

## Vers l'Observatoire Virtuel

Daniel Egret

*Centre de Données astronomiques de Strasbourg, Observatoire de  
Strasbourg*

**Abstract.** How will it be possible to manage, and make accessible to the end users, the increasing volumes of data produced by the major telescopes both ground-based and space-based? This is one of the challenges of the emerging *Virtual Observatory* projects. We present here the context of data management in astronomy, the specific needs of the large sky surveys and deep fields, and some of the solutions currently envisaged.

### 1. Introduction

Comment gérer et mettre à la disposition des scientifiques les très grands volumes de données résultant des observations astronomiques réalisées par les grands télescopes au sol ou dans l'espace? Cette question connaît actuellement d'importants développements qui ont fait émerger le concept d'Observatoire Virtuel : ce terme recouvre l'ensemble des infrastructures et des méthodologies nouvelles qui vont permettre d'organiser un accès efficace à ce gigantesque gisement d'observations astronomiques, en s'appuyant sur les technologies de l'information et du traitement statistique, au service de l'amélioration de notre compréhension de l'Univers.

### 2. L'ère des grands Observatoires

La production actuelle des données astronomiques est caractérisée par l'ère des grands Observatoires : une génération de nouveaux télescopes au sol avec des surfaces collectrices de 8 à 10 mètres de diamètre, des télescopes spatiaux de très grande capacité, et l'utilisation de détecteurs qui seront bientôt capables d'enregistrer des milliards de pixels simultanément. La prochaine génération des télescopes de 50 à 100 mètres de diamètres est à l'étude.

Les astronomes doivent donc se préparer à faire face à un véritable déluge de données : plusieurs téraoctets (soit mille milliards d'octets)

par an, par exemple, pour les télescopes VLT de l'Observatoire Européen Austral (ESO).

### 2.1 *Les données astronomiques*

Mais il convient d'abord d'expliciter ce que nous entendons par "données astronomiques" : il s'agit principalement des observations collectées en mesurant (le plus souvent grâce aux détecteurs placés dans le plan focal d'un télescope) la lumière reçue des astres lointains. Ce sont, par exemple, la position et la magnitude des étoiles et des galaxies, mais aussi les spectres qui nous apportent des informations sur la température et la composition chimique des astres, ainsi que sur leur vitesse radiale.

Mais les données astronomiques, ce sont aussi les images qui nous renseignent sur la morphologie des sources étendues et des nuages de matière, et qui servent de réservoir à la constitution de catalogues des sources observées.

Ce sont enfin les documents, tables et figures qui contribuent à informer la communauté des chercheurs en astronomie sur les résultats obtenus : articles de journaux, comptes-rendus de colloques, livres, documentation technique.

L'ensemble de ces données connaît une très forte progression. On a évoqué plus haut les développements technologiques qui multiplient la sensibilité des détecteurs numériques, mis au foyer de télescopes dont la taille est elle aussi allée en croissant de façon spectaculaire. La progression est aussi sensible en ce qui concerne la production des connaissances astronomiques, comme l'illustre la Figure 1 montrant l'augmentation du nombre d'articles publiés dans l'*Astrophysical Journal*, entre 1895 et 1994, juste avant que ne devienne inévitable le recours complémentaire à la publication électronique.

### 2.2 *La nécessité de l'archivage*

Face à ce déluge de données, se pose la question de l'archivage : et d'abord est-il possible, ou tout simplement raisonnable, de vouloir archiver l'ensemble des données collectées par les très grands télescopes ? La réponse à cette question a connu des évolutions récentes : dans le passé il était commun que les observateurs disposent de leurs données sous forme de clichés ou, plus récemment, de bandes magnétiques, et les emportent avec eux pour les analyser dans leur institut d'origine, sans qu'une archive en soit conservée par l'Observatoire producteur. La seule responsabilité de l'archivage était donc laissée à l'astronome observateur qui ne disposait généralement pas de l'infrastructure pour faire évoluer son mode de stockage en fonction de l'évolution des technologies.



