Chapitre 1

Infrarouge et Observatoire Virtuel

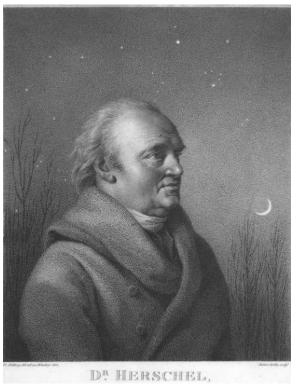
There are more things in heaven and earth, Horatio,
than are dreamt in your philosophy.
– W. Shakespeare, *Hamlet*

Au cours des dernières centaines d'années, la manière d'appréhender le cosmos s'est profondément modifiée, tant du point de vue scientifique que culturel. Les avancées technologiques ont contribué, et contribuent toujours, à changer notre façon de sonder les profondeurs cosmiques, obligeant sans cesse l'esprit humain à se forger de nouveaux concepts et à découvrir de nouveaux horizons. Les mythes créateurs et les croyances ancestrales (Reeves 1994) cèdent progressivement leur place face à ce que l'on se plaît à nommer des vérités scientifiques, certes toutes relatives. Sans cesse ébranlées par de nouvelles observations, les théories-Phœnix continuent de renaître et de s'élever sur les ruines des précédentes. Comme pour laisser une part de rêve au milieu de ce nouveau paradigme qui marque le tournant du troisième millénaire, les questions essentielles, parfois et même souvent inavouées, restent toujours sans réponse. L'astronomie est une science par essence purement observationnelle, en ce sens que son objet d'étude est inaccessible. Paradoxe suprême de la destinée humaine : l'Homme, qui fait partie intégrante de l'Univers et en est même le produit, est non seulement incapable d'en sortir mais encore incapable de pénétrer totalement les mystères de ses constituants essentiels que sont les étoiles. L'information astronomique n'est accessible que par l'analyse de la lumière (à l'exception des météorites qui parviennent jusqu'à la Terre). Mais nous allons voir combien les progrès sont rapides et dévastateurs. . .

Les Hommes ont découvert que la Terre tourne autour du Soleil, que le Soleil n'est qu'une simple étoile parmi les 200 milliards qui forment notre galaxie la Voie Lactée, et que de nombreuses galaxies peuplent l'Univers. Maintenant les fusées partent à l'assaut du ciel, on tente même quelques petits pas de danse sur la Lune et des sondes, puis des robots partent explorer de nouveaux mondes... Il y a à peine 200 ans, la Terre était encore considérée comme étant vieille de 6 000 ans seulement. Le physicien écossais James Hutton, un géologue amateur, fut le premier à déterminer son âge correct. Son étude des formations rocheuses sur Terre l'amena à conclure qu'elle était vieille de plusieurs millions d'années, faisant ainsi subitement plonger l'esprit humain dans les profondeurs du temps, jusque-là ignorées. Le géologie ne fut pas la seule science à changer radicalement notre vision du monde et à en élargir les frontières. Vers la même époque, l'astronomie allait connaître un bon prodigieux, grâce à la découverte de l'infrarouge. Riche de promesses et de réalisations, cette découverte passée est la source, encore aujourd'hui, de nombreux projets, tant à l'échelle internationale avec les grands projets observationnels tels que DENIS (*DEep Near-Infrared Survey of the Southern Sky*, Epchtein et al. 1997) et 2MASS (*Two-Micron All Sky Survey*, Skrutskie et al. 1997), qu'à l'échelle individuelle avec cette thèse. Des centaines de millions d'étoiles ont été recensées grâce aux grands relevés DENIS et 2MASS. L'acquisition et l'accumulation de ces masses de données ne sont pas limités au seul domaine infrarouge. En entrant dans l'ère du tout technologique, l'astronomie de la fin du XX^{e} siècle a connu une véritable révolution se traduisant par l'explosion du volume de données. Toujours à la recherche de nouvelles découvertes scientifiques, la volonté d'exploiter ces nouvelles mines de données et la nécessité de développer des outils innovants pour les traiter a conduit l'astronomie à un changement de paradigme qui s'incarne dans la notion d'Observatoire Virtuel.

Nous consacrerons cette introduction à une vue d'ensemble de l'astronomie infrarouge puis à l'émergence de l'Observatoire Virtuel aux abords du troisième millénaire.

1.1 Pionniers de l'infrarouge



En 1800, l'astronome et musicien anglais Sir William Herschel (ci-contre), remarqua que les différents filtres colorés au travers desquels il observait le Soleil laissaient passer différentes quantités de chaleur. Il décida alors de séparer la lumière du Soleil en ses différentes couleurs de l'arc-en-ciel à l'aide d'un prisme et de mesurer la température de chaque couleur. Herschel observa que plus il déplaçait le thermomètre vers les couleurs rouges, plus la température augmentait. Piqué de curiosité, il décida de mesurer la température du récepteur au-delà de la couleur rouge, là où il n'y avait plus de lumière visible, et quelle ne fut sa surprise lorsqu'il decouvrit que c'était précisément à cet endroit-là que la température était la plus haute ! Il en déduisit donc l'existence de rayons invisibles porteurs de chaleur : le rayonnement thermique ou chaleur calorifique. Aujourd'hui, ces rayons invisibles font partie de ce que l'on nomme le rayonnement infrarouge.

En 1847, A. H. L. Fizeau et J. B. L. Foucault montrèrent que la radiation infrarouge possède les mêmes propriétés que la lumière visible. Elle peut être réfléchie, réfractée, et capable de former des figures d'interférences. La découverte de l'infrarouge intervint 65 ans avant la théorie de James Clark Maxwell sur l'existence du spectre électromagnétique complet (Fig. 1.1).

Tous les objets émettent une radiation infrarouge et il est possible de déduire nombre d'informations sur la nature de ces objets à partir de l'analyse de ce rayonnement. Néanmoins, l'importance véritable du rayonnement infrarouge en astronomie ne fut comprise que tardivement. Bien que les premiers instruments pour détecter cette lumière in-

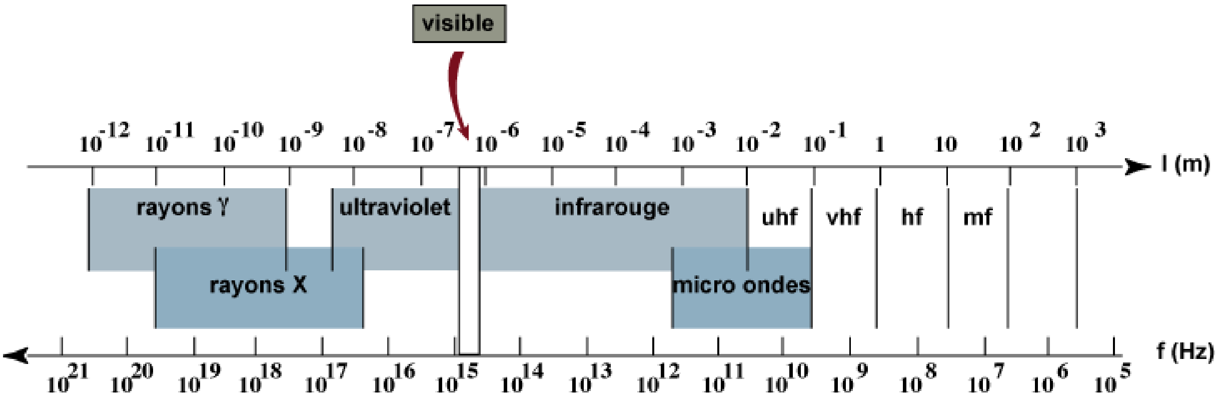


FIGURE 1.1 : *Spectre électromagnétique et principaux domaines de longueurs d'ondes.*

visible furent construits rapidement après l'expérience d'Herschel, pas même la Lune ne fut observée avant 1856 par Piazzi Smyth. En 1881, Samuel Pierpont Langley inventa le bolomètre dans le but d'étudier la radiation solaire. L'un de ses étudiants composa ces quelques vers afin de célébrer ses travaux :

Prof. Langley devised a Bolometer
It's really a sort of Thermometer
It'll detect the heat
Of a Polar Bear's feet
At a distance of Half a Kilometer !

De 1868 à 1890, Laurence Parsons, le quatrième Comte de Rosse entreprit de mesurer la chaleur émise par la Lune, l'objet le plus brillant du ciel après le Soleil. Il obtint une température de la Lune compatible dans les 8% avec sa valeur actuelle. Thomas Alva Edison inventa le tasimètre pour mesurer la température de la couronne solaire, notamment lors d'une éclipse totale du Soleil dans le Wyoming. Dans les années 1900, la radiation infrarouge fut détectée en provenance de plusieurs objets astronomiques : Jupiter, Saturne, Véga. En 1915, William Coblentz développe les détecteurs à thermopile et mesure la radiation infrarouge en provenance de 110 étoiles. Il est considéré comme le fondateur de la spectroscopie infrarouge moderne. Des observations plus systématiques furent entreprises entre autres par les astronomes américains Seth Nicholson et Edison Pettit dans les années 1920.

Pourtant, il fallut attendre plus d'un siècle et demi après l'expérience d'Herschel pour révéler aux astronomes l'impact extraordinaire que fut l'observation du ciel dans l'infrarouge.

1.2 Astronomie infrarouge moderne

1.2.1 Premiers développements technologiques

La toute première technologie infrarouge ne fut pas développée dans un but scientifique mais pour des applications militaires. Les détecteurs de lumière infrarouge peuvent "voir" même la nuit (ou dans le brouillard) les corps qui émettent de la chaleur, rendant ainsi possible la vision de nuit, qui fut la principale motivation de leur développement tout au long du vingtième siècle. Également l'équipement des têtes de missiles avec des systèmes de guidage par recherche d'émission thermique. La technologie infrarouge a ensuite été utilisée en médecine pour faire de l'imagerie thermique et détecter des zones corporelles plus ou moins chaudes. On dénombre bien d'autres applications de la thermographie : étude de la végétation, recherches archéologiques... La technologie des détecteurs infrarouges appliquée à l'astronomie a été décrite par McLean (1993) et de Graauw (1993).

En dépit d'un départ assez précoce et de quelques développements dans la technologie des détecteurs infrarouges (Edison, Golay), aucune percée véritable ne fut réalisée.

Les observations dans la partie de l'infrarouge comprise entre 700 et 2 500 nm se font pour l'essentiel comme avec la lumière visible. Un télescope optique ordinaire peut servir à focaliser la lumière et la seule modification consiste à utiliser un détecteur sensible à l'infrarouge. Gerard Kuiper a été le premier astronome à effectuer de telles observations, après la fin de la Seconde Guerre mondiale. Au-delà de 2 500 nm la lumière émise par l'atmosphère, le bâtiment, l'astronome et le télescope lui-même vient perturber considérablement les observations. Il faut donc refroidir le dispositif observationnel. Toutefois, on ne peut refroidir l'ensemble du télescope, car de la glace se formerait rapidement par condensation de la vapeur d'eau atmosphérique sur toutes les surfaces exposées. Pour contourner le problème, on peut masquer périodiquement la lumière en provenance de l'objet céleste observé. Pendant ce temps, le détecteur mesure le bruit de fond global lié à l'environnement, que l'on soustrait ensuite à l'intensité totale mesurée en présence de l'objet ; c'est la technique du *chopping*.

L'astronomie infrarouge ne débuta véritablement que dans les années 1950, quand les astronomes eurent accès aux détecteurs au sulfure de plomb (PbS), la sensibilité des détecteurs étant accrue en opérant un refroidissemet à 77 K à l'aide d'azote liquide. Aux alentours de 1960, Harold L. Johnson définit le premier système infrarouge de magnitudes. Son photomètre couvrait les bandes R, I, J, K and L (jusqu'à 4 μm). Johnson et son équipe mesurèrent ainsi plusieurs milliers d'étoiles.

Néanmoins les premiers détecteurs à être pointés vers le ciel ne le furent qu'en considérant l'apport de ces observations comme de simples ajouts ou extensions au spectre visible. Le riche potentiel scientifique véhiculé par les observations infrarouges était encore insoupçonné.

