

Compte-rendu de la réunion de suivi du CDS du 21 juin 2005

Présents: M. Allen, L. Arbousse, T. Boch, F. Bonnarel, C. Bot, L. Cambresy, S. Derrière, P. Dubois, P. Fernique, F. Genova, S. Lesteven, M. Louys, F. Ochsenbein, A. Preite-Martinez, A. Schaaff, M. Wenger

Excusé: B. Vollmer

Bibliographie

- Organisation de la répartition du travail d'A. Beyneix qui va partir à la retraite fin juin. Réunion le 21 mars dernier en vue de cette redistribution.
- A. Beyneix traitait 42 journaux d'où une nécessaire répartition ; 4 journaux sont restés à Paris.
(Compte-rendu sur le twiki à l'adresse suivante :
<http://cds.u-strasbg.fr/twiki/bin/view/Reunions/ReunionsBiblio2005>)
Réunion le 10 mai pour faire le point.
- A. Schreyeck a repris A&A complètement ainsi que 10 journaux dont le PASP et le PASJ.
- C. Brunet a rédigé les procédures correspondant au travail qu'elle assure. Il faut organiser une réunion de discussion (documentalistes et chercheurs) à la rentrée.
- Prochaine réunion bibliographie: en septembre (avec S. Borde).
- Problèmes avec les changements de format des données de New A (envois non conforme à la procédure mise en place).
- Visite de L. Romary, chargé de mission Information Scientifique et Technique au CNRS (auparavant au LORIA). L'idée est la mise en place d'une réelle politique d'IST au CNRS: présentation des activités du CDS (Centre thématique au sein d'un réseau international), discussion sur la politique du CNRS vis-à-vis des publications des chercheurs.
F. Laloë développe une base de données ouverte qui prend en charge le dépôt dans astroph lorsqu'on y dépose un article ; l'idée serait d'avoir au CNRS une BD éventuellement protégée consultable par les instances.
- S. Lesteven a identifié tous les noms à consonnance chinoise posant problème et les a fait vérifier par Chenzhou Cui (environ 150 noms).
- Acronymes: B. Vollmer en fait 20 par semaine. Il reste à les intégrer dans SIMBAD (action F. Genova).
- Les noms des planètes extrasolaires ont été corrigés dans SIMBAD (lettres minuscules dans l'acronyme).

- Nécessité de rediscuter les notes. S. Laloë revoit le contenu des notes. Il faut fixer des règles. Pour le moment, beaucoup d'intervenants considèrent que c'est trop compliqué. Structure hiérarchique et associations: sera pris en charge par Simbad4. Réunion spécifique sur les notes à organiser fin septembre - début octobre.
- IBVS: nettoyage (A. Holl). Beaucoup d'anciens IBVS ne sont pas dans SIMBAD.

Contenu

- ISO: réponse à faire à A. Salama pour dire que c'est presque terminé ; il ne reste plus que quelques corrections (mesures rentrées 2 fois).
- XMM: discussion avec M. Guainazzi ; il faut faire passer le raccordement sur le log XMM. Action pour A. Preite Martinez (trouver quel objet de SIMBAD on doit relier à une observation XMM puis discuter le résultat à l'ESA).
- Radio:
 - o B. Vollmer et F. Woelfel ont regardé les objets SPECFIND dont les identificateurs se trouvent dans deux blocs différents dans SIMBAD sans contradiction entre SPECFIND et SIMBAD puis ils ont décidé de les fusionner. Si la modification de raccord (groupe d'identificateurs) est prête en juillet, ils pourront aussi attaquer les objets où l'on trouve des contradictions entre SPECFIND et SIMBAD.
 - o Il y a une nouvelle colonne rajoutée au catalogue SPECFIND. En cliquant dessus, Aladin s'ouvre en Applet, charge une image radio NVSS, fait des contours, charge une image DSS et trace les sources SPECFIND et les contours NVSS sur l'image optique.
L'effet est très joli ! (merci à F. Ochsenbein et P. Fernique)
Avec le plot du spectre dans Vizier, cette image annotée permet à l'utilisateur d'avoir toutes les informations nécessaires pour décider s'il est d'accord avec la cross-identification. Il a aussi une idée sur la qualité des données.
 - o B. Vollmer a soumis une demande de temps de télescope pour observer les Gigahertz Peaked Sources (GPS) qu'il a trouvé avec SPECFIND au radiotélescope de 100m à Effelsberg (avec T. Krichbaum et A. Zensus). Ceux-ci sont très intéressés (A. Zensus est un des directeurs du MPIfR). T. Krichbaum a proposé d'inclure aussi les catalogues mm de l'IRAM 30m.
Cela est dans la ligne du projet sur les SED (voir ci-dessous).
A la rentrée, action à mener pour se mettre d'accord sur la procédure de traitement des catalogues radio qui arrivent.
 - o Projet SED (Spectral Energy Distribution):
Pour l'instant B. Vollmer n'interfère pas dans le projet. Il attend que la recherche de catalogues radio avec les UCD soit effective.
 - o Projet LSB (Low Surface Brightness Galaxies):
On détecte bien la LSB déjà trouvée par Sabatini et al. avec l'algorithme de C. Collet.
D. Balu, le stagiaire qui travaille sur le projet, progresse. Il extrait le profil radial de brillance de surface, dans lequel il faut d'abord trouver les bonnes valeurs pour le centre, l'angle de position et l'inclinaison.