Figure 1.: *Helmut Abt, éditeur scientifique de l’Astrophysical Journal, montre les piles correspondant aux volumes de publications annuelles pour un certain nombre d’années choisies entre 1895 et 1994.*

C’est l’arrivée dans l’ère des observations spatiales, avec des télescopes embarqués sur des sondes ou des satellites (à partir des années 1970), qui va amener graduellement un changement des approches et mettre en évidence les arguments en faveur d’un effort spécifique d’archivage des données :

- permettre une exploitation multiple des mêmes données pour valoriser de manière optimale les produits de coûteux investissements ;
- servir de point de référence pour étudier la variation des observations au cours du temps ;
- autoriser des approches complémentaires ultérieures de phénomènes astrophysiques complexes par le croisement avec des observations du même astre ou de la même région du ciel utilisant des instruments et des techniques différents.

### 2.3 Les Grands Relevés

Les réflexions actuelles sur la nécessité du développement de l’Observatoire Virtuel prennent en compte, de façon spécifique, les relevés systématiques du ciel. Un relevé (*survey*, en anglais) consiste en l’observation systématique du ciel avec un instrument donné, afin d’enregistrer un *ciel de référence* qui sera la source de nombreuses études et utilisations ultérieures.

Il s'agit d'une activité ancienne et traditionnelle de l'astronomie. Sans remonter jusqu'à Ptolémée et son *Almageste*, on peut rappeler que la réalisation de la *Carte du Ciel*, et du *Catalogue Astrographique* associé, a mobilisé des ressources considérables de l'astronomie, en Europe et dans le monde, pendant près d'un siècle de 1887 à 1962 : la coalition des efforts de dix-huit observatoires, la mesure de quatre millions de positions stellaires...

Plus récemment, l'ambition d'une cartographie photographique complète du ciel se concrétisa, dans les années 1950, par la réalisation d'atlas photographiques du ciel avec des clichés couvrant environ  $5^\circ \times 5^\circ$ , comme celui reproduit dans la Figure 2, obtenu au télescope du mont Palomar. Il faut environ 1500 clichés de cette taille pour une couverture complète du ciel.

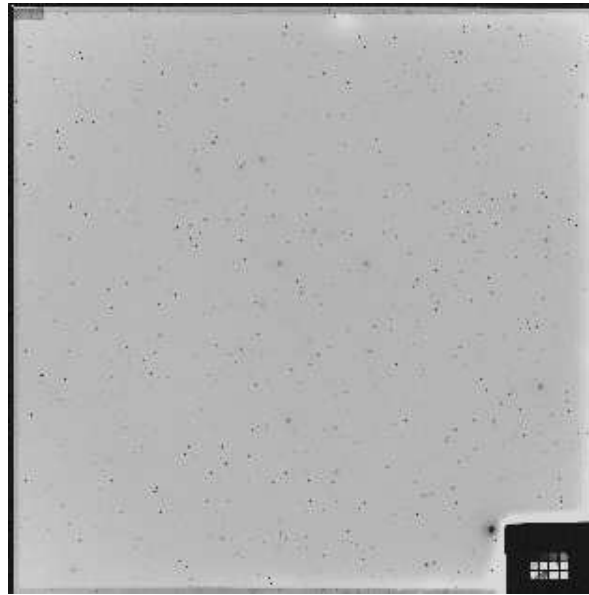


Figure 2.: *Cliché Palomar J779* : sur ce cliché on peut cataloguer environ 500 000 étoiles et galaxies.

Ce sont ces clichés sur plaque de verre, réalisés à partir des années 1950, complétés pour le ciel Sud par ceux obtenus au Télescope Anglo-Australien (SERC britannique et ESO), qui ont jusqu'à l'époque actuelle servi de "ciel de référence". Pour les besoins modernes, par exemple ceux du pointage du Télescope Spatial Hubble, il a été décidé de numériser ces clichés, et d'autres plus récemment obtenus dans d'autres couleurs, et de constituer à partir de ce travail des catalogues d'objets reconnus automatiquement et classifiés.

De nombreuses machines à numériser les clichés, aux Etats-Unis, en Grande-Bretagne, ou en France (Machine à Mesurer pour l'Astronomie, MAMA, au Centre d'Analyse des Images de l'Observatoire de Paris) ont ainsi dans les deux dernières décennies été presque exclusivement consacrées à la réalisation d'atlas numérisés. Les catalogues de sources extraites de ces relevés photographiques peuvent contenir jusqu'à 500 millions de sources, généralement identifiées comme étoiles, galaxies, ou objets étendus (nébuleuses, etc.).