Leur importance véritable ne fut comprise qu'en 1965, après le premier relevé infrarouge du ciel nord, à 2 μm et avec une résolution angulaire d'environ $4'$, réalisé par Gerry Neugebauer et Robert Leighton en Californie au Mont Wilson. Stephan Price effectua un relevé similaire dans le ciel sud. Les résultats de ce relevé attestèrent la présence d'une dizaine d'objets jusque-là inconnus des télescopes traditionnels munis de détecteurs adaptés à la lumière visible. Une prise de conscience s'opéra alors. Ces objets qui ne sont visibles

FIGURE 1.2 : *La nébuleuse d'Orion M42 - A gauche, la région de l'amas du Trapèze vue en lumière visible par le télescope spatial Hubble, à l'aide de la caméra grand champ WFPC2. A droite, la même région observée cette fois dans l'infrarouge proche à l'aide de la caméra NICMOS embarquée à bord d'Hubble. Cette formidable région de formation d'étoiles laisse apparaître dans l'infrarouge une myriade de toutes jeunes étoiles autrement enfouies dans les nuages de poussière et de gaz et indétectables dans le visible.*

que dans l'infrarouge soulevèrent nombre de questions et apportèrent un changement radical à la vision que l'on avait de l'Univers. Et si le ciel infrarouge se révélait différent du ciel visible si familier, aussi bien qualitativement que quantitativement ? (Fig. 1.2)

1.2.2 Découverte d'un nouveau monde

Le domaine infrarouge couvre un large éventail de longueurs d'ondes, de 700 nm (fin du spectre visible) à 100 000 nm (début du spectre submillimétrique). C'est un régime clef en astronomie. Le ciel infrarouge se distingue considérablement de ce que l'on peut voir dans le visible. En effet, les objets traditionellement visibles sont ceux qui sont assez chauds pour émettre dans le visible et ceux dont la lumière n'est pas absorbée par la poussière qui hante l'Univers. Alors que dans l'infrarouge, tout devient possible. On détecte les objets froids et les objets environnés de poussières. Mais l'infrarouge présente aussi des avantages dans d'autres domaines comme l'astrochimie et la cosmologie.

Poussières interstellaires

L'atténuation graduelle de la lumière des étoiles est due à une composante relativement uniforme et très diluée de la matière interstellaire : les grains de poussière interstellaire, composés de carbone, silicates, glaces et de fer. Ces grains représentent à peine 0.1% de l'ensemble de la matière interstellaire et leur taille varie de 100 à 1 000 nm. Au bout de 1 000 a.l. la lumière visible est atténuée d'un facteur deux. Le centre de la Voie Lactée étant situé à 26 000 a.l., la lumière visible qu'il émet est atténuée d'un facteur de plusieurs millions ! Il semble donc illusoire de vouloir étudier une telle région du ciel. Or ce phénomène d'atténuation que l'on nomme extinction, n'est vraiment efficace qu'avec la lumière visible. On comprend donc tout l'intérêt qu'il y a à observer le ciel avec la lumière infrarouge ! Les longueurs d'ondes infrarouges les plus longues pénètrent au travers des nuages de poussières interstellaires et laissent apparaître les objets qu'ils renferment alors que ces nuages sont d'ordinaire opaques aux longueurs d'ondes visibles.

L'extinction s'accompagne également d'un effet de rougissement interstellaire, car l'action des grains de poussières augmente rapidement avec les plus petites longueurs d'ondes. L'importance de la couleur rouge s'en trouve ainsi renforcée. La lumière provenant des étoiles est donc d'autant plus dominée par le rouge que celles-ci sont plus éloignées.

En infrarouge on observe aussi des objets plus rouges que prévus à partir de leurs spectres visibles. Ces objets, nommés à excès infrarouge, sont des étoiles environnées de coquilles de poussières ou bien accompagnées d'émission circumstellaire libre-libre ou *free-free*. Enfin les grains de poussière circumstellaires, quant ils sont chauffés par des étoiles jeunes, réémettent l'énergie qu'ils absorbent à des longueurs d'ondes moins énergétiques, dans l'infrarouge lointain. Ainsi la poussière est elle-même une source de rayonnement infrarouge.

L'infrarouge permet donc de voir plus d'étoiles rouges, de voir plus loin (notamment en direction du disque Galactique, d'ordinaire très obscurci dans le visible) et de détecter des régions de formation d'étoiles.

Objets froids

De plus, les objets de faibles températures peuvent être détectés grâce à leur émission infrarouge. Plus on examine le ciel à de grandes longueurs d'ondes, plus les objets qui apparaissent comme les plus proéminents ont tendance à avoir des températures plus basses. La branche la plus froide des populations stellaires devient prédominante. En effet, une étoile peut être considérée comme un corps noir et dans ce cas la longueur d'onde λ_{max} correspondant à son maximum d'intensité spectrale est donnée par :

$$\lambda_{\max} \times T = 2898 \text{ K.}\mu\text{m}$$

où T est la température de surface de l'étoile en Kelvins. Cette relation est connue sous le nom de loi de Wien et traduit le déplacement du pic d'émissivité spectrale (Fig. 1.3). À titre d'exemple, notons que la température de surface du Soleil est de $\sim$ 6 000 K, ce qui correspond à $\lambda_{\max} \simeq 0.55\ \mu\text{m}$ (lumière verte) qui est également le maximum de sensibilité de l'œil humain. En comparaison, une étoile plus froide que le Soleil aura son pic dans le rouge, alors qu'une étoile plus chaude aura son pic dans le bleu. Les étoiles froides ($\sim$

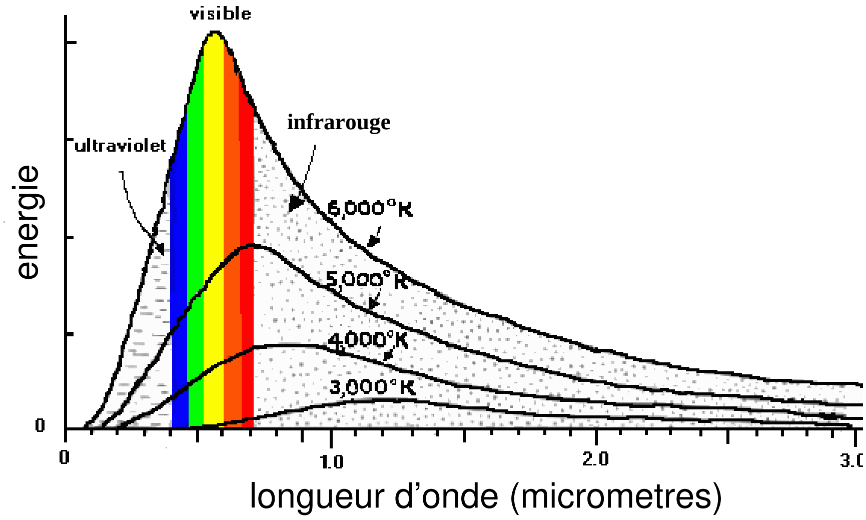


FIGURE 1.3 : *Loi de Wien : déplacement du pic d'émissivité d'un corps noir en fonction de sa température*.

3 000 K) ont le maximum de leur pic d'émissivité aux alentours de $1\ \mu\mathrm{m}$. La poussière a souvent une température de quelques centaines de degrés, avec un maximum à environ $10\ \mu\mathrm{m}$. Quant aux traces du Big Bang, observables sous la forme du rayonnement de fond cosmologique à 3 K, elles se traduisent aux alentours de $1\,000\ \mu\mathrm{m}$, juste au-delà des limites infrarouges et au commencement du domaine spectral radio.

Chimie spatiale

Les transitions rotationnelles et vibrationnelles des molécules, et en particulier la très abondante molécule de dihydrogène $\mathrm{H_2}$, produisent des photons infrarouges de basse énergie. L'infrarouge est riche en transitions atomiques et moléculaires et caractéristique de régions froides comme les nuages de poussières où l'on trouve souvent des agencements de molécules complexes tels les composés organiques.

Cosmologie

Pour finir, l'expansion de l'Univers décale vers l'infrarouge les spectres des galaxies les plus lointaines. Pour les plus distantes (et aussi les plus jeunes) d'entres elles, cet effet de décalage Doppler vers le rouge, dit de *redshift*, est si important que les galaxies ne deviennent plus détectables que dans l'infrarouge. Les relevés infrarouges profonds révèlent un Univers jeune, violent, évoluant très rapidement et avec un taux de formation d'étoiles important.

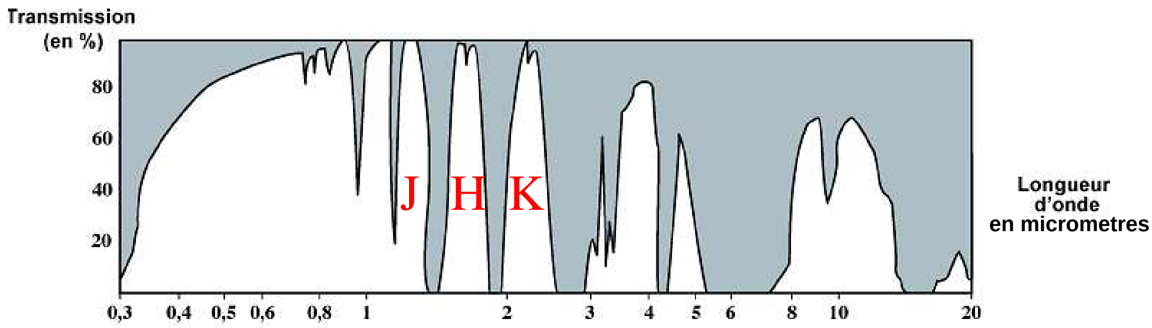


FIGURE 1.4 : *Fenêtres atmosphériques et position des filtres infrarouges sur ces fenêtres.*

1.2.3 Fenêtres atmosphériques et lente ascension vers l'espace

En 1961, à la Texas Instruments Corporation, Frank Low inventa le bolomètre au germanium dopé au gallium, rendant ainsi possible l'étude des longueurs d'ondes infrarouges plus longues, avec l'aide notamment des relevés en fusées, ballons ou avions. En 1970, les cellules en sulfure de plomb furent remplacées par des photodiodes à l'antimoniure d'indium (InSb) et l'amélioration des bolomètres et des détecteurs grâce à un dopage approprié en silicone permirent l'étude des longueurs d'ondes plus longues. En 1979, une nouvelle génération de télescopes (2-4 m) dédiés à l'astronomie infrarouge furent mis en opération : l'UKIRT (*United Kingdom InfraRed Telescope*) de 3.8 m et l'IRTF (*InfraRed Telescope Facility*) de 3 m de la NASA, tous deux basés à Hawaii sur le site du volcan éteint Mauna Kea à 4 200 m au-dessus du niveau de la mer.

Néanmoins des observations plus détaillées du ciel infrarouge nécessitent de relever plusieurs défis. Tout d'abord, l'atmosphère terrestre empêche le passage d'une grande partie de la lumière infrarouge du ciel. L'atmosphère terrestre n'est pas vraiment transparente aux radiations infrarouges. Seules quelques longueurs d'ondes permettent de regarder au travers de l'atmosphère (Fig. 1.4). Les principales sources d'opacité atmosphérique sont la vapeur d'eau (Fig. 1.5) et le dioxyde de carbone. La quantité de vapeur d'eau varie grandement non seulement en fonction de l'altitude et de la latitude, mais aussi de la température et donc de la période de l'année (saisons). Le contenu atmosphérique en vapeur d'eau chute avec l'altitude. C'est pourquoi la plupart des observatoires infrarouges sont situés dans des sites désertiques à haute altitude. En dessous et aux alentours de 2 μm le fond du ciel nocturne est dominé par la diffusion Rayleigh liée à la lumière de la Lune et des étoiles, auquel vient s'ajouter l'*airglow* lié à OH. Ensuite vient la principale source extraterrestre qui contribue au rayonnement de fond. Il s'agit de la lumière zodiacale liée aux poussières présentes dans le système solaire au niveau de l'écliptique. Au delà de 100 μm l'émission émise par la poussière interstellaire domine : ce sont les cirrus infrarouges. Pour finir intervient le rayonnement de fond cosmologique ou CMB (*Cosmic Microwave Background*), le fossile du Big Bang, à environ 300 μm.

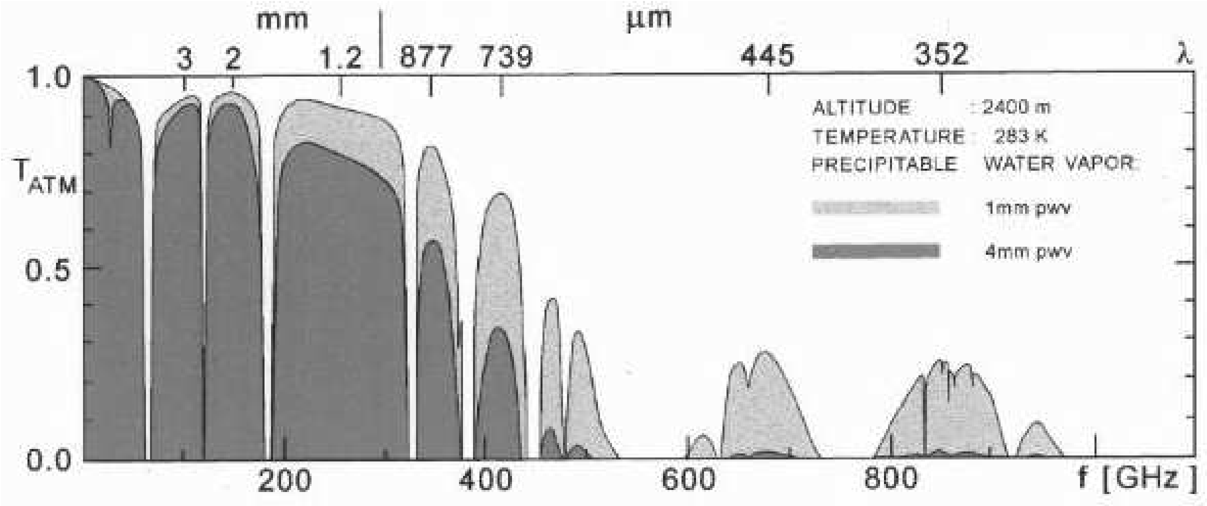


FIGURE 1.5 : *Variation de la transparence atmosphérique en fonction de la quantité de vapeur d'eau présente.*

Quand l'astronomie infrarouge décolle...