Travaille maintenant sur une procédure automatique pour fitter une exponentielle au profil. L'objectif est d'arriver à une procédure automatique d'identification des LSB.

B. Vollmer compte inviter W. van Driel après les vacances pour discuter des résultats.

- DENIS

La mise en ligne de l'information au CDS a été évoquée au Conseil Scientifique de l'Observatoire de Paris. Problème de crash du disque sur lequel se trouvait l'information copiée et mise en forme à Strasbourg.

Catalogues

- Poursuite des réunions régulières une fois par semaine ou tous les 15 jours, le mardi matin, pour mettre en place la nouvelle organisation du travail d'entrée des catalogues dans le serveur catalogues et VizieR.

M. Brouty a fait l'AJ 2004 et est en train de faire le premier trimestre 2005. Les catalogues sont mis en ftp. La mise dans VizieR est encore en cours d'apprentissage par P. Vannier.

M. Brouty commencera à s'occuper des =g= lorsque la situation des =f= sera stabilisée. Début de reprise des anciens AJ.

- UCD1+: ils sont prêts.

Il y a 150000 colonnes dans VizieR actuellement (augmentation de 1000 colonnes par mois). La finalisation de la liste UCD1+ est prévue pour la fin du mois de juin. Une nouvelle liste sera proposée à la fin de la semaine. Il va falloir mettre en place la nouvelle procédure d'assignation des UCD1+ dans VizieR.

- Arrivée de nouveaux gros catalogues:

- NOMAD (+500 millions d'objets): cross-identification de grands surveys fait à l'USNO.

- SDSS DR3 : environ 200 millions d'objets.

- Matériel:

- o Cluster: l'arrivée des nouveaux grands catalogues nécessite de doubler la place disque. 4 machines sont arrivées et serviront intégralement pour Aladin. Le reste sera libéré pour VizieR. Il sera possible de rajouter quelques disques si nécessaire. Le remplacement des éléments du cluster en panne est en cours. La répartition est de plus en plus compliquée à gérer. Le programme de gestion du cluster tient la montée en charge.

- o Achat machine ftp: on espère qu'il sera fait bientôt. Il s'agit de 2 machines: serveur ftp et serveur http et twiki ; disques en raid pour assurer la fiabilité nécessaire.

- o Réunion « matériel » générale à prévoir pour la rentrée ; à faire: bilan du matériel disponible et de celui que l'on souhaite acquérir dans un avenir proche afin d'avoir un schéma clair sur le matériel et son évolution (à prévoir avec J.-Y. Hangouet).

- Installation de machines en parallèle pour reprendre les numéros IP heartbeat en cas de panne (services sécurisés, tête du cluster, Web Services et http).
- Developer's corner: page à réorganiser et à enrichir. Très bonne visibilité en français ainsi qu'en anglais dans Google alors que le terme est assez générique.
Action: fabriquer des liens sur les programmes disponibles dans les autres services et reprendre le design des pages pour augmenter l'information sans que cela devienne trop lourd.
Rubrique VO Partners: à étendre.
- GAIA: projet de bibliothèque de logiciels astronomiques orientés simulation des données GAIA. F. Ochsenbein se rendra à la réunion organisée à Genève les 7 et 8 juillet (à suivre).