Une utilisation, familière aux astronomes professionnels, de ces atlas numérisés est illustrée par l'image ci-dessous, produite à partir de l'atlas interactif du ciel ALADIN du Centre de Données de Strasbourg (voir le chapitre de F. Bonnarel, dans ce volume).

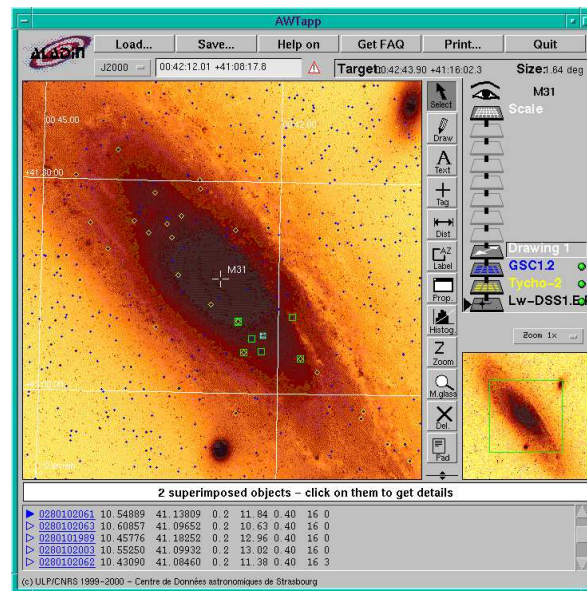


Figure 3.: Exemple d'utilisation d'une image du ciel numérisée, avec le logiciel Aladin (voir chapitre NN). La figure montre un champ de  $1.6^\circ$  centré sur la nébuleuse d'Andromède (M 31). Des symboles montrent la position des étoiles du GSC dans le champ.

Une nouvelle génération de relevés utilisant directement des caméras numériques est maintenant en train de prendre le relais : ainsi le programme Sloan prévoit de cartographier un quart du ciel en cinq couleurs et de cataloguer plusieurs centaines de millions d'objets.

#### 2.4 *Les Champs profonds*

Une approche complémentaire des grands relevés, est celle des champs profonds : il s'agit de programmes d'observation de zones restreintes du ciel, souvent réalisés avec des caméras à grand champ. La stratégie consiste à consacrer un temps d'observation suffisant pour détecter des photons émis par des objets extrêmement faibles, souvent très lointains et, donc porteurs d'une information cosmologique sur l'histoire de l'Univers. Le champ profond du Télescope Spatial Hubble est un exemple particulièrement frappant du succès de cette stratégie d'observation.

#### 2.5 *L'astronomie à plusieurs longueurs d'onde*

Une dimension importante de l'astrophysique contemporaine qu'il convient d'avoir à l'esprit lorsqu'on s'intéresse à l'Observatoire Virtuel est l'approche que l'on peut qualifier de "multispectrale" ou "panchromatique", c'est à dire prenant en compte la gamme complète du spectre de lumière. L'émission d'énergie, en provenance des astres, que nous détectons avec nos télescopes, a en effet des caractéristiques très différentes en fonction de la longueur d'onde : l'envoi de télescopes dans l'espace a ainsi permis de mesurer les rayonnements ultraviolet, infra-rouge, X ou gamma qui nous étaient masqués par le filtre de l'atmosphère terrestre.

L'observation du ciel dans plusieurs gammes de longueur d'onde permet d'accéder à une compréhension globale de la nature des phénomènes astrophysiques en œuvre dans les objets.

L'ouverture de ces nouvelles "fenêtres" sur l'Univers est aussi l'occasion de relevés complets du ciel.

L'une des principales entreprises actuelles concerne le domaine de l'infrarouge proche, qu'il est possible d'observer depuis le sol, pourvu que le télescope soit situé à une altitude suffisamment élevée. Deux relevés sont ainsi en cours d'achèvement. Le relevé Européen DENIS (voir Epchtein et al. 1999), réalisé depuis le Chili à partir d'un télescope de dimension modeste (1m de diamètre) couvre le ciel Sud. Le relevé 2MASS (Skrutskie 1998), réalisé sous la responsabilité de l'Université du Massachusetts, avec le soutien de la NASA, couvre tout le ciel à partir de deux télescopes identiques, l'un situé au Mont Hopkins, l'autre à Cerro Tololo (Chili).