Au rayonnement émis par la source astrophysique vient s'ajouter le fort rayonnement infrarouge émis par l'atmosphère elle-même, venant ainsi contaminer le rayonnement céleste. Or, depuis les télescopes au sol, les astronomes cherchent à observer des objets dont l'émission est jusqu'à un million de fois plus faible que celle de l'atmosphère. Cela revient à vouloir observer des étoiles en plein jour avec un télescope optique traditionnel, une tâche bien ardue. Les astronomes sont vite arrivés à la conclusion qu'il fallait observer au-dessus de l'atmosphère pour réaliser une exploration profonde du ciel infrarouge. À partir de ballons lancés à 40 km d'altitude et d'avions furent obtenues les premières images infrarouges du centre Galactique. À partir de 1974, l'avion à réaction C-141A *Kuiper Airborne Observatory* ou KAO de la NASA, basé en Californie au *Ames Research Center*, réalisa une série de vols pouvant atteindre 41 000 pieds avec à son bord un télescope de 0.9 m. Le KAO permit de découvrir entre autres les anneaux autour d'Uranus en 1977.

Mais ces méthodes ne permettant pas de s'affranchir totalement de l'atmosphère terrestre, on lança ensuite des fusées (250 km d'altitude). La première tentative eut lieu à bord d'une fusée Atlas de l'US Air Force. Cette mission échoua à cause du système de refroidissement. En effet, les détecteurs infrarouges doivent être refroidis à l'azote liquide (LN_2) jusqu'à atteindre de basses températures (-196 °C, 77 K) pour éviter toute contamination artificielle des sources astronomiques. On utilise aussi l'hélium liquide (LHe ; -269 °C, 4 K). Le refroidissement des détecteurs possède deux avantages. Tout d'abord il permet de s'affranchir du rayonnement thermique lié aux filtres, lentilles et des structures de soutien en métal qui entourent les détecteurs. Pour finir, les basses températures permettent de réduire grandement le bruit de fond thermique interne aux détecteurs, produisant ainsi un large gain en sensibilité. Malheureusement les fluides de refroidissement s'évaporent très rapidement et lors du premier lancement le vol fut retardé si bien que la température des détecteurs grimpa et qu'ils s'avérèrent hors de fonctionnement une fois la fusée lancée. Par la suite, les fusées permirent d'obtenir jusqu'à cinq minutes de temps (AFGL, *Air Force Geophysics Laboratory*) d'observation au-dessus de l'atmosphère mais très rapidement les astronomes pionniers de l'infrarouge aspirèrent à des temps plus longs, de l'ordre de l'année, que seul un véritable télescope spatial pourrait leur fournir. De plus, dans l'espace, seuls les cirrus infrarouges et l'émission zodiacale viennent augmenter le fond de ciel. Le télescope orbital présentant l'avantage d'être dans le vide spatial, il n'y a pas de vapeur d'eau et on peut donc le refroidir entièrement à de très basses températures sans craindre de condensation.

1.2.4 Ère spatiale et progrès technologiques

La technologie de la mise en orbite des satellites ayant été mise au point à la fin des années 50, le rêve des astronomes se concrétisa en 1983 avec le lancement d'IRAS (*Infra-Red Astronomical Survey*), dédié à la réalisation d'un relevé complet du ciel aux longueurs d'ondes 12, 25, 60 et 100 μm, et réunissant les États-Unis, le Royaume-Uni et les Pays-Bas. IRAS permit à l'astronomie infrarouge d'effectuer un véritable bond en avant. Son succès fut considérable et le télescope resta opérationnel 11 mois, jusqu'à épuisement des réserves embarquées d'hélium liquide et réchauffement du télescope et des détecteurs qui perdirent alors toute sensibilité. IRAS catalogua environ 250 000 sources infrarouges. Les observations spatiales infrarouges ont aussi l'avantage de s'affranchir du rayonnement thermique lié à l'optique du télescope, qui vient rehausser le bruit de fond de façon considérable.

Une nouvelle ère de l'astronomie infrarouge débute en 1986 avec les premiers résultats astronomiques associés à une nouvelle génération de détecteurs regroupés en matrice. L'accumulation de pixels augmente grandement l'efficacité des observations et contribue au développement des caméras infrarouges permettant de produire rapidement des images. Jusqu'alors, bien que sensibles, les détecteurs n'étaient pas particulièrement adaptés pour prendre de larges images du ciel comme dans le visible On produisait des cartes obtenues point par point sous forme de trames (*rasters*), à l'aide d'un détecteur unique dont le champ était typiquement de l'ordre de 2 à 3 secondes d'arc de diamètre, et que l'on déplaçait d'avant en arrière.

Dans les années 90, le développement de l'optique adaptative permet de corriger les images des distorsions induites par la turbulence atmosphérique. Ce système a considérablement amélioré la résolution des images prises à partir du sol, sachant que ce système est d'autant plus efficace que la longueur d'onde observée est grande et donc particulièrement propice à l'infrarouge.

1.3 Diverses bandes photométriques

L'infrarouge proprement dit, technologiquement parlant, démarre à $1.1~\mu\mathrm{m}$, là où les technologies appropriées au visible, telles que la photographie, cessent d'être utilisables. Cela correspond également à la limite au delà de laquelle l'œil humain perd toute sensibilité. L'infrarouge est traditionnellement divisé en trois catégories : l'infrarouge proche (0.75 à 5 $\mu\mathrm{m}$), moyen (5 à 25 $\mu\mathrm{m}$) et lointain (25 à 350 $\mu\mathrm{m}$). Au delà, on utilise généralement des récepteurs radio et le rayonnement reçu a tendance à être traité non plus comme une succession de photons individuels mais plutôt comme une onde.

La photométrie infrarouge au sol se divise en deux régimes autour de 5 $\mu\mathrm{m}$, justifiés par le type de détecteurs utilisés :

- les bandes J (1.25 $\mu\mathrm{m}$), H (1.65 $\mu\mathrm{m}$), K (2.2 $\mu\mathrm{m}$), L (3.5 $\mu\mathrm{m}$) et M (4.8 $\mu\mathrm{m}$) pour les plus petites longueurs d'ondes,
- et les bandes N (10 $\mu\mathrm{m}$) et Q (20 $\mu\mathrm{m}$) pour les plus grandes.

Chacune des bandes correspond plus ou moins à une fenêtre de la transmission atmosphérique (Fig. 1.4).

La première photométrie publiée dans ces bandes a été faite par Johnson en 1962 pour les bandes K et L. Plus tard en 1964, il publia un échantillon photométrique pour 256 étoiles définissant ainsi le système Arizona $JKLM$ standard. La bande H fut introduite plus tardivement en 1968. Avec les progès des détecteurs, la bande L fut transformée en bande L' (3.8 $\mu\mathrm{m}$). Des variations de la bande K existent aussi maintenant sous les noms K' (bande passante 1.9 – 2.3 $\mu\mathrm{m}$, dédié au site de Mauna Kea) et K_s (K short, bande passante 2.0 à 2.3 $\mu\mathrm{m}$). Elles ont une coupure en longueur d'onde plus courte afin de s'affranchir plus clairement du rayonnement de fond thermique terrestre qui commence à devenir important dans cette région du spectre. Alors que les bandes J et H ne sont pas affectées par les radiations thermiques environnantes, les sensibilités des bandes K, L et M s'en trouvent dégradées.

Les versions modernes du système introduit par Johnson dérivent des travaux de Glass. Le système $JHKLM$ est moins bien standardisé que les autres systèmes. Bien souvent, chaque observatoire définit son propre système afin de se caler au mieux sur les fenêtres atmosphériques. Chaque système diffèrera donc légèrement de ceux en vigueur dans les autres observatoires. Ces différences sont liées aux comportements variables des fenêtres atmosphériques infrarouges en fonction du lieu géographique et donc des observatoires. Les différents systèmes photométriques existants sont donnés dans le Tab. 1.1. Comme pour le traditionnel système Johnson-Morgan, le point zéro du système $JHKLM$ est défini de façon à ce qu'une étoile non rougie de type spectal A0 aie la même magnitude dans toutes les couleurs : $J = H = K = L = M$ $(= U = B = V)$. L'étoile standard utilisée est Véga de la Lyre.

En raison des spécificités inhérentes à chaque système photométrique, un grand soin doit être apporté à la comparaison de données en provenance de deux observatoires différents. En matière de calibration photométrique et pour plus de détails, on pourra voir Bersanelli et al. (1991), Bouchet et al. (1991) et Straizhis (1992). Simons & Tokunaga (2002) tentent de standardiser les différents systèmes.

Système	Institut	Bandes	Référence
Arizona	Lunar and Planetary Lab.	$JKLM$	Johnson (1964)
SAAO	South African Astron. Obs.	$JHKL$	Glass (1974)
		$JHKL$	Carter (1990)
ESO	European Southern Obs.	$JHKLM$	Wamsteker (1981)
		$JHKL$	Engels et al. (1981)
AAO	Anglo-Australian Obs.	$JHKL'$	Allen & Cragg (1983)
MSO	Mount Stromlo Obs.	JHK	Jones & Hyland (1982)
CIT	California Inst. Technol.	JHK	Frogel et al. (1978)
		$JHKL$	Elias et al. (1982, 1983)
UNAM	Univ. Autonoma de Mexico	$JHKL'M$	Tapia et al. (1986)
UKIRT	Joint Astron. Centre, Hawaii	JHK	Casali & Hawarden (1992)
		JHK	Hunt et al. (1998)

TABLEAU 1.1 : *Principaux systèmes photométriques infrarouges, d'après Bersanelli et al. (1991).*

1.4 Grands relevés infrarouges

Les premiers travaux dans l'infrarouge s'attachèrent à étendre la couverture spectrale des objets bien brillants dans le visible. Cependant, pour apprécier tout le contenu du ciel infrarouge, des relevés complets du ciel sont nécessaires. Parmi les principaux relevés qui permirent de dévoiler les secrets du ciel infrarouge, on distingue les relevés au sol des relevés spatiaux.

1.4.1 Relevés au sol

Relevé TMSS

Le *Two-Micron Sky Survey* ou TMSS (relevé à deux microns), fut réalisé par Neugebauer & Leighton (1969). Ce fut la première tentative de cartographie systématique du ciel infrarouge. Les déclinaisons inférieures à $-33°$ ne furent pas couvertes. Le catalogue qui résulta de ces observations contient 5 562 étoiles plus brillantes que la magnitude $K = 3$, avec une précision sur les positions de l'ordre de la minute d'arc. Le relevé fut entrepris dans les bandes I (0.84 μm) et K (2.2 μm) afin de pouvoir estimer la couleur des sources. En dépit d'une assez faible sensibilité, ce relevé donna d'étonnants résultats, avec notamment la découverte d'un certain nombre de sources très brillantes dans l'infrarouge, bien que faibles dans le visible : des étoiles jeunes enfouies dans des nébuleuses mais aussi des étoiles évoluées telles que des AGB carbonées, totalement obscurcies dans le visible par une épaisse coquille de poussière interstellaire.

DENIS

À la fin des années 80, il devint évident qu'un relevé complet du ciel était techniquement possible. L'accession aux technologies militaires sur les détecteurs infrarouges adaptables à l'astronomie fut cruciale. De même que les avancées informatiques pour autoriser un traitement efficace des futures masses de données et leur distribution. En Europe, plusieurs laboratoires, menés par l'Observatoire de Paris-Meudon, engagèrent leurs efforts en direction d'un même projet : le *DEep Near-Infrared Survey of the Southern Sky* (relevé profond du ciel sud dans le proche infrarouge), dénommé DENIS. Le Ministère français de l'Éducation finança les premières études de faisabilité. Le Département de Recherche Spatiale à Meudon s'occupa du *design* des deux caméras infrarouges présentes dans DENIS. Dans le même temps, plusieurs laboratoires européens offrirent leur expertise technique et les ressources humaines pour recevoir les fonds nécessaires dans le cadre du plan *Science* de la Commission Européenne et ensuite dans le cadre du programme *Human Capital and Mobility*. Le relevé démarra en 1996 et a consisté en la couverture complète du ciel sud pour des déclinaisons inférieures à deux degrés, dans les deux bandes photométriques infrarouges J (1.22 μm) et K_s (2.15μm), mais également dans la bande visible I (Gunn–i, 0.79 μm), ce qui fait toute l'originalité du projet DENIS. Les observations furent conduites à partir du télescope de 1 m de l'ESO à La Silla au Chili, alors entièrement dédié à ce projet. La résolution était de $1''$. DENIS a conduit à une amélioration de plus de quatre ordres de magnitude dans la bande K par rapport à l'IRC, l'*InfraRed Catalogue* issu des observations du TMSS.