Logiciel

- Simbad4: état d'avancement de Simbad au 20 juin 2005.
La présentation de l'état d'avancement du travail change de perspective, en décrivant l'état d'avancement vu d'un point de vue « utilisateur » et non plus « système »:

- 1) Accès utilisateurs
- 2) Accès par programmes
- 3) Application de mise à jour
- 4) Applications annexes
- 5) Données
- 6) Déploiement
- 7) Présentation CDS

- 1) Accès utilisateurs

Toutes les requêtes (hors accès par applications) se feront par le web:

- accès normal
- accès par listes d'objets
- accès différé pour les longues listes

Suppression des accès telnet et mail (pour l'instant ?).

En plus des types de requêtes standards, il est possible de définir des expressions (régulières ou jokers) sur les identificateurs et les bibcodes. Il est aussi possible de sélectionner des références sur des auteurs, des mots du titre, des journaux et toujours des années.

Les données à afficher sont paramétrables de manière souple dans une page web. Il sera en plus possible de définir des formats entièrement personnalisés.

Les sorties peuvent être:

- en format html
- en format ASCII dans la page web ou dans un fichier
- en format VOTable

Le site Web et les pages Web fonctionnent pour les requêtes de base.

Ce qui reste à faire:

- les requêtes par critères: création d'un langage simple à traduire en requête SQL. Ensuite on retombe dans des fonctions opérationnelles.
- reprendre et réviser la sortie VOTable (UCD1+ et sans doute quelques évolutions du logiciel).
- lien avec le service biblio
- page simbo
- lien aladin + plot en remplacement des plots actuels de la carte des objets Simbad.

2) Accès par programmes

Les possibilités suivantes sont prévues:

- Accès « direct » par requêtes SQL: pour des cas particuliers pour le moment, à usage interne.
- API Simbad: accès par programmes Java, à réserver de préférence pour un usage local. La connexion au serveur se fait par RMI (mécanisme de type Corba propre à Java).
- Simcli: opérationnel. Les applications (en C) tournant actuellement n'ont rien à changer, pas besoin de recompilation.
- Accès http: pas encore examiné, mais sans doute possible. Avec des sorties en simple ascii plus exploitables que les sorties Simbad3. Un encapsulage devrait être possible.
- Accès par web service: prévu pour les requêtes standards. A développer.

3) Application de mise à jour

L'interface graphique est actuellement en phase de mise au point. Tests et corrections de bugs. La présentation à l'équipe de mise à jour reste à faire courant septembre. Il faudrait qu'une partie au moins de l'équipe de Paris y assiste.

Mise à jour batch: à tester.

Deux autres applications graphiques sont en cours de développement:

- acrobuilder : construction des acronymes (50% achevés).
- maintenance de la table des journaux. Quasiment achevée.

L'accès à ces fichiers par un éditeur ne sera pas la procédure courante: la lecture et la réécriture après modifs se fera par le réseau, avec verrouillage du fichier pendant la mise à jour, et contrôle des données par l'interface.

Ce qui reste à faire:

- déboguer et figoler l'interface de mise à jour
- terminer acrobuilder. Il faut en particulier conserver les commentaires qui figurent dans les fichiers de codage et la liste des acronymes.
- mettre au point les procédures de correction batch de bibcodes et d'identificateurs (changement d'écriture).

4) Applications annexes

Les applications annexes suivantes sont ou seront adaptées à Simbad4:

