d'un filtre (menu Popup transformé en boutons) (TB)
- Ajouter SpecView dans les pluggins (PF+TB)

Documentation: P. Fernique a relancé O. Bienaimé qui est surchargé jusqu'en février. T. Boch prépare une version de base.

Prévoir une « démonstration maison »: voir s'il y a un créneau un vendredi matin. Il faudrait prévoir, en parallèle, certaines fonctionnalités pour le personnel du CDS.

- SIA: les extensions sont restées en stand by ; il y a maintenant urgence avant la réunion de Kyoto.
- Data Model: il faut, tout d'abord, pouvoir avancer sur SIA sur la caractérisation des images. A. Micol a validé le travail accompli et a eu un apport significatif depuis un an. Discussion prévue lors de l'AAS.
Responsable: M. Louys. Documents à écrire d'ici fin février.
F. Bonnarel se rend à Baltimore la semaine prochaine (A. Micol et D. Durand) pour tester SIAP extension (vendredi + samedi matin). C'est une manipe précurseur pour le CGPS et le CFHLS. Réunion CFHLS le 4 février.

- Groupe images

- J.-J. Claudon travaille sur l'intégration des programmes du groupe image.
- Projet de stage avec C. Wolf sur la classification des galaxies (sur données Slezak). E. Slezak va venir 2 semaines en mars pour préparer le stage.
- Détection des LSB. W. van Driel va contacter ses collaborateurs qui ont travaillé sur les images INT pour valider les techniques et traiter éventuellement de nouvelles images ; c'est une action entre les mains de B. Vollmer. C. Collet a un stagiaire qui va démarrer en avril (récupérer les informations concernant le stage).
- Thésitifs: M. Petremand – aspect visualisation d'images multi-bandes (avec L. Cambresy).
Visite de J. Blaizot du groupe Gallix (qui fait les simulations). Minimum de 70 plans différents jusqu'à plusieurs centaines – données simulées.
- Visite de G. Lagache sur images de type IRAS en prévision des besoins de Planck.
- A. Lançon voudrait relancer une collaboration sur de nouveaux types de données.

De façon générale, toute collaboration interdisciplinaire génère une tâche lourde ; il faut définir les créneaux où les nouvelles méthodes apportent un plus.

Peut-on organiser plus efficacement la collaboration ? Discussion prévue avec C. Collet.

- VO-TECH: réunion à Cambridge fin novembre.

Le projet devrait commencer au 1^{er} février puis phase de 2 mois pour sa mise en place. Pour ce qui est des réunions MDA: il faudrait orienter certaines réunions techniques MDA vers des thèmes VO-TECH ; parler de visualisation multidimensionnelle. Prendre contact avec J. Pety (IRAM).

- IDHA

Faire un bilan de ce qui a été fait.

- Japon (visite de F. Genova en décembre 2004).

Des développements ont été fait sur la mise en œuvre de la grille de calcul. Les japonais seraient intéressés vers une orientation de leur grille vers une grille de données.

VOQL à discuter avec VOJapon – interrogation ADQL/VOQL de Vizier. Urgent (A. Schaaff et F. Ochsenbein).

- Démonstration AVO

Réunion AVO Démonstration à Madrid du 25 au 26 janvier. Une autre réunion a été ajoutée le 17 janvier à Garching, suite au problème avec les développements AstroGrid, pour finaliser la démonstration. Téléconférence prévue le 7.

- Workflow:

Problème: l'utilisation des workflow n'est pas pertinente pour la démonstration (sert peu pour les cas scientifiques retenus).

Démonstration AstroGrid: intégration d'un plugin avec une présentation intéressante, qui permet d'avoir la documentation sur tous les paramètres d'entrée et de les modifier.

Dans la démonstration, ceci n'est pas utilisé: on part d'un catalogue et on va lancer des modèles d'évolution de galaxies ; ce n'est pas géré dans le workflow AstroGrid. Le workflow ne peut faire une opération que sur tout un catalogue, et ne peut pas faire de boucle sur les objets à l'intérieur d'un catalogue.

Interface Aladin / My Space: OK (écrit par T. Boch).

- Registry

L'accès au registry UK est délicat. M. Hill doit pouvoir transformer une requête ADQL simple en requête Vizier pour récupérer un résultat en VOTable.

- Développements CDS:

Il reste à finaliser le xmatch par liste.

- Organisation du CDS

Le travail sur l'organigramme est en cours.