

Compte-rendu de la réunion de suivi du CDS du 8 avril 2005

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

Informations générales

- Démonstration à Marseille (F. Bonnarel + un des astronomes): reprendre l'organisation fin mai.
- Ecole CNRS astrométrie: les transparents de S. Derrière sont sur le twiki CDS (exposé de 3 heures avec modem 56k).
Public divers (aussi science de la terre). Fort intérêt.

Bibliographie

- Réunion avec A&A début février
Information réciproque sur les procédures. Sévérité accrue de A&A envers les auteurs qui n'envoient pas les tables.
Remarque: 15 à 20% des liens vers les sites non CDS qui sont donnés par les auteurs ne marchent plus après quelques mois, ce qui montre bien les problèmes de maintenance long terme pour les sites qui ne sont pas sous la responsabilité d'un centre de données.
- Réunion bibliographie du 21 mars: en 2 parties
(Compte-rendu sur le twiki à l'adresse suivante :
<http://cds.u-strasbg.fr/twiki/bin/view/Reunions/ReunionsBiblio2005>)
 - o Les procédures et leur révision
 - o Tour de table:
 - problème des tables qui existent dans la littérature mais qui ne passent pas par Vizier ; peut-on les entrer dans SIMBAD sans les faire passer dans Vizier (entre 50 et 100 objets ce qui est assez long à rentrer à la main). Faut-il passer systématiquement par =e= =f= ? Faut-il faire systématiquement par raccord pour les cross-identifications (=g=) ?
Rappel: =e= est une table à récupérer en ligne ou auprès des auteurs
=f= les données ont été récupérées et une fois traitées, vérifiées et intégrées dans le serveur catalogues et Vizier, elles passent en =g=
=g= reste à intégrer les données dans SIMBAD
 - Acronymes: beaucoup de retard ; faut-il faire un tri ?
 - Critères de tri à l'entrée des tables électroniques dans SIMBAD → prévoir une réunion avec C. Brunet et F. Woelfel (pour fin avril).
- Organisation du travail

Une réunion bibliographie réduite est prévue pour le lundi 11 prochain pour la répartition du travail d'A. Beyneix (40 journaux, environ 2000 références par an, en plus du MNRAS déjà repris par P. Vonflie).

- Etudier le besoin pour un poste d'AI documentaliste ; revoir l'ordre de priorité des activités d'entrée des journaux.
- Retour de G. Chassagnard le 15 février
- Acronymes: B. Vollmer a commencé à voir la création d'acronymes avec F. Genova. Ils ont fait 2 séries de tests.
- A. Schreyek reprend progressivement les A&A (1 sur 2)
cf: <http://cds.u-strasbg.fr/twiki/bin/view/Reunions/ReunionsBiblio2005>
- Les informations bibliographiques du PASJ arrivent maintenant directement et peuvent être incluses dans le serveur bibliographie (action lancée en décembre lors de la visite de F. Genova au Japon).

Contenu

- Radio: le gros des sources radio a été ajouté dans SIMBAD.
(200000 objets ajoutés, 540000 ids, 136000 nouveaux objets, 64000 objets modifiés).
Les problèmes sont inférieurs à 10% ; F. Woelfel trie les différentes sortes de problèmes qui sont réglés progressivement. Problème le plus important: les modifications de raccord qui doit pouvoir regrouper des identificateurs en tenant compte de la fréquence des catalogues pour calculer des scores.
Réunion à prévoir pour définir la stratégie à adopter pour ajouter les catalogues radio: avec P. Dubois, F. Woelfel, E. Davoust, F. Ochsenbein.
Les 2 articles sont acceptés. Réponse du MRC lors de la mise en ligne sur astroph.
- Tables radio dans Vizier
Il faut rentrer pour chaque catalogue la fréquence et les résolutions attachées à chaque fréquence. C'est un problème plus général du VO: cet item (fréquence/longueur d'onde, résolution) pourrait être standardisé dans le ReadMe, même si l'information ne pourra être rendue disponible que pour une petite fraction des catalogues.
- Specfind:
Peut-on mettre une référence sur Specfind dans la bibliographie SIMBAD et un lien vers Vizier? Faut-il créer un identificateur ou une mesure « référence pour les cross-identifications » ? Il s'agit encore une fois d'un prototype d'une fonction générale.
- ISO
Message d'A. Salama fin février faisant remarquer qu'il manquait des mesures des observations ISO. P. Dubois les complète avec l'aide d'E. Davoust et ce travail sera repris par F. Woelfel.
- New A
Encore des problèmes ; nous n'avons plus rien depuis octobre. F. Genova va relancer K. Schwarz début mai.