Ces deux relevés constituent les premières entreprises de cartographie du ciel à partir de caméras numériques. Le projet 2MASS implique, ainsi, la collecte d'environ 24 Téraoctets, et l'identification sur les images ainsi collectées, d'environ 500 millions d'objets (étoiles et galaxies).

Les études qu'il est possible de réaliser à partir d'un tel relevé touchent un grand nombre de domaines : étude des astéroïdes de notre système solaire, détection d'étoiles de faible masse ou des fameuses naines brunes, analyse des populations stellaires dans la Voie Lactée et dans les Nuages de Magellan, observation détaillée des galaxies proches de la nôtre, observation de noyaux actifs de galaxies distantes, et du fond cosmique de l'Univers dans le proche infrarouge.

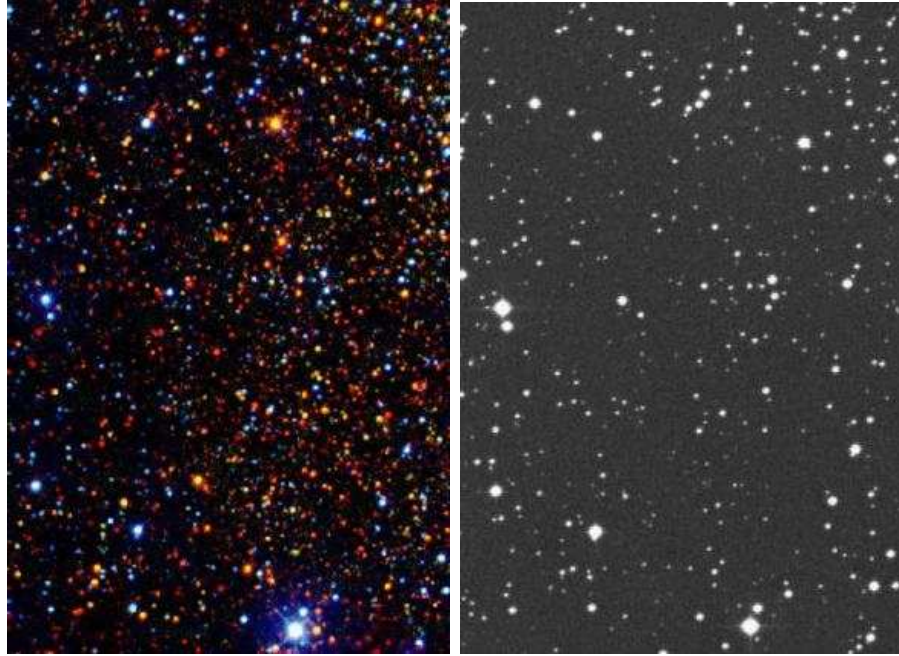


Figure 4.: Images d'une région du ciel proche du centre de notre Galaxie, à la latitude  $l = 20^\circ$ , observée dans l'infrarouge proche (DENIS) et dans le visible (cliché photographique SERC-J numérisé). (Montage photographique : S. Derrière).

Les images ci-dessus, représentent la même région du ciel, proche du centre de notre Galaxie, observée dans l'infrarouge proche avec DENIS (à gauche) et dans l'optique (cliché photographique numérisé, à droite). L'image infrarouge révèle un grand nombre d'étoiles rouges, de faible luminosité, qui n'étaient pas visibles sur le cliché photographique, à cause de l'absorption interstellaire.

De plus en plus, l'astrophysique se nourrit de ces visions complémentaires : l'approche multi-longueur d'onde peut seule nous donner une vision globale d'un objet astronomique complexe.

### 3. Archives et Bases de données

#### 3.1 *Le Centre de Données de Strasbourg (CDS)*

Les astronomes n'ont pas attendu l'arrivée de ce raz-de-marée d'informations pour imaginer et développer des solutions pour l'archivage et l'accès aux données, à travers un réseau de bases de données et de systèmes d'information.