- a) interrogation « batch »
Opérationnel. Les requêtes de base se font en tapant juste un identificateur, des coordonnées ou un bibcode. Des commandes nouvelles permettent des requêtes plus sophistiquées ou des définitions de sortie différentes (nouveautés/simbad3).
- b) raccord
A tester. Devrait marcher puisqu'il s'appuie sur simcli. Adaptation à prévoir si des commandes de mise à jour générées changent. Un « raccord 2 » qui pourrait être une application graphique sera éventuellement étudié plus tard.
- c) info
Info permet des requêtes sur le dictionnaire de nomenclature, de quelques fichiers divers : cela restera tel quel pour l'instant.
- d) Gsc4sim
Ce script interroge VizieR pour certains grands catalogues, et retourne des commandes de mise à jour qui sont copiées/collées dans la mise à jour de Simbad. La partie client a été faite en Java et permet la communication avec un serveur « aserver » quelconque. Les infos de gsc4sim seront donc récupérables et applicables directement dans l'interface de mise à jour.
- e) findXXX (findgsc, findppm, etc...)
Ces programmes accèdent VizieR depuis des scripts clients. A priori, cela ne changera pas. Le client « aclient » java développé pour gsc4sim serait utilisable s'il fallait accéder à ces scripts dans des applications.
- f) Simbo
Va devenir une commande standard de Simbad.
- g) Simbib & simref
Les fonctionnalités de simbib & simref sont réalisables par l'api standard de Simbad.
L'idée est de laisser les clients qui utilisent ce script tels quels, et de fournir un serveur (type « aserver ») qui reprend ces fonctions en utilisant le protocole aclient/aserver.
Les deux applications peuvent être regroupées dans le même serveur.
- h) Astropos
Refait sous forme d'interface graphique. Y compris les calculs de coordonnées dans un fichier avec rajout/remplacement des coordonnées existantes.
Peut être facilement rajouté : la même chose avec des paramètres en ligne de commande.
- i) Adsbib

C'est la liste des nouvelles références récupérées par ads tous les 10 jours.

Encore à examiner.

Nouvelles applications simbad4 :

Tests de données

Une application comportant un ensemble d'outils permettant de vérifier des codages de données va bientôt être mis à la disposition des personnes intéressées.

Discussion en séance:

- les fonctions suivantes sont très utilisées:
 - programme de correction d'un bibcode
 - détection des modifications, statistiques, ... par rapport à la version de la veille
- Miroirs
Ils sont maintenus via les programmes de mise à jour en batch plus les logs. La possibilité d'avoir un miroir complet sur place est de nouveau évoquée.

5) Les données

Création d'une base de données actualisée cette semaine (20-25 juin).

Mise à jour des acronymes.

Correction des derniers problèmes de décodage de données existants (type spectraux).

Production d'un rapport sur des données incorrectes dans Simbad3.

Reste à faire:

- l'exploitation des liens hiérarchiques et associatifs.
- l'inclusion dans les requêtes de différentes tables annexes (DOI, bibcodes synonymes, etc ...).

6) Déploiement

Les PC dédiés au web et au serveur simbad sont commandés et devraient arriver prochainement. L'ensemble des trois machines devrait donc être opérationnel d'ici fin juillet.

L'ensemble des applications pourra tourner sur deux quelconques des serveurs, assurant ainsi une redondance matérielle.

Etudier la possibilité d'installer une sauvegarde dans un autre bâtiment.

Simbad4 comprend plusieurs serveurs, répartis sur les 3 machines: serveur postgresql, simbad, simcli, simref/simbib, log, web/tomcat/axis.

Les moniteurs de surveillance de ces applications sont opérationnels: moniteur général de l'ensemble des serveurs, moniteur de pilotage du serveur log, moniteur de contrôle du serveur simbad (version non graphique).

Les moniteurs suivants seront développés:

moniteur du serveur postgres (appli open source disponible), moniteur graphique du serveur simbad.

Reste à faire:

- l'implantation des enregistrements de logs
- l'optimisation de certaines requêtes courantes

Proposition: installer bugzilla, qui permet à tous les utilisateurs de soumettre des bugs et des demandes de modification, et entre ces demandes dans une base de données. Aussi en cours de réflexion pour Aladin.

7) Calendrier

Base opérationnelle début septembre.

Tests intensifs (pour le moment il faut tester l'interface Web)

Basculement. Octobre ? Il faut prévoir un apprentissage préalable de la mise à jour.

Aladin

- Aladin Java

- o Le lancement officiel fin avril d'Aladin V3 s'est bien passé ; il n'y a eu que quelques peaufinages durant les semaines qui ont suivi, indolores pour les utilisateurs.

Le nouveau code source fait 105000 lignes soit 695Ko de byte code et fichiers associés à télécharger.

- o Le développement, actuellement, porte sur:
 - la reprise du code nécessaire à la recalibration des catalogues
 - l'amélioration de la console
 - la correction des bugs résiduels.

Dès que ce sera fait, il faudra organiser une réunion à la rentrée avec les bibliographes. Période de tests.