- Premières tables du PASJ (voir ci-dessus), tables de Baltic Astronomy. Nous les avons reçues ; des mises à jour ne devraient également pas tarder.
- V. Avedisova devrait envoyer des descriptions de catalogues. Il est possible de demander des informations en provenance de revues russes quand celles-ci ne sont pas disponibles à Strasbourg ou à Paris.
- A. Holl (à Strasbourg du 21 au 26 février), a développé les liens de l'IBVS (International Bulletin on Variable Stars). Il va digitaliser une partie des fichiers de mesures d'étoiles variables. Est-il envisageable de faire appel à une société extérieure pour faire scanner l'ensemble de ces documents ?
- DENIS
Les 3732 strips qui ont été validés, ont été récupérés. Côté image: tous les en-tête des images correspondant à 2000000 d'images ont été récupérés et seront transmis à F. Bonnarel pour création de pointeurs. Ils apparaîtront dans l'arbre de métadonnées du serveur Aladin, et l'url appelé sera à Paris. La fiabilité du serveur est à surveiller.

Catalogues

- Mise en place de réunions régulières (L. Cambresy, M. Brouty, F. Ochsenbein, P. Vannier) le mardi matin, tous les 15 jours.
- Passage aux UCD1+
Ce sera fait pour Kyoto ; il faut que la liste soit stabilisée. Les outils d'assignation et de traduction développés par A. Preite-Martinez fonctionnent bien.
- Statistiques
Le nombre de questions posées au service Vizier a augmenté de 50-100% en 1 an ; il en est de même pour les copies miroir.
- Matériel:
 - o La nouvelle machine va être mise en service
 - o Cluster:
Dans le cadre de MDA, achat de 4 nouveaux biprocesseurs qui entraîne une réorganisation complète du cluster. Spécialisation du cluster pour Aladin et le calcul image. Le reste sera utilisé par Vizier exclusivement
Des réunions sont prévues entre A. Schaaff, J.-Y. Hangouet, T. Keller, P. Fernique et F. Ochsenbein pour la réorganisation de l'affectation du parc et la sécurisation du fonctionnement: mise en place de redondances
Une réflexion est menée pour voir s'il vaut mieux avoir un seul "mulet" plutôt que plusieurs ; tout ceci sera utilisé dans un premier temps par les Web Services, puis le service ftp.
Pour la demande CNES, il faut prévoir le changement d'Aladin (option pour le moment station SUN) d'où faire un estimatif du coût. A prévoir également une nouvelle machine pour le serveur Web du CDS.

Les crédits obtenus de l'INSU vont permettre d'acheter le matériel SIMBAD, un nouveau firewall, et de remplacer une partie des Terminaux X.

Nécessité de mettre en place des réunions hardware/systèmes (réseaux/matériels) régulières.

- Vizquery: il n'y a pas assez de publicité qui est faite, il faudrait qu'il y en ait au moins dans le FAQ
- Web Service VizieR
Il est en phase de développement. Ne pas oublier que l'on peut aussi utiliser Vizquery.
- Sesame
Donner une meilleure visibilité au service.
- Sloan: reste à faire
Le catalogue a été rapatrié mais il doit encore être mis en forme.
- Interrogation de VizieR par ADQL: en cours (A. Schaaff) et va être étoffé pour Kyoto, ce qui devrait intéresser beaucoup de monde.
Les aspects VOQL ne sont pas très avancés pour le moment. Discuter avec M. Ohishi à propos du test d'implémentation ADQL/VOQL.
- Utilisation d'Aladin pour donner les positions de tables x, y. Il faut que ces tables puissent être rendues visibles dans Aladin même si elles n'ont pas de coordonnées absolues.
- UCD1+
Il y a eu encore des modifications de la liste (faites par S. Derrière et A. Preite-Martinez) pour répondre à des commentaires, à des demandes (RFC et demandes individuelles) et améliorer l'homogénéisation. A. Preite-Martinez a mis au point un outil d'assignation qui, à partir d'une description, trouve le bon UCD. Il a également relancé le groupe de travail IVOA.
S. Derrière a développé des outils utilisant la liste de référence (définition, validation,...). Il faudra implémenter ces fonctions en Web Services (pour VO-TECH).