2MASS

À la fin des années 80, la combinaison des premiers résultats scientifiques IRC dans le domaine stellaire et des résultats liés à émanant d'IRAS pour le domaine extragalactique, convint les astronomes de la nécessité d'un nouveau relevé dans le proche infrarouge, mais cette fois beaucoup plus profond. Le développement technologique, tant d'un point de vue de la sensibilité des détecteurs que de l'informatique et de l'accélération des réseaux, rendit les choses envisageables. Un tel relevé fut classé priorité numéro un au *Decadal Survey* en 1991. Le projet fut soumis par l'Université du Massachussetts et l'IPAC (*Infrared Processing and Analysis Center*) en réponse à cette recommandation. C'est ainsi que naquit le *Two-Micron All Sky Survey* (relevé complet du ciel à deux microns), dénommé 2MASS. Il réalisa des observations uniformes du ciel complet simultanément dans les bandes J (1.24 μm), H (1.66 μm) et K_s (2.16 μm) avec une résolution de $2''$ par pixel. Les sources plus brillantes que 1 mJy dans chaque bande furent détectées avec un rapport signal sur bruit plus grand que 10, ce qui représente une amélioration en sensibilité de 80 000 par rapport à l'IRC. Les observations 2MASS furent menées sur des télescopes de 1.3 m. Le ciel nord fut observé à partir du Mont Hopkins en Arizona de juin 1997 à décembre 2000 et le ciel sud à partir du Cerro Tololo au Chili de mars 1998 à février 2001. Au total, 24.5 To de données ont été collectées sous la forme d'images brutes, avec un taux de couverture de 99.998%.

Mais aussi…

SPIREX ou *South Pole InfraRed EXplorer* est un des premiers observatoires construits au pôle sud pour tester la pertinence d'un tel site pour des observations infrarouges. Le télescope de 60 cm et la caméra infrarouge proche furent mis en opération en 1994.

Le projet de télescope infrarouge pour l'Antarctique, IRAIT ou *Italian Robotic Antarctic Infrared Telescope*, est un instrument destiné principalement à l'imagerie dans le domaine spectral 2.5-20 μm. Il vise à réaliser les meilleures images grand champ jamais obtenues dans ce domaine spectral au sol. Il existe une base scientifique au pôle sud à 2 800 m d'altitude sur le plateau antarctique, la base Concordia, sur le site du dôme C. Cette station est située à l'intérieur des terres antarctiques, à 950 km de la côte, en face de l'Australie et résulte d'un coopération franco-italienne. L'atmosphère particulièrement stable, pure et sèche ainsi que la faible température de surface, -60 °C, en font un site idéal pour l'astronomie infrarouge.

1.4.2 Relevés spatiaux

AFCRL et AFGL

Le laboratoire de recherche de Cambridge de l'US Air Force (*US Air Force Cambridge Research Laboratories* ou AFCRL, qui fut remplacé par l'*Air Force Geophysics Laboratory* ou AFGL et actuellement *Philips Laboratory*) réalisa une série d'expériences entre 1970 et 1984. Il utilisa une série de fusées contenant des télescopes de 16.5 cm pour effectuer des relevés dans les trois bandes 3–5 μm, 8–14 μm et 16–24 μm. Le produit de ces observations donna naissance au catalogue AFGL, qui permit de mettre en évidence de nouvelles classes d'objets très froids tels que des proto-étoiles, des régions HII et des nébuleuses proto-planétaires.

IRAS

Le lancement du satellite IRAS, *InfraRed Astronomical Satellite*, a été motivé par la volonté d'observer le ciel infrarouge sur des temps plus longs que ceux autorisés par les ballons ou les fusées. Une autre des raisons essentielles était d'étendre le domaine spectral observable à des longueurs d'ondes supérieures à 50 μm. Les résultats des observations IRAS furent publiés sous forme de catalogues de sources ponctuelles (plus de 254 000) et de sources étendues, de spectres basse résolution, d'images et de diverses cartes du ciel. La NASA créa l'IPAC (*Infrared Processing and Analysis Center*) afin d'exploiter dans toute leur ampleur ces bases de données.

		IRAS	ISO	NICMOS
Hauteur	(m)	3.6	5.3	2.2
Largeur	(m)	3.2	3.5	0.9
Profondeur	(m)	1.7	2.8	0.9
Miroir	(m)	0.57	0.60	2.54
Poids	(kg)	1 000	2 400	370
Longévité		25.01.83 à 23.11.83	17.11.95 à 16.05.98	02.97 à ?
Orbite		$\sim$ circulaire	elliptique	circulaire
altitude	(km)	900	1 004 à 70 611	593
Température	(°C)	-269	-269	-213
Longueur d'onde	(μm)	8 à 120	2.5 à 240	0.8 à 2.5
Collaboration		EU, RU, Pays-Bas	ESA, ISAS, NASA	ESA, NASA

TABLEAU 1.2 : *Comparaison des caractéristiques de trois satellites spatiaux infrarouges : IRAS, ISO et NICMOS. Pour NICMOS, seul le poids de l'instrument est comptabilisé.*

COBE

Le satellite COBE, *COsmic Background Explorer*, développé par la NASA, a fourni les bases de données les plus complètes sur les rayonnements de fond infrarouges. Il fut lancé en 1989 et confirma le rayonnement à 2.7 K.

Autres exemples : IRTS, MSX,...

Les Japonais lancèrent un satellite infrarouge, IRTS (*InfraRed Telescope in Space*), en mars 1995. Sa mission ne dura que 28 jours et il observa $\sim$ 7% du ciel. Ce fut le premier télescope orbital japonais dédié à l'astronomie infrarouge. Il réalisa des observations simultanées du ciel sur un domaine allant de 1 à 1 000 μm.

Le satellite MSX (*Midcourse Space Experiment*) a été lancé par les Américains le 24 avril 1996. Il pesait 2 700 kg et fut en opération pendant 10 mois. Son domaine spectral s'étend de 4.2 à 26 μm, avec une résolution beaucoup plus élevée qu'IRAS. Plus de 200 Go (Giga-octets, 10^9) de données furent produits. L'objectif principal des expériences astronomiques à bord de MSX était de compléter les recensements célestes dans l'infrarouge moyen et notamment les zones manquées par IRAS et COBE/DIRBE et celles où la sensibilité d'IRAS s'était trouvée dégradée, entre autres dans les régions à forte densité spatiale.

La mission spatiale WIRE (*Wide-field InfraRed Explorer*) est l'une des premières missions du programme *Origins* de la NASA. Elle consistait à faire un relevé infrarouge dans les bandes centrées sur 12 et 25 μm sur une durée de quatre mois et qui aurait détecté principalement les galaxies à flambées d'étoiles, dites *starburst*, qui émettent la plupart de leur énergie dans l'infrarouge lointain. WIRE aurait couvert 100 degrés carrés du ciel et détecté des sources 200 à 500 fois plus faibles qu'IRAS pour 25 μm, et 500 à 2 000 fois plus faibles qu'IRAS à 12 μm. Malheureusement, après son lancement en mars 1999, WIRE fut incapable de mener à bien sa mission première et a été reconverti dans le suivi à long terme des étoiles brillantes avec pour buts l'astérosismologie et la recherche de planètes.

1.5 Observatoires spatiaux

Des projets récents de satellites infrarouges ont été conçus dans le but d'être utilisés comme de véritables observatoires terrestres : ils ne sont pas dédiés uniquement à la réalisation d'un relevé global du ciel, mais peuvent aussi être pointés en direction de sources cibles.

Satellite ISO

Avant même le lancement d'IRAS, l'Agence Spatiale Européenne (ESA, *European Space Agency*) avait déjà entrepris des études préliminaires pour un possible successeur, qui fut accepté peu de temps après le lancement d'IRAS. Ce satellite, nommé ISO, *Infrared Space Observatory*, a été lancé le 17 novembre 1995 par un lanceur Ariane depuis Kourou. Il devait rester opérationnel 20 mois mais grâce à une technologie rigoureuse, la durée de vie du satellite atteignit 28 mois. Il a cessé d'être opérationnel en avril 1998, après épuisement de ses réserves d'hélium liquide et fut éteint le 16 mai après avoir observé la supergéante η Canis Majoris. Le satellite fut brûlé dans l'atmosphère afin de préserver l'environnement spatial. ISO opéra de 2.5 à 240 μm avec un miroir primaire de 60 cm de diamètre. ISO disposait d'une sensibilité 1 000 fois meilleure qu'IRAS et d'une résolution angulaire 100 fois meilleure à 12 μm.

Caméra NICMOS à bord de Hubble

NICMOS (*Near-Infrared Camera and Multi-Object Spectrometer*) est un instrument qui a été installé à bord du télescope spatial Hubble, le HST ou *Hubble Space Telescope*. NICMOS couvre un domaine spectral de 0.8 à 2.5 μm et permet à la fois de faire de l'imagerie et de la spectroscopie.

1.6 Futurs projets

Avant de passer en revue les futurs projets, qui sont quasiment tous spatiaux, notons qu'il existe toujours l'alternative des observatoires embarqués à bord de ballons ou d'avions. Cette méthode d'observation a l'avantage d'avoir un coup de revient peu cher. De plus, l'instrumentation peut être facilement changée. En revanche de tels observatoires ne peuvent fonctionner aux basses températures rendues possibles pour les télescopes spatiaux grâce à la cryogénie.

SOFIA

SOFIA, *Stratospheric Observatory For Infrared Astronomy*, est une coopération entre la NASA et l'Allemagne : un Boeing 747SP modifié pour accueillir un télescope de 2.5 m. Il viendra remplacer le KAO. Ce sera le plus grand observatoire aéroporté du monde. Il devrait commencer à voler en 2004. Il offrira aux astronomes une plate-forme unique pour étudier le ciel aux longueurs d'ondes comprises entre 5 et 300 μm, donc dans le moyen et lointain infrarouge.

VLTI

Le VLTI de l'ESO, *Very Large Telescope Interferometer*, est la combinaison des quatre unités du VLT (quatre fois 8 m) et de plusieurs télescopes auxiliaires de 1.8 m. Une fois totalement opérationnel, le VLTI fournira une résolution angulaire de l'ordre de la milli-seconde d'arc ainsi qu'une grande sensibilité grâce à des lignes de base atteignant 200 m de long.

ASTRO-F

ASTRO-F, auparavant dénommé IRIS, *InfraRed Imaging Surveyor*, est la deuxième mission japonaise spatiale infrarouge. Ses buts scientifiques sont l'évolution des galaxies, la formation des étoiles et des planètes, les naines brunes et leur lien avec la matière noire. Le lancement est prévu pour février 2004 sur une orbite à 745 km. ASTRO-F est équipé d'un télescope de 67 cm refroidi à 6 K et est est composé de deux instruments. Le premier effectuera un relevé complet du ciel dans l'infrarouge lointain dans le domaine 50–200 μm avec des résolutions angulaires de 30 à 50″. Le deuxième instrument prendra des images profondes de régions prédéfinies du ciel, dans l'infrarouge proche et moyen. Le champ sera de $10'$ et la résolution spatiale d'environ $2''$.

SIRTF

SIRTF, *Space InfraRed Telescope Facility*, a été lancé de Cap Canaveral en Floride le 25 août 2003. Sa mission durera deux ans et demi. Son domaine spectral s'étend de 3 à 180 μm. Le télescope fait 0.85 m de diamètre et est le télescope infrarouge le plus grand jamais lancé dans l'espace. SIRTF est la dernière mission du programme *Great Observatories* de la NASA, visant à couvrir l'observation de l'Univers dans quatre grands domaines : infrarouge, visible, X et Gamma. SIRTF fait aussi partie du programme *Astronomical Search for Origins* dédié à la compréhension de nos origines cosmiques, à la formation et à l'évolution des galaxies, des étoiles et des planètes.