- o Développements prévus pour l'été:
 - Proposition alternative d'utilisation d'Aladin via Web Start (action VO-TECH)
L'étude Web Start est terminée. Volonté d'abandonner l'applet non signée, cela éviterait d'avoir à installer le GLU. L'utilisation de l'applet signée est en augmentation: cependant, 20% des gens utilisent encore l'applet non signée. Le chargement est un peu lent.
 - Visualisation spécifique pour les données IMCCE
 - Prototypage des cartes de champs pour SIMBAD nouvelle mouture
 - Incorporation de la nouvelle classe Astropos de F. Ochsenbein
 - Amélioration des "layouts" des formulaires d'Aladin (que le Meta-data tree puisse être "extensible")
 - Prototypage d'accès à des noeuds SkyNode (VOTech 2e tranche ?)
- o Les collaborations sont en bonne voie:
 - Les sites exploitant Aladin ont incorporé la nouvelle version sans soucis (NED, LEDAS, CADDC, IBVS et Datascope) soit immédiatement, soit courant mai. Il faut relancer L. Michel pour que XMM passe également à la nouvelle version.
 - ESO utilise désormais officiellement Aladin V3 pour les cartes de champs de ses observations
(voir http://archive.eso.org/eso/eso_archive_main.htm)
 - ESA (pour XSA) est prêt à l'utiliser. Date de lancement inconnue.
 - STScI (pour MAST) a fait des tests. A relancer pour voir ce que ça devient.

- CFH (pour QSO) est en train de développer des extensions propres à leurs besoins basées sur la version 3 d'Aladin. Sortie prévue avant l'été. Beaucoup de questions du nouveau développeur: mail de P. Martin ; quelques soucis avec le clone du QSO.
- o Suivi des bugs et commentaires: réflexion en cours sur l'utilisation de bugzilla (voir ci-dessus).
- o Quelques statistiques du mois de mai:
 - 865 installations de la nouvelle version (580 en mars)
 - Nombre de sessions: 19922
 - 7583x 38.0% signed applet
 - 6215x 31.1% applet
 - 6124x 30.7% standalone
 - Origine particulière du lancement de l'applet:
 - 5891x 29.5% Simbad
 - 2139x 10.7% VizieR
 - 765x 3.8% NED
 - 27x 0.1% ESO_Archive
 - Non encore recensés : CADC, IBVS, CFH, LEDAS, Datascope et XMM
- Serveur image
 - o Les soucis sur la rapidité du serveur image devraient être levés par le transfert des images de Schmidt vers le nouveau serveur.
 - o La partie client a été faite en décembre.
 - o Développement du logiciel à faire pour décrire les header fits puis il va falloir faire les scripts: ce sera fait courant août par F. Bonnarel.
 - o J. viendra en septembre.

OV / MDA / VO TECH

- Data Model / DAL
 - o Modèle Caractérisation:
 - Discussion avec A. Micol à Garching pour achever une unification des idées. Les aspects plus pratiques XML Schema, document,... doivent être achevés avec J. McDowell pour les présenter sous forme de working draft à Madrid. Visite de J. McDowell prévue en septembre.
 - o SIA:
 - Beaucoup de choses ont avancé. Nouvelle version (F. Bonnarel, T. Boch) pour le CGPS et le CFHTLS qui sera discutée avec D. Durand la semaine prochaine, ce qui devrait conduire à une release du CGPS. CFHTLS a avancé politiquement (accord D. Schade, travail avec L. Sinard, démo réunion CFHTLS), possibilité de convergence assez rapidement.
 - Le document SIA extension a été publié en note IVOA: ce sera discuté avec D. Tody la semaine prochaine.
 - o Visite à Marseille le 16 septembre (F. Bonnarel, B. Vollmer, voir aussi ci-dessous) pour parler de la mise en ligne de leurs données sur le VO.
- Outils VO et Web Services
 - o SAVOT: le nouveau SAVOT est sorti ; il reste encore à le faire évoluer. Ce parser est très utile parce qu'il est optimisé par rapport aux parsers générés

automatiquement: les performances sont meilleures: accent à mettre sur la gestion de binaires.