Logiciel

- Logiciel Simbad4: état d'avancement de SIMBAD mi-avril 2005: le rapport respecte le découpage suivant:
 1. le serveur de données (base, schéma et postgres)
 2. le serveur Simbad
 3. les clients
 4. les utilitaires divers
 5. le développement et la documentation
 6. le déploiement final
 7. les réunions à prévoir
 8. un planning

1. Le serveur de données

Quelques derniers ajustements du schéma sont en cours et seront terminés d'ici une semaine.

Une nouvelle version de la base est en cours de chargement (copie du 29 mars). En dehors de quelques erreurs provenant de Simbad3, le seul problème restant concerne des types spectraux non reconnus.

La machine Simdbs contient une base complète utilisable.

2. Le serveur Simbad

Reste à développer: 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

Statut des différents clients en développement:

Client Java de mise à jour

Le développement est terminé dans son ensemble. Des tests sur une base réelle ont commencé. Certains aspects opérationnels sont étudiés: encapsulation des communications avec le serveur dans ssh pour les utilisateurs distants, maintenance de miroirs par passages des commandes SQL journalisées.

Client WEB d'interrogation

Une première version est opérationnelle. Par rapport au stage de Julien Deprez, l'utilisation du langage JSP a été abandonné: il introduit une complexité qui est inutile dans notre contexte, en plus d'être un langage peu élégant et lisible. De plus, de nombreuses alternatives voient le jour et sa pérennité sur le long terme n'est pas assurée. Les JSP génèrent des 'servlets', la solution adoptée est d'écrire les servlets directement.

Restent à compléter: l'interrogation par échantillonnages et les sorties autres que HTML. Remarques bienvenues d'ici quinze jours (fonctionnalités (ascii, votable, utilisation d'un fichier xsp pour sorties personnalisées), ergonomie). Toutes les fonctions de SIMBAD3 (+ d'autres) existeront dans SIMBAD4.

Client batch et langage de script

Rien de neuf.

Simbad natif:

Pas d'équivalent pour le moment mais peut être développé assez facilement. Important: visualisation des commentaires de travail, nombreuses fenêtres côte à côte.

Package simcli en langage C

Le package client actuel, qui est utilisé dans de nombreuses applications, est utilisable sans aucune modification depuis Simbad4.

Pour cela un serveur particulier gère le dialogue spécifique imposé par les routines clients. L'ensemble est opérationnel et attend d'être testé avec raccord.

Quelques remarques point de vue performances:

- c'est pour l'instant le seul exemple d'une application qui peut être utilisée de manière identique sur les deux systèmes
- pour une longue connexion comportant de nombreuses requêtes, l'accès à Simbad3 est plus rapide. Cela peut encore être amélioré dans une certaine mesure en limitant les jointures dans une première requête.
- pour une utilisation avec connexion unique (la plus courante pour les services extérieurs), l'accès à simbad4 est plus rapide car on économise à chaque appel l'initialisation du serveur Simbad.

Web services

Il va être nécessaire de définir les Web Services souhaitables dès le début, puis, de les développer...Attention: Sesame est une application opérationnelle.

4. Les utilitaires divers

generation des commandes de mise à jour de SIMBAD
Développement A. Oberto.

raccord

Le raccord actuel devrait pouvoir être testé avec le serveur Simcli. L'utilisation de raccord devrait être plus lent qu'avec le serveur actuel. Or raccord fait des requêtes éventuellement massives et c'est une application importante pour la chaîne bibliographique. Il va falloir étudier le problème.

log

L'ensemble des informations à logger reste à définir et les instructions correspondantes à implanter.