JWST

Maintenant appelé *James Webb Space Telescope* ou JWST, le *Next Generation Space Telescope* ou NGST sera composé d'un miroir de la classe des 6 m doté de grands boucliers solaires et devant être placé au point de Lagrange L2, avec un lancement prévu pour 2009. Le point de Lagrange L2 se situe au-delà de la Lune, loin de la Terre à une distance de 1.5 millions de kilomètres, dans la direction opposée au Soleil. L'imagerie et la spectroscopie couvriront le domaine 0.6–28 $\mu\mathrm{m}$. Avec une durée de vie minimale de 5 ans, l'objectif serait de le faire fonctionner pendant 10 ans. JWST est spécialement conçu dans le but de découvrir et de comprendre la formation des premières étoiles et galaxies, de mesurer la géométrie de l'Univers et sa distribution en matière noire, d'étudier l'évolution des galaxies et la chimie des étoiles ainsi que les processus de formation des étoiles et des panètes. JWST fait partie du progamme *Origins* de la NASA.

FIRST/Herschel

L'observatoire spatial Herschel, auparavant FIRST, sera dédié à la photométrie et la spectroscopie dans le domaine 60–670 $\mu\mathrm{m}$. Son lancement est prévu pour 2007 sur Ariane 5 et il devrait se mettre autour du point de Lagrange L2 pour ensuite être opérationnel sur une durée minimale de trois ans. C'est la quatrième pierre angulaire (CS4) du programme scientifique de l'ESA. C'est le seul dispositif dédié à l'observation dans la partie sub-millimétrique et infrarouge lointain du spectre électromagnétique. Il devrait permettre de découvrir les plus jeunes proto-galaxies. Les objectifs scientifiques d'Herschel sont tournés vers la formation d'étoiles et des galaxies, mais aussi la physique du milieu interstellaire, l'astrochimie et l'étude du système solaire.

Planck et WMAP

Planck Surveyor est la troisième mission de taille moyenne (M3) dans le cadre du programme scientifique Horizon 2000 de l'ESA. Le nom d'origine de la mission était COBRAS/SAMBA mais une fois le projet accepté (fin 1996), il fut renommé en honneur au célèbre scientifique allemand Max Planck (1858-1947), prix Nobel de Physique en 1918. La principale motivation scientifique derrière Planck est la cosmologie. Il s'agit d'un relevé destiné à cartographier le ciel entier dans le millimétrique. Planck mesurera les anisotropies du rayonnement de fond cosmologique avec une précision jusque-là inégalée, à l'image de son homologue WMAP ou *Wilkinson Microwave Anisotropy Probe* de la NASA. Il permettra de tester les théories sur l'origine des structures cosmiques. Planck devrait permettre également de mieux comprendre la Galaxie, grâce au relevé complet du plan Galactique. Initialement prévu pour 2003, le lancement est maintenant annoncé début 2007, sur un vol partagé avec Herschel. Planck se placera également autour du point de Lagrange L2, mais sur une orbite différente de celle d'Herschel.

Darwin et TPF

Darwin sera composé d'une flotille de six télescopes spatiaux, chacun d'au moins 1.5 m de diamètre, fonctionnant sur le mode interférométrique afin d'obtenir une grande résolution spatiale et de pouvoir occulter la lumière éblouissante de l'étoile centrale. Ensemble ces six télescopes vont scanner l'Univers proche (les 1 000 étoiles les plus proches) à la recherche de vie sur des planètes de type terrestre. Le lancement est prévu aux alentours de 2015. Darwin observera dans l'infrarouge moyen. À ces longueurs d'ondes, comparé au visible, le contraste planète/soleil est diminué d'un facteur un million, autorisant ainsi une détection nettement plus envisageable de telles planètes. Darwin sera équipé d'un spectromètre afin de pouvoir analyser les différents constituants atmosphériques de ces planètes. En effet, sur Terre, la présence d'eau, ainsi que l'activité biologique (présence d'ozone, de dioxygène) se traduisent dans le domaine spectral infrarouge. On comprend donc la nécessité pour cette mission de se dérouler dans l'espace afin de se soustraire au rôle bloquant de l'atmosphère terreste à ces longueurs d'ondes et afin de bénéficier d'un refroidissement possible complet de l'instrument pour pouvoir détecter la faible lueur des planètes lointaines. Darwin sera placé au point de Lagrange L2. Darwin permettra de voir en détails la formation des étoiles et de leurs planètes au sein des nuages de poussières.

La NASA dispose d'un projet similaire, *Terrestrial Planet Finder* ou TPF, prévu pour 2012-2015.

ALMA

ALMA, *Atacama Large Millimeter Array* est un télescope millimétrique. C'est un projet commun à l'Europe et l'Amérique du Nord, qui devrait arriver à terme en 2011. Il s'agit d'un réseau de 64 antennes de 12 m situées à 5 000 m d'altitude près de Cerro Chajnantor au nord du Chili, à 40 km à l'est du village historique de San Pedro de Atacama et à 275 km du port d'Antofagasta. Le réseau pourra prendre des configurations allant de 150 m à 10 km. Le domaine couvert s'étend de 10 mm à $250\ \mu\mathrm{m}$ et toutes les fenêtres atmosphériques seront couvertes. Il s'agira de l'instrument le plus grand et le plus sensible du monde dans le domaine millimétrique et sub-millimétrique. La résolution spatiale est de 10 millisecondes d'arc, 10 fois mieux que le VLA et le HST. La résolution en vitesse sera de $0.05\ \mathrm{km.s}^{-1}$. ALMA sera capable de prendre des images de source pouvant s'étendre sur quelques degrés avec une résolution de $1''$.

1.7 Observatoire Virtuel

Dans la section précédente, on a pu se rendre compte que le nombre de missions dans le domaine infrarouge allait croissant, générant toujours plus de données, toujours plus vite. Le même phénomène a lieu pour les autres longueurs d'ondes. Cette inflation est le fruit d'un long cheminement technologique.

Les premières observations astronomiques ont été faites à l'œil nu et enregistrées sur des tablettes en terre cuite à l'époque babylonienne et des papyrus dans l'ancienne Égypte. Répertorier les étoiles suivant leur positions dans le ciel et les classer suivant leur magnitude constituent la fondation de l'étude de l'Univers depuis les temps les plus anciens. Puis la science a avancé main dans la main avec le progrès technologique. De quelques centaines d'étoiles répertoriées dans l'*Almageste* de Claude Ptolémée, 150 après J.-C., s'appuyant sur les travaux d'Hipparque de Nicée et présentant pour la première fois le distinction des étoiles en six grandeurs ou magnitudes suivant leur éclat, on passe à des dizaines de milliers avec l'*Uranographia* de Johann Bode, 1801, l'atlas le plus grand jamais publié, puis des centaines (*Durchmusterungen catalogues*, *Henry Draper Catalogue*) de milliers d'étoiles avec le développement des lunettes et des télescopes. Puis, avec l'avènement des plaques photographiques, on atteint le million d'étoiles (*Astrographic Catalogue*). À la fin du XX^e^ siècle, l'arrivée de l'informatique couplée à l'électronique (CCD, *Charged Coupled Device*), la conquête du spatial et la numérisation des anciennes plaques photographiques caractérisent les grands catalogues modernes qui flirtent avec le milliard d'étoiles (GSC II). Entre les premiers travaux d'Hipparque, qui a jeté les bases de la trigonométrie, et Hipparcos, la première mission spatiale dédiée à l'astrométrie et à la mesure des distances, 2 000 ans d'histoire astronomique se sont écoulés et les données se sont accumulées. Par ailleurs, l'astronomie a une longue tradition en matière d'archivage et de préservation des données car ses objets d'étude sont inaccessibles, souvent variables avec le temps. Garder trace au maximum des événements célestes passés est nécessaire à leur compréhension.

Répartis de part le monde, quelques centres de données, à l'instar du Centre de Données astronomiques de Strasbourg (CDS, Genova et al. 2000), sont chargés d'archiver le produit des observations et de faciliter leur accès à la communauté astronomique mondiale. Or chaque centre dispose de sa propre façon de gérer les données. Dès lors, afin d'exploiter le maximum de leur valeur scientifique, on comprend qu'il devient nécessaire de regrouper ces données dans un cadre commun, afin de standardiser ce mélange hétérogène et rendre l'interaction possible entre les différents services. C'est ainsi qu'émerge le concept d'Observatoire Virtuel (Szalay & Gray 2001, Brumfiel 2002).

1.7.1 Contexte scientifique et technologique

Ces vingt dernières années, la manière dont la science est effectuée a été bénéficiaire et actrice de deux révolutions majeures. Elles constituent les racines d'un véritable changement de paradigme.

Informatique et information

Tout d'abord la révolution informatique des années 80 et 90 a rajouté le calcul informatique comme troisième composante incontournable de la méthodologie classique de la recherche qui contenait jusqu'alors théorie et observation. Les ordinateurs sont devenus plus rapides, moins chers et facilement disponibles, autorisant une part grandissante de chercheurs à inclure la modélisation, l'analyse et la visualisation de données comme outils quotidiens de recherche.

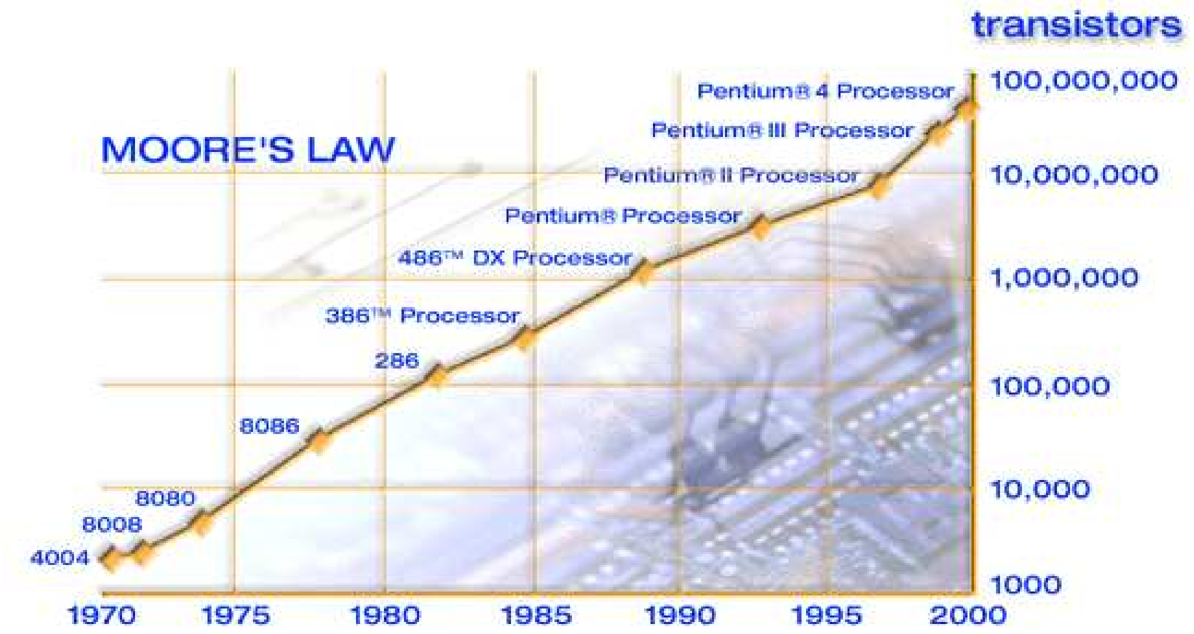


FIGURE 1.6 : *Loi de Moore, d'après* `http://www.intel.com`. *Moore estime qu'elle se poursuivra jusqu'en 2017, date à laquelle elle devrait rencontrer des contraintes physiques liées à la taille des atomes. Il faudrait donc chercher une alternative à la toute puissance du silicium...*

La croissance informatique, formalisée par la loi de Moore, a joué un rôle crucial dans l'histoire de l'informatique. Énoncée pour la première fois en 1965 par Gordon E. Moore, cette loi postule une croissance exponentielle de la puissance des ordinateurs, se traduisant par le doublement annuel des performances des circuits intégrés (mémoires et processeurs). Cette loi a donc contribué à l'accélération du rythme de l'innovation car les ingénieurs de recherche ont alors conçu des systèmes très en avance sur leur temps par rapport aux capacités du moment, en se basant sur cette loi. Moore a revu son estimation en 1975 : le doublement aurait lieu tous les 18 mois et non tous les ans, ce qui correspond tout de même à une croissance très rapide. Cette loi, fondée sur un constat empirique, a été vérifiée par la suite (Fig. 1.6).

Ensuite la révolution de l'information des années 90 a permis aux chercheurs de partager leurs efforts et leurs programmes de recherche sur toute la surface du globe, grâce au développement rapide et peu cher du World Wide Web ou WWW, dont l'infrastructure permet de connecter et distribuer l'information et la puissance logicielle à des flots d'utilisateurs, individus et organisations, tant du point du vue scientifique que commercial ou culturel. Notons également l'importance du paradigme *open source* UNIX/LINUX.

Ces deux révolutions se nourrissent l'une l'autre et contribuent à l'avalanche de données à laquelle l'astronomie observationnelle est actuellement confrontée.