- Web Services Vizier:
La nouvelle version en cours du WS Vizier (beta testeur: M. Comparato de l'INAF Catane) a été rediscutée avec F. Ochsenbein.
- MDA ontologies / VO-TECH
 - Toulouse : un fichier corrigé a été renvoyé à N. Hernandez ; en attente du retour.
 - LORIA : contact avec Roccia qui travaille sur la recherche de termes dans les textes. Second envoi d'articles complets.
 - Rapport de stage de DEA d'A. Richard: en cours de relecture par S. Derrière (ontologie des UCD, assignation à partir des descriptions de colonnes) ; organiser une réunion avec lui à Strasbourg début juillet.
 - B. Gassmann a commencé à faire un état de l'art sur les registry IVOA. Implémentation test d'une base de données purement XML (cf Carnivore, le registry de CalTech). Il implémente des requêtes XQuery sur le contenu de la base XML. Les requêtes par UCD permettent d'extraire les tables qui contiennent un flux radio, des positions,... La réponse est suffisamment rapide, même pour des requêtes complexes. Toutes les tables identifiées par B. Vollmer sont retrouvées, plus d'autres. En cours: extraction des champs intéressant des tables précédemment localisées. Très bon exemple d'utilisation des UCD1+.
- MDA image / VO-TECH
 - Stage de D. Ballu sur les LSB: voir ci-dessus.
Thèse M. Petremand. Données simulées de Gallics: organisation d'une rencontre avec J. Blaizot à Paris. E. Slezak passe la semaine prochaine. Une réunion MDA image est prévue vendredi après-midi. Travail de M. Petremand sur l'aspect visualisation, intégration à OTIAM.
OTIAM: interface avec Aladin, a révélé des bugs dans OTIAM et MARSIAA qui ont été corrigés. Il reste un petit développement sur lequel J.-J. Claudon va travailler jusqu'à fin août (son contrat est prolongé pendant encore 6 mois).
Travail en cours avec C. Collet sur la documentation. OTIAM est, pour le moment, un prototype ; il reste beaucoup de choses à explorer dans les 6 prochains mois: montée en charge, performances (réencapsulation de Matlab)...
Workflow: stage de C. Pestel. Développement d'un outil de gestion graphique des workflow. Le travail avance bien. C. Pestel travaille avec J.-J. Claudon sur l'interface avec OTIAM.
- Extension VO-TECH: les réunions des DS et du TAP auront lieu en septembre
- Architecture du VO/VO-TECH
 - Stage de T. Hertenstein qui va travailler avec A. Schaaff et T. Boch sur AstroGrid (« Exploration du projet AstroGrid »).
 - Voir aussi OTIAM pour les aspects workflow.
Tout ce que nous avons proposé pour l'extension du cycle 1 a été retenu. Il faut travailler sur nos propositions pour le cycle 2.
Voici un petit résumé des extensions qui ont été proposées par le CDS :

DS3:

- Amélioration du prototype "GUI workflow builder" (A. Schaaff et son stagiaire C. Pestel)
- Stage d'évaluation des composants d'Astrogrid (stagiaire : T. Hertenstein, encadré par A. Schaaff et T. Boch)
- Collaboration Aladin/Astrogrid Dashboard : N. Winstanley viendra 4 jours mi-août pour travailler avec T. Boch

DS4:

- Fournir un accès Java Web Start pour Aladin (P. Fernique)

DS5:

- SED construction : premier prototype permettant de localiser les ressources pertinentes (pour fin août) (B. Gassmann, S. Derriere, T. Boch)
- Rendre Aladin compatible UCD1+ (T. Boch, P. Fernique)

DS6:

- Faire passer le document sur "characterization model" de l'état draft à l'état working draft (F. Bonnarel, M. Louys).

Demos CDS, actions OV France

- SF2A: dix minutes de démonstration en séance plénière (L. Cambresy, A. Oberto). Démonstration du multi-vue, de la recalibration, des « regions of interest ». Un transparent sur le travail avec IMCCE (visualisation des astéroïdes sur une image donnée).
- JENAM / EAS: stand démo (M. Allen, B. Vollmer, M. Wenger). Posters en cours de préparation.
- LAM Marseille le 16 septembre: présentation des services CDS en parallèle avec la discussion sur les services LAM (B. Vollmer, F. Bonnarel).
- Tutoriel OV France: il serait probablement utile d'organiser un second tutoriel débutant pour les participants à OV France (demandé lors de la réunion de démarrage OV France). Possible en janvier 2006.