L'une des pièces maîtresses du dispositif actuel est le Centre de Données astronomiques de Strasbourg (CDS), créé il y a près de trente ans à l'initiative de quelques astronomes clairvoyants, et en particulier de Jean Delhaye, alors responsable de l'astronomie au sein de l'Institut National de l'Astronomie et de la Géophysique. Situé à l'Observatoire astronomique de Strasbourg, le CDS abrite un ensemble de bases de données de référence utilisées par l'ensemble des astronomes professionnels au niveau mondial. La plus connue de ces bases de données est SIMBAD, qui est la base de référence pour les objets astronomiques, permettant d'identifier les astres déjà observés et mesurés, et d'obtenir immédiatement la liste de tous les articles publiés dans des journaux astronomiques professionnels mentionnant ces objets.

Les deux autres services phares du CDS sont la base de données Vizir qui rassemble les tables et catalogues d'observations, et l'atlas interactif Aladin des images du ciel (voir plus haut).

Ces services, comme ceux développés par d'autres centres, par exemple la base de données extragalactique NED, ou les centres d'archivage de la NASA aux Etats-Unis, s'appuient sur de nombreuses collaborations internationales puisque, pour être réellement complet il importe de collecter l'essentiel des informations produites par l'ensemble des télescopes et instruments astronomiques de par le monde.

Un autre exemple de service d'information, vite devenu indispensable pour les astronomes professionnels, est la base de données ADS (Astrophysics Data System) qui recense la totalité des articles publiés dans la littérature astronomique. ADS fournit les résumés et, le plus souvent aussi, le texte complet numérisé à partir de l'édition sur papier (pour les articles anciens) ou disponible en ligne sur les serveurs des éditeurs de journaux (pour les articles les plus récents). ADS recense ainsi près de deux millions d'articles. ADS est donc une bibliothèque virtuelle, qui fournit, en réponse aux requêtes de ses utilisateurs, plus de dix millions de références chaque mois. Ce total est déjà bien supérieur à ce que fournissent, à ceux qui franchissent leur porte, l'ensemble des bibliothèques *réelles* des observatoires et instituts d'astronomie.

Notons enfin qu'il existe bien d'autres centres d'information dédiés à des thématiques particulières : bases de données solaires, centre de données de physique des plasmas spatiaux, base de données des galaxies,



pour citer quelques exemples dans lesquels l'astronomie française est particulièrement active.

### 3.2 Intéropérabilité

Il nous faut maintenant introduire le terme d'interopérabilité : il s'agit là en effet d'un concept clef de l'Observatoire Virtuel.

Prenons d'abord un exemple. Imaginons un chercheur qui s'intéresse à la nébuleuse de l'Aigle, aussi appelée Messier 16 : il peut interroger la base SIMBAD du CDS et y trouver un certain nombre d'informations de référence (telles que la position sur le ciel ou les autres identificateurs) ainsi que la liste complète des articles mentionnant cet objet. Or, nous avons vu que le serveur ADS est à même de fournir le texte complet de ces articles. Si donc nous organisons les services d'information de façon qu'une référence trouvée dans la base de données des objets astronomiques (SIMBAD) puisse être automatiquement soumise au serveur d'articles (ADS) nous économiserons bien du temps au chercheur pour obtenir l'information dont il a besoin.

C'est exactement ce qui a été fait, et ceci est un exemple élémentaire d'interopérabilité : c'est à dire la capacité pour un système (base de données ou plus généralement service d'information) d'interagir avec un autre système, souvent situé à un autre endroit dans le monde, pour élaborer la réponse complète à une question complexe, pour le plus grand bénéfice de l'utilisateur. L'infrastructure actuelle commune de presque tous les serveurs d'information astronomique (accessibles par Internet à travers le Web) facilite l'interopérabilité. Mais il faut aussi un travail de fond sur la normalisation des données et des formats d'échange. Dans le cas décrit ci-dessus, par exemple, le travail est rendu facile parce que l'ensemble des partenaires des serveurs d'information astronomique se sont mis très tôt d'accord sur une codification simple et lisible des références bibliographiques.