Conséquences astronomiques

Les avancées technologiques autorisent la production de To (Tera-octets, 10^{12}) (plusieurs milliers de milliards) d'images et de catalogues. L'augmentation du volume de données s'accompagne d'un accroissement dans la complexité et la qualité de ces données. Chaque objet est représenté par une dizaine, voire des centaines de paramètres. L'avènement de relevés synoptiques produisant des Po (Peta-octets, 10^{15}) d'informations est imminente (10 To par jour). À titre d'exemple, le taux de production du HST est de 5 Go par jour. Ces avancées technologiques se situent à différents niveaux :

- ouverture du ciel à toutes les longueurs d'ondes du spectre électromagnétique grâce aux télescopes spatiaux et au sol ;
- instrumentation de plus en plus sophistiquée et passage du tout analogue au tout digital ;
- efficacité grandissante des détecteurs permettant de gagner en sensibilité plusieurs ordres de magnitudes ;
- augmentation des vitesses de calcul et acquisition plus rapide des données ;
- disponibilité de logiciels et de supports informatiques capable d'acquérir, réduire et archiver les masses de données produites ;
- capacité de stockage informatique accrue ;
- amélioration de la couverture réseau et de la transmission des données par les réseaux ;
- développement de chaînes de traitement rapides et automatisées permettant un accès public efficace.

Vers une nouvelle donne de l'information

Toutes ces modifications influencent profondément la façon dont la recherche est menée. Pendant des centaines d'années, la science a été le domaine privilégié de quelques individus. Plus récemment, un projet scientifique était limité à un petit nombre d'objets astronomiques, et les observations étaient réalisées par une seule personne, voire une petite équipe. Puis, poussés par leur quête de l'insondable, à force de toujours vouloir extraire les données les plus enfouies, toujours plus lointaines, les astronomes ont contribué à faire émerger une nouvelle tendance. Maintenant de nombreux projets se font à l'échelle nationale, européenne, si ce n'est mondiale, et correspondent à des relevés complets du ciel, toujours plus profonds. Les produits de ces observations sont tellement importants en termes de masses de données que les capacités de stockage, de réduction, d'analyse et d'exploitation d'une simple équipe ne suffisent plus à les valoriser. Le stockage de gros volumes de données n'est pas non plus sans coût. Il faut mettre en place une hiérarchie dans l'archivage de données ainsi que des techniques de compression avancées. Toutes ces conditions se trouvent réunies au sein des centres de données, qui se transforment progressivement en véritables nœuds névralgiques de l'information astronomique. En plus du stockage et du traitement, ils rendent l'accès aux données possible par le biais d'interfaces webs. D'où le résultat majeur qui se dessine suite à toutes ces transformations : l'accès aux données fondamentales n'est plus réservé aux simples observateurs. À terme, le cadre général de l'Observatoire Virtuel devrait permettre de surmonter les difficultés techniques et culturelles qui consistent à partager encore plus largement l'information et à démocratiser l'astronomie. Cependant, ce cadre de pensée novateur nécessite un changement de paradigme, qui n'est pas sans soulever de nombreux défis et susciter des enjeux majeurs.

1.7.2 Problèmes et motivations

Ces 30 dernières années, le nombre de pixels dédiés, par l'intermédiaire du CCD, à l'imagerie du ciel a augmenté 100 fois plus vite que la surface optique disponible en termes de miroirs (Szalay et al. 2001). En effet, la surface optique double tous les 25 ans tandis que le nombre de pixels double tous les deux ans. Ce rythme s'accélère, l'efficacité des instruments et des détecteurs doublant tous les ans. Or le rythme de croissance actuel de la puissance informatique, des capacités de stockage et des vitesses de réseau est à peu près trois fois trop lent pour gérer cette explosion du volume de données disponible. L'élaboration du ciel virtuel pose donc un sérieux problème. Le doublement des performances informatiques n'est réalisé que tous les 18 mois et, pire encore, la capacité des réseaux ne double qu'à peu près tous les 20 mois (loi de Nielsen) ! Quant à la globalisation de l'information, elle entraîne un doublement de la taille du WWW tous les 9 mois (Hobbes' Internet Timeline, http://www.zakon.org/robert/internet/timeline). Les taux de transfert actuels sont de l'ordre de 1 To de données pour 20 jours. Dans l'avenir, des taux de transfert beaucoup plus élévés devraient permettre l'accès généralisé des données à une grande partie de la communauté qui auparavant n'avait pas d'accès direct à ces données. Néanmoins, comme le volume d'information croît plus vite que les performances d'une station de travail, il faut nécessairement se tourner vers de nouvelles stratégies pour gérer le traitement et l'organisation de l'information. La communauté scientifique (mais pas seulement) internationale est mobilisée pour unifier et réconcilier ces deux révolutions, en développant une nouvelle infrastructure technologique.

Afin de réduire les coûts d'observation (temps, argent et ressources humaines), les données collectées à grand prix sur un télescope doivent être archivées de manière à pouvoir être réutilisées facilement dans l'espace-temps par des équipes différentes (ou identiques) ayant des buts scientifiques multiples, et ceci afin d'exploiter pleinement le potentiel des données. Les données seront réutilisables et transmises avec un maximum de précisions aux générations futures et serviront au-delà de leur buts originels. Au besoin, elles pourront même être retraitées, mieux calibrées avec l'émergence de nouveaux algorithmes. Répertorier toutes les données existantes devrait aussi permettre une meilleure gestion des observations, afin d'éviter des observations multiples inutiles et maximiser le potentiel de découverte. L'Observatoire Virtuel, projet par essence à l'échelle mondiale, permettra aux communautés scientifiques économiquement faibles et dans l'incapacité de financer et gérer des infrastructures observationnelles complexes, de bénéficier de cette immense source de données. La recherche devrait s'en trouver dynamisée.

La construction de ce nouvel espace de recherche n'est possible que dans le cadre novateur proposé par la technologie GRID ou Globalisation des Ressources Informatiques et des Données. Mais il ne s'agit pas simplement de problèmes technologiques à résoudre.

Les astronomes ne sont pas formés à travailler avec de telles avalanches de données, et l'esprit humain encore moins apte à appréhender et analyser les millions de données disponibles simultanément. Certes, les buts scientifiques sont clairement définis, et les données présentes, mais il reste encore à construire les méthodes et les outils qui permettront d'extraire l'information pertinente cachée au milieu de ce flot de données.

1.7.3 GRID : une nouvelle infrastructure technologique

La technologie GRID (Foster et al. 2001) consiste à partager des ressources au sein d'une même communauté. Les ressources, distribuées de part le monde, regroupent tous les maillons de la chaîne d'informations : informatique, stockage, réseau, visualisation,. . . Le succès réside dans la formule suivante :

$$\frac{\text{temps de communication}}{\text{temps de calcul}} \ll 1$$

Les systèmes GRIDs ne sont assimilables ni à du calcul distribué, ni à des ordinateurs mobiles, ni à du calcul parallèle. Il y a trois types de GRIDs (Krauter et al. 2000) :

- Grilles de calculs :
  - aggrégat de puissance de calcul ;
  - distribution des ressources pour former un ordinateur virtuel géant.
- Grilles de données :
  - bases de données à grande échelle, partagées ;
  - catalogues de bases de données ;
  - base de donnée virtuelle.
- Grilles de services :
  - temps réel : changer en direct les paramètres d'une expérience ;
  - laboratoires ou observatoires virtuels avec des grilles multi-média (réalité virtuelle) et des grilles sur demande (services dynamiques).

En France, le ministère de la recherche a lancé une ACI (Action Concertée Incitative) intitulée GRID, afin de dynamiser les activités de recherche sur le développement de logiciels permettant la conception et la mise en œuvre de ces infrastructures globales et de développer des actions pluridisciplinaires en vue de l'utilisation de ces équipements. Les capacités de calcul basées sur les technologies GRID émergentes devraient permettre la création d'un environnement de calcul géant composé d'un ensemble distribué de fichiers, bases de données, ordinateurs, instruments et appareils scientifiques. Pour plus de précisions sur cette vision GRID ainsi que sur les dernières techniques en termes de meta-calcul, on pourra se référer au livre de Foster & Kesselman (1999).
Sur le site web du ministère de la recherche[1], on trouve la définition suivante GRID :

[1] `http://www.recherche.gouv.fr/recherche/aci/grid.htm`

La Globalisation des Ressources Informatiques et des Données ou GRID consiste en la possibilité d'offrir aux utilisateurs des ressources informatiques *virtuellement illimitées*, tant en capacité de stockage qu'en capacité de calcul, en les répartissant de manière dynamique sur un ensemble non figé d'équipements et en offrant à l'utilisateur un accès transparent à ces ressources. Cela correspond donc à une évolution des réseaux de stations de travail vers des réseaux dynamiques, hétérogènes et transparents, dits *grilles de calcul*, construits en interconnectant des équipements informatiques standard, des plates-formes expérimentales et des équipements de calcul ou de stockage, à travers un réseau à haut débit. La répartition et le couplage de très grandes bases de données posent des problèmes qui sont, pour partie, étroitement reliés à ceux qui sont étudiés dans le cadre de la problématique *grille de calculs*. Par ailleurs, on assiste à une évolution forte des ordinateurs portables avec l'intégration des capacités de télécommunication ; interconnecter des réseaux sans fils d'équipements mobiles et les utiliser comme des ressources globales est un sujet d'actualité.

Un autre projet, DataGrid, financé par l'Union Européenne à la hauteur de 9.8 millions d'euros et mené par le CERN (Centre Européen pour la Recherche Nucléaire), a pour but la mise en place d'une grille de ressources destinée à l'analyse des données en provenance des expériences scientifiques. La science requiert maintenant le partage coordonné des ressources et une collaboration étroite dans le traitement et l'analyse des masses de données produites et stockées dans de nombreux laboratoires différents, appartenant à des institutions différentes.

Les Po de données distribuées et les dizaines de milliers de disques et processeurs seront utilisés simultanément par des milliers d'utilisateurs aux quatre coins du monde. L'initiative DataGrid développe et teste l'infrastructure technologique qui rendra possible l'implémentation de collaborations scientifiques où les chercheurs pourront réaliser leurs activités quelle que soit leur situation géographique. L'interaction et le partage des instruments et des données se fera à une échelle jusque-là inégalée, mondiale.

Les ressources informatiques sont développées à des endroits spécifiques pour servir les besoins des utilisateurs sans avoir la nécessité d'améliorer les ressources au niveau individuel. Les gros volumes de données n'ont pas besoin d'être envoyés à l'utilisateur, mais uniquement aux centres de données, là où leur réduction et leur traitement aura lieu. D'autre part, il n'est pas nécessaire de centraliser toutes les données dans un seul centre. Mieux vaut les laisser aux mains des experts qui savent les traiter. De plus, en termes de rapidité de calcul, il est plus efficace et moins cher d'avoir l'accès local aux données. De même, il n'est pas raisonnable de maintenir des copies des données dans tous les centres dispersés à la surface du globe, les copies miroirs des données n'étant justifiées que pour éviter une perte des données à long terme, en cas de catastrophe majeure dans l'un des sites. À la place, il faudra développer des stratégies efficaces de transmission de l'information et être capable de fournir des taux de transfert élevés entre les différents centre de données pour contrecarrer la répartition géographique des données. On parle alors de données distribuées. On aboutira à des nœuds névralgiques de données et de ressources spécialisées. Finalement, ne sera envoyé à l'utilisateur que les sous-échantillons nécessaires à son étude, après traitement dans les centres de ressources informatiques, de manière complètement transparente.

Notons que ce changement de paradigme a déjà eu lieu dans le secteur commercial pour résoudre des problèmes similaires. La technologie de l'information trouve un intérêt tout particulier dans les données astrophysiques en raison de leur taille, de leur complexité et de leur nature non-propriétaire. On peut s'attendre à ce que la technologie GRID utilise certains des développements commerciaux.

1.7.4 Nouveaux outils

La stratégie de développement des programmes GRID est basée sur l'interopérabilité, le calcul parallèle et la hiérarchie de services et de données. Historiquement parlant, les archives ont été construites indépendamment les unes des autres pour répondre à des objectifs particuliers. Les différences d'approche dans la gestion des données est particulièrement frappante suivant le domaine de longueur d'onde considéré. Les données sont stockées sur différents supports et ne passent pas par les mêmes logiciels. Il y a cependant des caractéristiques communes mais leur interopérabilité à distance est loin d'être évidente et efficace. Les archives de données actuelles sont pour l'instant comme des îles perdues au milieu de l'océan. Elles n'interagissent pas entre elles et il n'existe pas de service permettant de les comparer, de les corréler, de faire des requêtes simultanées en direction de plusieurs archives. Il faut trouver un moyen de rassembler, d'unifier, de fédérer toutes ces structures disjointes, pour assurer le bon fonctionnement de la technologie GRID et donc de l'Observatoire Virtuel. Le maître mot dans ce domaine est l'interopérabilité.