Tests de données

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

Gestion des tables

Les principales tables à maintenir concernent le codage des acronymes et la gestion des journaux: deux applications graphiques sont en cours de développement pour assurer cette maintenance.

Contrôles

Un moniteur de surveillance des différents serveurs (postgres, web, simbad, simcli, log, ...) est opérationnel.

simref, lookbib, info

A traiter ... Ces applications opérationnelles utilisent la version « à plat » de SIMBAD produite quotidiennement.

5. Développement et documentation

Rien de neuf

6. Le déploiement final

Le site web est disponible provisoirement à l'adresse :

<http://makino.u-strasbg.fr/simbad/>

Il utilise makino comme serveur web et exécution des servlets, mhd comme serveur Simbad et simdbs pour la base de données postgres.

Un serveur Simcli peut être atteint sur le port habituel de simcli 1674 de la machine mhd.u-strasbg.fr. Les userid/passwd ne sont pas utilisés.

Une réflexion est en cours pour l'implantation des différents serveurs sur les trois machines prévues: en installant deux services sur chaque machine, il devient facile de redéployer les fonctions si une machine tombe en panne. Cela impliquerait entre autre de maintenir un miroir (mis à jour la nuit) sur une deuxième machine.

7. Calendrier

Il devient sans doute raisonnable d'inclure la période d'été dans la phase de test de SIMBAD4, et de tabler sur un démarrage à l'automne (septembre-octobre).

- avril/mai

Implantations de web services, sortie VOTable actualisée (UCD 1+).

Développement du moniteur graphique du serveur Simbad, indispensable pour un contrôle fin de son fonctionnement.

Développement de la partie échantillonnage.

Correction des derniers problèmes de données.

- juin/septembre

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

- Divers:

- ergonomie

Prévoir une réunion de remue-méninges dans l'amphi. Est-il envisageable/souhaitable d'augmenter les ressemblances avec VizieR ?

- Miroir

Commencer à prévenir ADS du changement du logiciel et du calendrier.

Aladin

- Aladin Java

1) Le lancement d'Aladin V3 multiview est prévu pour la semaine prochaine (11 avril 2005):

. Le manuel d'utilisation sera prêt en même temps (merci à O. Bienaymé)

. Les dernières modifications portent sur :

- la simplification de l'usage des filtres (merci à T. Boch)

- La réécriture complète des tracés de grilles

- L'introduction d'une "Console" pour le mode script

- Le débogage sur les diverses configurations (TX, Mac, Linux, Window)
- . Le lancement se fera en 3 temps comme d'habitude
 - uniquement notre site (période de validation - 2 à 3 semaines)
 - puis sur les clones d'Aladin (sites VizieR)
 - puis enfin par sur les sites partenaires (NED, CFH,...)

- 2) Une grande nouveauté : l'applet est désormais "signée" (proposition de C. Arviset) ce qui signifie qu'il n'y a plus de réelles différences fonctionnelles entre la version applet d'Aladin et la version standalone. La certification de la signature est faite actuellement par nos soins (affichage du panneau "this applet is certified by "Centre de Données astronomiques de Strasbourg"... do you accept this certification..."). Dans le cas où l'utilisateur refuse le certificat ou si son browser est trop vieux, Aladin démarre en applet "bridée" (comme cela était le cas précédemment).

L'applet signée permettra plus facilement les collaborations (voir ci-dessous) car un site Aladin ne nécessite désormais que l'installation d'un seul fichier sur le serveur HTTP concerné (ni GLU, ni scripts perl additionnels). D'autre part, cela simplifiera grandement la gestion des bases de données à accès sécurisés (ex: CFH-LS).