Les obstacles à vaincre pour autoriser l'interopérabilité de plusieurs systèmes sont de différents ordres :

- techniques : le rôle primordial d'Internet et du Web est un atout important pour faciliter le transport très rapide des (petites) quantités d'information nécessaires au dialogue des serveurs ;
- commerciaux : cet obstacle ne joue guère en astronomie, où les données sont très généralement publiques et disponibles à tous après, éventuellement, une année de protection. Ce n'est de loin pas le cas dans la plupart des autres sciences (songeons aux intérêts en jeu pour l'accès aux données en chimie ou en pharmacie) ou dans le monde commercial (édition, vente par correspondance). Toutefois de nombreuses entreprises utilisent ce concept pour échanger des

- données essentielles avec des entreprises du même secteur (banques, agences de voyage, etc.) ;
- stratégiques : l'interopérabilité ne s'impose que dans un contexte de systèmes hétérogènes et distribués. Dans tous les cas où un site unique centralisé est possible, cette dernière solution pourra apparaître plus favorable. Pour l'astronomie, l'exigence de garder les données auprès des centres d'expertise qui les ont produites (observatoire, centre de dépouillement d'une mission spatiale) et qui continuent à les alimenter par des mises à jour régulières, a conduit à imposer le maintien de sites divers et complémentaires ;
  - sociologiques : faire dialoguer des systèmes d'information, et trouver un accord sur les formats, les protocoles, les standards, cela suppose des collaborations, des réunions de coordination, l'acceptation de compromis. L'existence de positions dominantes, l'inertie de certaines institutions, sont parfois des obstacles à l'interopérabilité.

#### 4. L'Observatoire Virtuel

Les premières parties de cet exposé nous ont permis de tracer les contours attendus de l'Observatoire Virtuel. Nous avons vu le contexte, celui des très grands volumes d'information produits par les télescopes modernes. Nous avons vu les enjeux scientifiques, avec l'exploitation des grands relevés qui cartographient le ciel à la recherche de vues globales de l'Univers, ou à la recherche des objets les plus rares. Nous avons évoqué la nécessité de corrélérer entre elles des vues obtenues à différentes longueurs d'onde, qui captent l'énergie émise par des phénomènes astrophysiques divers. Nous avons enfin montré que les technologies actuelles et le concept d'interopérabilité avaient ouvert le chemin vers des systèmes capables de mieux faire face à la grande diversité des informations disponibles. L'Observatoire Virtuel est la réponse organisée à ces besoins et à ces défis.

L'Observatoire Virtuel vise à donner accès aux observations du ciel collectées, numérisées et archivées, en utilisant les capacités d'interconnexion et de transmission des données, et à favoriser leur exploitation par le développement de méthodologies innovantes et l'utilisation des grilles de calcul.

Les centres de données et systèmes d'information qui existent en astronomie depuis près de trente ans étaient déjà porteurs de cette préoccupation. L'élément nouveau est l'irruption dans le paysage des grands relevés complets du ciel, observés à différentes longueurs d'onde. Le volume et la complexité des données mises en jeu impliquent un changement d'échelle dans les efforts réalisés jusque là.

Le développement, à l'échelle internationale, de l'Observatoire Virtuel (ou des Observatoires Virtuels) est encore dans une phase préliminaire, aussi bien en Europe qu'aux Etats-Unis ou en Australie. Mais, comme nous l'avons montré tout au long de cet exposé, un certain nombre de pièces du puzzle sont déjà opérationnelles. Ainsi, les archives des missions spatiales se conforment déjà aux exigences de l'Observatoire Virtuel : accessibilité, standardisation, interopérabilité. Il en est de même, quoique avec de notables exceptions, des archives des grands observatoires au sol (comme le VLT).

Une importante leçon du passé, après des tentatives infructueuses pour créer de gros systèmes centralisés, est qu'il est extrêmement important que les données résident là où se trouve l'expertise à leur sujet. Le réseau de centres de données, collectant l'ensemble des résultats et informations produits par les chercheurs, est lui aussi en place depuis plusieurs décennies, avec une tradition forte de coopération et d'échange de l'ensemble des données collectées.

Les développements logiciels concernant le traitement, la manipulation, et la visualisation des données font traditionnellement l'objet de développements beaucoup plus compétitifs, impliquant souvent l'utilisation de systèmes commercialisés, coûteux et sensibles aux brutales évolutions du marché informatique. Enfin les méthodologies d'analyse de données et d'extraction des connaissances font l'objet de recherches communes avec les statisticiens et les spécialistes du traitement de l'information.