Pour faire interagir les archives, il est nécessaire de répertorier, classer et indexer de manière uniforme les données. L'Observatoire Virtuel vise à rendre l'utilisation de données astronomiques plus facile grâce à l'introduction de nouveaux outils en amont du projet et la définition de normes et de standards. Tout ceci facilitera grandement l'inclusion des futures nouvelles données. Le développement de toutes ces capacités requiert une interaction et une collaboration proches de la communauté de la technologie de l'information. Cette richesse d'information, afin d'être exploitée au mieux suppose la maîtrise d'outils techniques et méthodologiques faisant appel aux sciences de l'information, à la manipulation et l'accès de données, au *data mining* et à l'analyse statistique. Ces outils, en raison de la taille, la dimension et la complexité des données, seront qualitativement différents de tout ce que l'on connaît actuellement.

La première démonstration de faisabilité de ce vaste projet s'est tenue lors de la conférence de presse à l'observatoire radio de Jodrell Bank près de Manchester, RU, les 20 et 21 janvier 2003. Une version prototype de l'AVO, l'*Astrophysical Virtual Observatory*, y a été présentée. Ce prototype est composé d'un échantillon d'outils consistant en trois composantes logicielles :

– l'interface type Aladin pour la recherche des meta-données (*metabrowser*) et la vi-

nom de colonne	fréquence	nom de colonne	fréquence
Vmag	657	mag1	9
mag	22	MAG	9
V	19	<Vmag>	8
Mag	16	otherMag	6
Vmag2	13	Vder	4
vmag	12	VHB	4
mv	10	mag3	4
Vmax	10	mag4	4
mvis	9	Vt	3
mag2	9	mvB	3

TABLEAU 1.3 : *Statistiques pour l'UCD* PHOT_MAG_V. *Les 20 plus fréquents (parmi 93) noms de colonnes donnés à cet UCD pour les catalogues de la base VizieR. Cet UCD est utilisé dans 926 colonnes, dans 531 catalogues différents (842 tables).*

sualisation ;
- l'ACE ou *Astronomy Catalogue Extractor*, une application SExtractor publiée sous forme de service web ;
- l'utilitaire SED ou *Spectral Energy Distribution* qui permet d'obtenir graphiquement un spectre basse résolution d'un objet astronomique, résultat de la compilation de mesures photmétriques prises dans des bandes différentes par un ou plusieurs instruments/télescopes et donc avec des résolutions, des précisions et des unités différentes.

Ces logiciels ont été réalisés respectivement par le CDS, le consortium Astrogrid et l'ESO. Ces trois modules échangent l'information astronomique par l'intermédiaire du format VOTable. Les VOTables[2] sont des tables de données dans un dialecte XML adapté à l'astronomie. L'innovation majeure dans le cadre du VO consiste en l'inclusion des tags UCDs ou *Unified Content Descriptors* dans ces VOTables. Cette petite révolution est ausi importante que l'introduction du standard FITS (Wells et al. 1981) pour les images il y a de cela plus de 30 ans et constitue une pierre angulaire de l'Observatoire Virtuel international. Les UCDs[3] sont des meta-données représentant la signification sémantique des données physiques, c'est-à-dire des mots-clefs bien définis valables pour toutes les applications et les centres de données. Auparavant, les services de données astronomiques, à la suite d'une requête, renvoyaient une information dont le contenu était lisible par l'homme mais qu'une machine ne pouvait pas interpréter, rendant ainsi difficile l'interaction de services distribués dans le cadre des réseaux et de GRID. En revanche, avec l'introduction du standard international d'échange de données VOTable, les ordinateurs peuvent facilement lire, filtrer et renvoyer l'information à d'autres ordinateurs, permettant ainsi de relier et fédérer les diverses ressources. Les applications et les données sont connectées de façon automatique et intelligible.

Les UCDs (Ortiz et al. 1999, Ortiz & Ochsenbein 2001, Ochsenbein & Ortiz 2001) ont été établis à partir d'une analyse du contenu des catalogues présents dans la base VizieR

[2] http://cdsweb.u-strasbg.fr/doc/VOTable/
[3] http://vizier.u-strasbg.fr/UCD/

(Ochsenbein et al. 2000). Chaque ligne d'un catalogue correspond à une source ou à un enregistrement. Chaque ligne est composée de plusieurs colonnes ou champs, les paramètres qui caractérisent chaque source. Il peut s'agir de son numéro `ID`, de son ascension droite `RA`, de sa déclinaison `Dec` ou encore de sa magnitude `V`. La base VizieR contient plus de 10^5 colonnes de ce type. Certains catalogues possèdent le même paramètre, néanmoins la description syntaxique qui en est faite dans les `ReadMe` peut varier grandement. Ainsi, pour l'UCD `PHOT_MAG_V`, représentant la magnitude visuelle, on trouve 93 noms de colonnes différents : `V`, `Vmag`, `vmag`, `mv`... Des outils permettent d'obtenir une statistique plus précise (Tab. 1.3). Si l'on souhaite traiter automatiquement le contenu de ces colonnes, il faut donc standardiser ces meta-données et uniformiser la description du contenu des catalogues. Cela facilitera également l'interopérabilité. L'analyse des différentes colonnes dans VizieR a conduit à la création de 1 500 termes dédiés à la description de ces colonnes, chacun ayant un sens unique. Ce sont les UCDs, répartis suivant une structure hiérarchique pour faciliter leur lecture. Cet arbre des UCDs permet aussi de sélectionner des catalogues sur la base des UCDs qu'ils contiennent et ensuite de lancer une requête en direction de ces catalogues. Cette description standardisée et concise du contenu sémantique d'un champ permet de comparer automatiquement le contenu de deux catalogues distincts possédant le même UCD et on peut envisager une conversion des unités si nécessaire. Les UCDs trouvent aussi leur application dans Aladin grâce à sa nouvelle fonctionnalité de filtres : ne sont superposées sur les images que les sources satisfaisant aux conditions imposées sur les valeurs prises par les UCDs. Leur visualisation se fait par l'intermédiaire de symboles dont la taille, la forme et la couleur peuvent être contrôlées via les valeurs prises par les UCDs. Notons également la création de l'utilitaire graphique VOPlot[4], résultat d'une collaboration VO-India/CDS, qui permet de visualiser des données astronomiques stockées au format VOTable.

L'objectif scientifique de la démonstration VO, basée sur les données GOODS (*Great Observatories Origins Deep Survey*), était de mettre en évidence les nouvelles capacités de fusion, de manipulation et d'analyse de structures de données hiérarchisées, multi-longueurs d'ondes et multi-instrumentales, comprenant à la fois des catalogues et des images, tout en offrant un large choix dans la configuration de méchanismes de filtres.

Sur la normalisation de la syntaxe utilisée pour échanger de l'information entre les différents services et archives via le protocole HTTP, on notera le développement de la norme ASU, *Astronomical Server URL*[5]. Dans le processus de standardisation, on citera aussi le système de fichiers ascii `ReadMe`[6], qui fournit une description homogène de tous les catalogues présents dans VizieR. Ces fichiers descriptifs sont des meta-données nécessaires à la bonne compréhension du contenu des catalogues de données astronomiques. Il y a aussi le `Bibcode` (Schmitz et al.1995), un code bibliographique tel que 1933C&T....49..194D, qui facilite les échanges entre les bases de données comme NED, SIMBAD et ADS. Au chapitre du développement web, nécessaire à l'unification de l'accès aux bases de données distantes, le CDS a développé le système GLU[7] ou Générateur de Liens Uniformes (Fernique et al. 1998). Ce système a pour but la maintenance des codes informatiques vis-à-vis du codage des URLs *Uniform Resource Locators*. À leur place, des

[4] http://vo.iucaa.ernet.in/~voi/voplot.htm
[5] http://vizier.u-strasbg.fr/doc/asu.html
[6] http://vizier.u-strasbg.fr/doc/catstd.htx
[7] http://simbad.u-strasbg.fr/glu/glu.htx

noms symboliques sont utilisés pour coder les ressources et les URLs sont générés au vol sur la base d'un dictionnaire GLU maintenu à jour permettant de faire le lien entre nom symbolique et URL.

Avec l'explosion du nombre de bases de données et de catalogues en ligne, il est devenu de plus en plus difficile de répertorier toutes les possibilités d'effectuer une requête en position. Par ailleurs, chaque formulaire de requête dispose de sa propre interface et de sa propre syntaxe. Pour pallier ces difficultés, Astrobrowse (Heikkila et al. 1999) se charge de répertorier toutes ces bases de données, de convertir le nom de l'objet fourni par l'utilisateur dans le format et l'époque requis par le service de requête distant et de fournir les résultats. Notons aussi SkyQuery (Budavári et al. 2003), un système de requête distribué pour les catalogues astronomiques. Urania (Boyce 1998, Eichhorn & Kurtz 2002). L'arrangement HTM ou *Hierarchical Triangular Mesh* (Kunszt et al. 2001) permet le partitionnement d'un catalogue de sources ponctuelles en triangles sphériques et la mise en place de requêtes spatiales complexes à diverses résolutions. Également la structure mathématique HEALPix (O'Mullane et al. 2001) ou *Hierarchical Equal Area isoLatitude Pixelisation* dont le but est de faciliter de manière rapide et précise l'analyse statistique et astrophysique d'échantillons massifs de données couvrant tout le ciel.
Plus généralement, les outils permettant de sélectionner l'information parmi les résultats d'une requête doivent être développés : *Topic Maps* (Mahabal et al. 2001), cartes de Kohonen (Poinçot et al. 1998),...

Le produit de tous ces efforts sera une astronomie multi-longueur d'onde, présentant les différentes facettes d'une même réalité. N'oublions pas que l'Observatoire Virtuel devra inclure toutes les données auxiliaires : les données relatives aux télescopes et aux instruments, les données environnementales comme le *seeing*. Mais aussi la dimension temporelle, les simulations, les tables de valeurs des paramètres stellaires, et surtout les services bibliographiques !

Loin d'être une simple collection d'archives, il devra être capable de les faire communiquer entre elles. Le tout transformant les parties, l'Observatoire Virtuel sera un environnement de recherche complet pour l'astronomie, orchestrant tout un ensemble de données massives et complexes, distribuées de par le monde, fournissant les outils d'analyse, de visualisation et de classification nécessaires à leur exploitation. Tout ceci depuis une simple station de travail.

1.7.5 Nouvelle science

Fouiller les archives de ce ciel digital peut être considéré comme réaliser une observation, au même titre qu'à l'aide d'un télescope physique bien réel. Seuls le mode operatoire et la terminologie changent. Le télescope est remplacé par un réseau d'archives et les instruments par des logiciels capables de faire des requêtes sur les données, d'analyser les résultats et de répondre à une question scientifique spécifique posée par l'observateur. L'Observatoire Virtuel doit devenir aussi compétitif que les grands observatoires actuels tels le VLT ou le HST. Lors de sa mise en œuvre il devra, lui aussi, satisfaire à un certain nombre d'exigences scientifiques, sous peine de perdre son intérêt premier : apporter une réponse à un problème scientifique posé. Mais d'ores et déjà, il est riche de promesses.

Les grands relevés produisent des flots de blocs de données par nature cohérents, souvent multi-longueurs d'ondes, multi-époques et couvrant de larges portions du ciel, atteignant une richesse et une profondeur inégalées. La taille, l'uniformité et l'intégrité des relevés permet de les cross-comparer sur le temps, les longueurs d'ondes (l'énergie spectrale) et la résolution. L'exploration de l'Univers peut maintenant se faire de façon panchromatique, sur des domaines spatiaux et temporels plus larges, permettant de s'affranchir des traditionnels biais et limitations observationnels, mais aussi des effets de sélection générés par une coupure en longueur d'onde. De nombreuses questions astrophysiques restées en suspens, qui auraient nécessité de gros volumes de données et donc incompatibles avec les temps d'observation limités disponibles avec les grands télescopes vont maintenant trouver une réponse.

De nouveaux champs d'investigation s'ouvrent (Djorgovski et al. 2001), un nouvel âge de découverte, grâce au *data mining*, à la reconnaissance de formes sophistiquées, aux espaces à plusieurs dimensions et au *clustering*, aux réseaux de neurones, à la statistique et aux identifications croisées multi-longueurs d'ondes à grande échelle (voir Chapitre 2). De nouvelles découvertes se font par l'analyse statistique de grands échantillons de données et la recherche de variations temporelles. S'ensuit alors la découverte d'objets rares tels des supernovae, d'objets non-fixes tels des astéroïdes. La recherche d'objets exceptionnels ou d'événements rares, en dehors des normes, nécessite de scruter des milliards d'observations individuelles pour finalement en trouver une qui sorte de l'ordinaire, qui vienne vérifier une théorie ou défaire les lois de la physique. La compréhension des phénomènes astrophysiques, par nature complexes, requiert l'exploitation de données complexes, nombreuses et multi-dimensionnelles et la création d'outils adaptés, y compris de visualisation. De nombreux outils sont présentés dans le compte-rendu de la conférence *Astronomical Data Analysis* du SPIE, Starck & Murtagh (2001).