- 3) Collaborations :

- a) Plusieurs nouvelles collaborations en bonnes voies :

- avec l'ESA pour la consultation de leur archive XSA (contact: Guillermo San Miguel) - collaboration très productive (cf 2) !
 - avec le STScI pour MAST (contact: Randall Thompson)
- Dans les deux cas, leurs tests avec Aladin V3 sont quasiment terminés. Ils attendent la sortie de notre nouvelle version pour basculer. La visibilité d'Aladin va être certainement grandement augmentée suite à ces deux nouveaux points d'accès (directement et indirectement via Google)
- avec le CFH pour leur nouvelle campagne d'observation (WIRcam) (contact : Pierre Martin) => nécessitera un peu de développement de notre coté
 - avec l'IMCCE pour la visualisation des objets du système solaire (contact : Jérôme Berthier). Tests préliminaires concluants => nécessitera un peu de développement de notre coté et attente du "remplissage" de la base solaire de l'IMCCE (plusieurs mois de calcul)
 - avec l'équipe italienne de VOTech pour les interactions VisiVO et Aladin (pas de développement particulier pour nous)

- b) Deux nouvelles collaborations désormais opérationnelles :

- avec l'IBVS (Andras Holl) pour la visualisation des cartes
- avec le CADC pour la construction systématiques de cartes de champs (Contact: Daniel Durand) => charge sur VizieR (ils utilisent leurs propres images + catalogues VizieR)

- 4) Le développement d'Aladin utilise désormais CVS (via eclipse) système de gestion de travaux collaboratifs. Très pratique et sans mauvaise surprise jusqu'à présent.

- 5) Quelques chiffres sur Aladin V3 :
- nombre de lignes source : 105000

- nombre de classes : 268
- taille de la nouvelle applet (Aladin.jar) : 658 Ko
- Serveur image
 - Données: mise en service d'IRAS IRIS.
 - Les mosaïques ne fonctionnent pas (bug à corriger).
 - Tests mosaïques 2MASS par XMM: problème dans la région du pôle.
 - DENIS, ROSAT: à faire.
 - SDSS: à reprendre
 - Entrée des plaques de Schmidt dans le nouveau service: doit être repris. Beaucoup de bugs astrométriques ont été corrigés en décembre.

OV / MDA

- Data Model / DAL
 - Proposition d'extension DAL (avec ESA): va paraître comme Note IVOA. Ce sera discuté à Kyoto.
 - Le pré-draft du modèle Caractérisation a été envoyé au sous-groupe de travail (en attente des commentaires). Evolution vers un compromis avec B. Thomas.
 - Implémentations:
 - extension de DAL-CGPS: collaboration D. Durand qui souhaite une implémentation dès que possible
 - CFHTLS: message à D. Schade. Objectif: présentation d'une implémentation prototype au « user meeting » en mai à Paris.
 - associations WFPC2: D. Durand, A. Micol. A. Micol veut une compatibilité avec Caractérisation. Retouches du logiciel Aladin par T. Boch.
 - peut-être cubes de données IFU (I. Chilingarian).
 - visite de G. Lemson ; discussion UML, XML schema, ... Discussion sur validation
- VOTable
 - SAVOT: A. Schaaff a amélioré pour T. Boch certains aspects du parser (signalement de problèmes concernant la validité des documents VOTable traités). Il conserve tout de même l'objectif initial qui était de fournir un parser léger. Travail actuel sur la prise en charge de données binaires, pour sortir par la suite une nouvelle version (améliorations et prise en compte des aspects binaires).
 - Specview: Ivo Busko utilise SAVOT (vérification en dézipant le .jar de la dernière version annoncée le 7/04). C'est compréhensible car la taille de SAVOT est négligeable dans Specview alors que JAVOT induirait une augmentation d'environ 30% et une capacité amoindrie au niveau de la gestion des gros fichiers VOTable.

- Web Services:
 - Colloque OVF : de nombreuses personnes sont intéressées par l'implémentation de WS ; beaucoup de questions sur le sujet et des mails à venir.
 - Une nouvelle version du WS Vizier est en cours (beta testeur : Marco Comparato de l'INAF Catane).
 - WS Unit est presque terminé. Une nouvelle version du WS UCD sera sans doute disponible (voir S. Derriere et A. Preite Martinez). VO WS Basic Profile 0.2 en cours pour Kyoto.