#### *4.1 Les projets en cours*

Les premières mentions d'un projet d'Observatoire Virtuel sont apparues dans le rapport décennal (juin 2000) de l'Académie des Sciences aux Etats-Unis : le National Virtual Observatory y est décrit comme la première priorité de la décennie à venir dans la catégorie des "petites missions", avec un budget affiché de 60 millions de dollars.

En Europe, l'Agence Spatiale Européenne (ESA) et l'Observatoire Européen Austral (ESO) ont d'abord proposé une phase exploratoire avec le programme AstroVirtel qui organise un soutien pour l'accès aux archives de l'ESO et de l'ESA. Dans une deuxième phase, c'est un projet d'Observatoire Astronomique Virtuel (Astrophysical Virtual Observatory) qui vient d'être approuvé (juillet 2001) par la Commission Européenne. Ce projet ne constitue encore qu'une phase d'étude. Il permettra de réaliser des prototypes de serveurs, et de préparer le cahier des charges et l'argumentaire scientifique pour justifier le passage à une phase opérationnelle de plus grande ampleur.

#### 4.2 Observatoire Virtuel : la mise en œuvre

Les tâches prioritaires à réaliser pour mettre en œuvre l'Observatoire Virtuel ont été identifiées : un inventaire des données existantes à l'échelle internationale ; le déploiement d'un ensemble cohérent d'archives, de relevés, de services, et de dictionnaires de référence ; la mise en place de modes d'accès normalisés aux données ; la prise en charge de la dimension éducative et culturelle permise par l'ouverture large de l'accès aux données et aux images. Un autre élément essentiel de la réussite de l'Observatoire Virtuel sera la capacité de développer les outils de traitement et d'analyse, en particulier dans le domaine de l'analyse statistique, de la fouille des données et de l'extraction des connaissances. Il conviendra, par exemple de réaliser l'identification croisée des sources observées à plusieurs longueurs d'onde, au sein de listes comprenant des milliards d'observations, et localisées en différents endroits. Un autre défi à réaliser sera la découverte, au sein de ces très vastes ensembles de données, de groupes particuliers partageant les mêmes caractéristiques, la recherche d'objets rares, voire très rares, etc. Cela imposera sans doute toute une gamme de travaux préalables sur la description et la classification des données.

### 5. Conclusion

L'Observatoire Virtuel vient d'entrer dans une phase de développement avec l'objectif d'offrir un accès global et performant à la grande richesse des données et informations disponibles actuellement ou dans un avenir proche.

L'ambition des promoteurs de ce projet, dont plusieurs éléments clefs sont d'ores et déjà disponibles, est de rendre possible une nouvelle approche de l'astrophysique, et de favoriser notre compréhension de l'Univers à différentes échelles.

L'Observatoire Virtuel deviendra alors, on peut le penser, un outil omniprésent dans le travail quotidien des astronomes professionnels, qui disposeront sur leur bureau ou leur espace de travail d'une fenêtre ouverte en permanence sur l'Univers observé.

### Références

- [1] Bonnarel, F., Fernique, P., Bienaymé, O. 2000, A&AS 143, 33 (Aladin)
- [2] Brunner R., Djorgovsky S.G., Szalay A.S., eds., 2001, *Virtual Observatories of the Future*, ASP Conf. Ser. 225

- [3] Epchtein, N., Deul, E., Derriere, S., et al. 1999, *A&A* 349, 236 (DENIS catalogue)
- [4] Genova, F., Egret, D., Bienaymé, O., Bonnarel, F., Dubois, P., Fernique, P., Jasniewicz, G., Lesteven, S., Monier, R., Ochsenbein, F., & Wenger, M. 2000, *A&AS* 143, 1
- [5] Gilmore G., 2001, in “Mining the Sky”, *ESO Astrophysics Symp.* XV, p. 689
- [6] NVO Interim Steering Committee, 2001, in “Virtual Observatories of the future”, *ASP Conf. Ser.* 225, p. 353 (White Paper)
- [7] Skrutskie M., 1998, in “The impact of near-infrared surveys on galactic and extragalactic astronomy”, *Kluwer Publ.*, *ASSL* 230, 11 (2MASS)