Le *data mining*, aussi connu sous le nom de "découverte de connaissances dans les bases de données" ou KDD, *Knowledge Discovery in Databases*, se définit comme suit : extraction non triviale d'information implicite, auparavant inconnue et potentiellement utile, à partir de données (Frawley et al. 1992). Comme son nom l'évoque, le *data mining* s'apparente à l'évaluation des ressources minières d'un nouveau territoire et se divise en trois étapes successives (Wright 2001) :
- exploration, cartographie et relevé ;
- prospection et découverte ;
- exploitation.

Borne (2001) distingue deux modèles de *data mining* : descriptif et prédictif. Il distingue également *event-based mining* de *relationship-based mining*.

L'avènement des grands relevés a d'ores et déjà permis de faire des découvertes majeures : quasar connu le plus distant (Fan et al. 2000) pour le SDSS, découverte de naines brunes avec DENIS (Delfosse 1997), 2MASS et SDSS (Burgasser 2000, Strauss 1999, Dahn 2002), ce qui a mené à la définition des nouveaux types spectraux L et T (Kirkpatrick 1998).

Afin de valoriser et faire fructifier les données, il faut les mettre dans une perspective qui transcende leur destination originelle. On peut partir à la recherche de trésors en fouillant les mines de données et créer de nouvelles données résultantes : les meta-données. Et pour la première fois de l'histoire de l'astronomie, de tels échantillons de données vont pouvoir être comparés aux simulations numériques et l'on pourra produire des analyses statistiques multi-variées complètes.

Et finalement, mais pas des moindres, en plus des nouvelles découvertes, viendront s'ajouter de nouveaux types de découvertes, obtenus par la façon nouvelle de combiner les données, grâce aux portes ouvertes par les nouvelles méthodes d'analyse. Le potentiel scientifique véhiculé par tous ces relevés est énorme. Leur combinaison devrait faire émerger des résultats entièrement nouveaux et inattendus, d'importance majeure, que leur simple étude individuelle n'aurait pas rendus possibles.

Tous ces changements vont changer de façon spectaculaire et irréversible la nature de la recherche en astronomie. Plus important encore, la prise de conscience que l'acquisition, l'organisation, l'analyse et la dissémination des données est un moteur essentiel d'une croissance robuste de la science et de la technologie.

1.7.6 Différents acteurs et partenaires

L'Observatoire Virtuel est une initiative de la communauté astronomique internationale. Pour plus d'informations, on pourra se référer aux compte-rendus des premières conférences qui furent entièrement dédiées à ce sujet et qui marquèrent le lancement de ce projet, à savoir : *Virtual Observatories of the Future* (Brunner et al. 2001a) tenue à Caltech, Pasadena en Californie en juin 2000 et *Mining the Sky* (Banday et al. 2001) tenue à Garching en Allemagne en août 2000.

L'idée d'un ciel électronique n'a rien de nouveau. Le CDS, créé en 1972, a une longue expérience dans ce domaine. Son ancien directeur, C. Jaschek en avait déjà une conscience aiguë comme le témoigne son livre *Data in Astronomy* (1989) qui donne une vision des prémices de l'Observatoire Virtuel. Des précurseurs ont vu le jour avec les projets NASA ADS (Good 1992, Kurtz et al. 2000), ESA ESIS (Giommi & Ansari 1994), mais qui ne remplissaient pas les attentes réelles de l'Observatoire Virtuel. Le CDS travaille depuis plus de 10 ans dans cette direction et son travail de pionnier est la pierre d'angle nécessaire à la construction de l'Observatoire Virtuel. Les archives et les outils de recherche deviennent de plus en plus populaires au sein de la communauté astronomique, comme le montrent les statistiques sur les services du CDS au début de l'année 2003. Il y a 13 000 requêtes externes par jour sur SIMBAD, 8 500 sur VizieR et 2 500 sur Aladin, soit 24 000 requêtes externes par jour, issues de plusieurs milliers de nœuds. Le nombre d'accès privilégiés qui n'utilisent pas le service de pages web comme par exemple le projet XMM-Newton, est de 100 000 par jour. Le nombre de copies de catalogues dans VizieR en un an est de 500 Go, soit 30 fois le volume total de l'archive qui contient 15 Go.

L'Observatoire Virtuel est formalisé en Europe sous le nom d'*Astrophysical Virtual Observatory* ou AVO (Quinn et al. 2002). C'est un projet soutenu par la Commission Européenne de la recherche dans le cadre FP5 et piloté par l'ESO (P. Quinn) en partenariat avec 5 autres organisations : le ST-ECF (P. Benvenuti), ASTROGRID (A. Lawrence), le CDS (G. Genova), TERAPIX (Y. Mellier) et l'Observatoire de Jodrell Bank (P. Diamond). Le projet AVO[8] a pu naître grâce à un réseau d'organisations astronomiques européennes regroupées sous le nom d'OPTICON ou *OPTical Infrared COordination Network for astronomy* (Gilmore 2001). Avec un budget de 7.2 millions d'euros, la phase A (2002-2004) de ce programme consiste à développer trois axes majeurs, à savoir : *i)* poser les exigences scientifiques requises pour l'AVO, grâce notamment à l'expérience acquise dans le projet ASTROVIRTEL (Pierfederici et al. 2001) *ii)* définir des standards astronomiques pour rendre possible l'interopérabilité *iii)* évaluer l'implication dans la technologie GRID pour un bon fonctionnement de l'AVO.

De plus, un consortium, formé autour du noyau des membres de l'AVO et composé des principaux centres de données européens (au total une vingtaine de sites), va soumettre un projet dans le cadre FP6 pour la création d'un Observatoire Virtuel pan-européen d'ici la fin 2007. Ceci afin de réunir, intégrer et fédérer tous les acteurs au sein d'un même EURO VO, pour pouvoir concentrer les efforts et faciliter le développement et la distribution d'outils communs. Cette entreprise prend le nom de EVOE, *Enabling a Virtual Observatory for Europe* et devrait permettre de mener à bien la phase B du plan AVO (2005+). On peut espérer que la construction de l'Observatoire Virtuel rendra possible la fortification et l'expansion d'un véritable espace européen de la recherche en ouvrant l'accès à des données précieuses aux états associés mais non-membres.

Il existe un projet similaire du côté américain, appelé NVO ou *National Virtual Observatory*[9], avec un budget alloué NSF de 60 millions de dollars. Ce projet (*White Paper*, Brunner et al. 2001b) est considéré comme la première des priorités au chapitre des petites initiatives dans le cadre du plan décennal *Astronomy & Astrophysics in the New Millenium*.

L'Observatoire Virtuel a pour but la collaboration globale des communautés astronomiques présentes en Europe, en Amérique du Nord et du Sud, en Asie et en Australie sous l'égide de la nouvellement créée alliance internationale des Observatoires Virtuels, l'IVOA ou *International Virtual Observatory Alliance*[10].

Il est nécessaire de construire l'Observatoire Virtuel dès maintenant. Plus l'on attendra, plus il sera difficile et long d'inclure les données a posteriori et moins la pertinence qu'il y a à combiner ces données entre elles sera justifiée, car les résultats de ces combinaisons devront servir à guider les observations à venir. Et puis il serait dommage de laisser inexplorées et inexploitées ces immenses gisements de données. Plus l'Observatoire Virtuel sera opérationnel tôt, plus cela contribuera à son succès. Il faudra s'atta-

[8] http://www.euro-vo.org
[9] http://us-vo.org/
[10] http://www.ivoa.net

Acronyme	Observatoire Virtuel
ASTROGRID	Grande-Bretagne
Aus-VO	Australie
AVO	Europe
China-VO	Chine
CVO	Canada
GAVO	Allemagne
JVO	Japon
KVO	Corée
NVO	États-Unis
RVO	Russie
VO India	Inde

TABLEAU 1.4 : *Liste des organisations membres de l'IVOA.*

cher à reconnaître la motivation essentiellement scientifique derrière la mise en œuvre de cet Observatoire Virtuel de l'ère du tout technologique. Il faudra maximiser la participation de la communauté scientifique dans cet effort, rendu d'autant plus probable que l'implémentation de services et outils fonctionnels se fera rapidement et régulièrement. L'accès aux nouvelles fonctionnalités de l'Observatoire Virtuel ne se fera pas brutalement mais régulièrement.

On comprend dès lors son importance en matière d'éducation scientifique en direction du public et des astronomes amateurs. L'empire croissant du WWW a ouvert les portes de la science à de nouvelles audiences. Le très populaire projet SETI@home[11] a réuni 2.3 millions de personnes dans un même projet en partageant leur puissance informatique (Sullivan et al. 1997). Le public veut de plus en plus, non seulement comprendre, mais s'impliquer dans la recherche. Il pourra créer son propre télescope, définir son propre programme de recherche et même faire des découvertes qui auraient autrement été impossible sans l'aide de cet effort distribué d'explorateurs de données enthousiastes. Le retour scientifique et culturel d'une entreprise telle que l'Observatoire Virtuel sera d'autant plus important que le public sera partie prenante. Le succès d'une telle entreprise se mesurera à l'augmentation de la productivité scientifique, pour tous les astronomes et pas à un petit groupe restreint d'utilisateurs. Des projets qui auraient pris des mois, des années à être mené à bien s'accompliront en quelques minutes, quelques heures. L'Observatoire Virtuel sera considéré comme un élément essentiel à la fois de la recherche scientifique et de l'éducation. Il devra se situer au centre des dernières avancées scientifiques et permettre de minimiser l'écart de connaissances entre la recherche de pointe et l'enseignement, afin d'intégrer la recherche dans l'espace public.

D'un point de vue budgétaire, des grands projets tels le VLT ou ALMA engloutissent de grandes parts du budget total de la recherche sur des périodes de l'ordre de la dizaine d'années. Le retour scientifique de tels investissements doit être maximisé. Ainsi, pour le VLT, 20% du coût du projet sert à assurer le bon fonctionnement des archives. Les grands observatoires ont investi du temps de télescope pour produire des archives avec un potentiel suffisant pour servir une grande variété de programmes scientifiques. L'Observatoire Virtuel est rendu possible grâce aux progrès de la technologie mais il est motivé essen-

[11] http://setiathome.ssl.berkeley.edu/

tiellement par la science. Sa structure et son évolution sont guidées par les besoins de la communauté scientifique.

Notons que ces considérations sont très proches des défis rencontrés par des disciplines connexes telles les sciences de la terre mais aussi les sciences physiques, biologiques, sociales et médicales. La taille du génôme humain (GenBank) s'étend sur environ 3 Go. Les outils et techniques développées pour gérer les masses de données astronomiques trouveront clairement des applications dans d'autres domaines. Mais tout cela ne se fera pas sans le développement de nouveaux outils et de nouvelles structures institutionnelles qui pourront consolider toutes ces bases de données et catalogues disparates, autoriser leur accès, définir une législation et placer les outils d'analyse appropriés dans les mains d'une communauté de scientifiques imaginatifs.

1.7.7 Projet ASTROVIRTEL

Dans la progression vers l'AVO, l'interaction entre les chercheurs et les archives est crucial pour garder le développement d'outils logiciels sophistiqués proche des exigences scientifiques actuelles, comme l'a démontré le projet ASTROVIRTEL[12] (outil Querator, Pierfederici 2001).

ASTROVIRTEL signifie *Accessing Astronomical Archives as Virtual Telescopes* ou encore utilisation des archives astronomiques en tant que télescopes virtuels. Ce projet est financé par la Commission Européenne sous contrat FP5 No. HPRI-CT-1999-00081. C'est une action du programme *Improving Human Potential* qui vise l'accès aux infrastructures de recherche. ASTROVIRTEL est géré par le ST-ECF sous l'égide de l'ESA et l'ESO.

Le projet ASTROVIRTEL a pour but l'exploitation des archives ST-ECF/ESO. Ces archives contiennent plus de 10 To d'informations scientifiques obtenues avec le HST de l'ESA/NASA et le VLT, NTT et WFI de l'ESO. Le taux de croissance de ces données est de 4.5 To par an, allant même jusqu'à 6 To pour ces deux dernières années. On s'attend même à ce que ces grands ensembles de données s'accumulent encore plus rapidement au cours des dix prochaines années d'observation HST. Cette collection de données peut être vue comme un Observatoire Virtuel, aussi bien capable de répondre aux mêmes requêtes d'observation que s'il s'agissait d'un véritable télescope. L'intérêt du projet ASTROVIRTEL est donc de valoriser et augmenter le retour scientifique de ces archives, par le développement d'outils de *data mining* adaptés, sur la base d'un projet de recherche bien défini. Une partie de cette thèse a bénéficié du soutien ASTROVIRTEL (Chapitre 2).

12 http://www.stecf.org/astrovirtel