- MDA
 - visite d'E. Slezak (2 semaines). Utilisation, réduction, segmentation pour faire de la classification de galaxies (spécification, méthodes de normalisation). Visite également d'E. Slezak et de M. Petremand à Garching pour discuter avec J. Blaizot, ce qui permettra de tester la chaîne de traitement sur des données simulées.
 - travail B. Vollmer, F. Flitti: article accepté. La présentation à OV France a suscité de l'intérêt pour plusieurs applications.
 - David Ballu, stagiaire de C. Collet, M. Louys, F. Bonnarel et B. Vollmer démarre la suite du travail sur les LSB.
 - travail J.-J. Claudon suivi par A. Schaaff. Un prototype tourne (accessible par CGI) et peut être appelé depuis Aladin.
 - Journée Images: la journée du 15 mars s'est bien déroulée ; il y avait une trentaine de personnes. Certaines personnes ont trouvé qu'il était dommage que les travaux réalisés dans le cadre de MDA n'aient pas été montrés aux extérieurs.
 - Réponse : ce n'était pas le but initial (ouverture vers les STIC et invitation de gens extérieurs à MDA) de la journée.
 - Journée Grilles: 8 intervenants (M. Ohishi, G. Rixon, C. Vuerli et G. Taffoni, F. Capello, E. Jeannot, P. Valiron, R. Keller, R. David) et 33 inscrits sans beaucoup de publicité, à 2 mois de la journée.
 - Une matinée a été consacrée aux Grilles dans le VO et une après midi orientée expérimentation et grands projets (Grid5000, etc.).
 - Le programme est donc susceptible d'être intéressant pour les personnes des communautés astro et stic.
 - Il faut prendre soin dès à présent de ne pas dépasser la limite des 40 participants (45 semblerait plus raisonnable car cela laisserait une douzaine de places pour répondre aux besoins OV France et quelques autres collègues).
 - ontologies
 - Le stagiaire A. Richard débute à Nancy. Faire converger la vision d'A. Napoli avec nos besoins (recherche de voisins faite par construction). Réunion le 16 mars.
 - Réponse à l'IRIT mais pas encore de retour ; il faut relancer une réunion au moins avec l'IRIT.

- MDA / VO-TECH
 - Ontologies des types d'objets (VO-TECH) : travail avec A. Preite Martinez pour créer une ontologie prototype.

Protégé à partir de la liste des types d'objet SIMBAD. La hiérarchie SIMBAD n'en est pas une du point de vue des ontologies.

Intérêt: liens transverses et principes génériques (ex: candidat=propriété plutôt que type d'objet séparé).

- VO-TECH Tools: cross-id: S. Derrière a travaillé sur les différentes approches possibles (qu'est-ce-qu'on entend par cross-id ?).

Evolution pour juin=implémentation d'ellipses d'erreur.

- Workflow: A. Schaaff rédige actuellement un document sur le sujet. Eric Slesak est très intéressé par la participation aux discussions Workflow. Par ailleurs lors du colloque OV France, de nombreuses personnes semblaient intéressées et certaines souhaitent participer aux discussions (C. Surace, Chevallier, Fenouillet, ...). Une réunion a eu lieu (André, François B. et Mireille) afin d'initier les discussions.

Stagiaires:

Cyril Pestel a démarré le 1er avril son stage « Etude et réalisation d'un outil de gestion visuelle de Workflow » d'une durée de 5 mois.

Paul Huchelmann et Christian Ficheira sont à mi-parcours du projet tutoré « Etude de C-JDBC »

Ils feront une présentation aux gens intéressés fin mai.

- VizieR et les spécification du WG VOQL : en cours, ADQL dans un premier temps.

Matériels : proposition de réunions régulières pour renforcer la synergie (mutualisation des idées, travaux, tests) entre les personnes du CDS ayant des responsabilités au niveau des serveurs et J.-Y. Hangouet et T. Keller.

- OV France

La réunion de démarrage (4-7 avril, Paris) a rassemblé plus de cinquante personnes. Plusieurs laboratoires s'impliquent sérieusement dans la réalisation de services pour l'OV. Bilan positif du tutoriel d'octobre 2004, mais besoin de discussions techniques (mise en place d'une liste de discussion par P. Le Sidaner à l'Observatoire de Paris) et peut-être de réunions ciblées (ex: Web Services) à proposer au prochain appel d'offre. L'école thématique (à direction des scientifiques) sera organisée à Obernai du 7 au 9 novembre.