

国立天文台 データベース天文学推進室
検討報告書

Japanese Virtual Observatory の構築に向けて

平成 14 年 3 月 7 日

目 次

概 要	1
1. 天文データベースの現状	2
1.1 世界の天文データベースの現状	2
1.2 日本の天文データベースの現状	3
1.3 アーカイブを用いた天文学研究の現状	4
2. 既存の天文データベース・データアーカイブの分析	5
2.1 役割	5
2.2 形態	5
2.3 機能	6
2.4 問題点	6
3. 仮想天文台(Virtual Observatory)の構築	6
3.1 天文学としての要請	6
3.2 大望遠鏡や専用望遠鏡による大規模サーベイ観測による研究	7
3.3 多波長観測データを利用する研究	7
3.4 計算機性能の加速度的向上	8
3.5 観測データの爆発的増加と人間の処理能力の限界	9
3.6 高速ネットワークの普及	9
3.7 情報学研究, 計算機科学, 特にデータベース技術の進展	10
3.8 JVO(Japanese Virtual Observatory)の構築: 望遠鏡や観測装置の抽象化	10
3.9 世界における VO に向けた流れ	12
3.10 JVO の他にない特色	13
4. JVO に必要な機能	13
4.1 データフォーマットの共通化機能	13
4.2 論理的数値宇宙の構築機能	14
4.3 数値宇宙に対する多様な検索機能	14
4.4 大量データの解析機能	15
4.5 大規模多次元データの可視化機能	15
4.6 より高度な解析機能	15
4.7 観測所(実望遠鏡)との連携と役割分担	15

5. JVO の必要技術	15
5.1 基盤技術	15
5.2 解析の自動化	17
5.3 Move Not the Data But the Programs and the Results	17
5.4 データの質の保証：望遠鏡との連携	18
5.5 多様な解析	18
5.6 可視化	18
6. JVO システムの構築	18
6.1 JVO の機能概要	18
6.2 開発計画	21
6.3 開発体制	23
6.4 プロトタイプ作成計画	24
6.5 利用法のイメージ	25
7. その先にあるもの	26
7.1 VO は全ての「実観測」のインターフェース	26
7.2 世界の VO との連携	26
7.3 大学・大学院教育への波及効果	26
7.4 小中高における理科教育への波及効果	27
7.5 Public Outreach	27
8. データベース天文学の世界	27

概 要

近年の望遠鏡技術，検出器技術の向上により，すばる望遠鏡などの大型望遠鏡による高精度の観測データ，スローン・デジタル・スカイ・サーベイ（SDSS）などの専用望遠鏡によるサーベイデータなど，我々は良質で大規模なデータが大量に生み出される時代を迎えている。しかし，これらのデータは大規模であるがゆえに，そこに含まれる情報をあますところなく利用し最大限の科学的成果を挙げるためには従来の解析方法では不十分であると考えられる。また，単一の望遠鏡のデータだけでなく複数の波長のデータを使った統計的な研究の重要性は多くの人の認めるところであるが，実際にそのような研究を行うには多大な労力が必要であるのが現状である。

一方，計算機の能力，ネットワークのスピードなどは加速度的に向上しており，計算機の利用方法も従来とは異なる形態が模索されている。情報学，計算機科学の分野で研究が進められており，代表的なものとして分散した計算機リソースを互いに結合させる GRID 技術と大規模なデータから新しい発見をおこなうデータマイニング技術，データベース技術である。

この天文学的なパラダイム転換と計算機利用のパラダイム変換をうまく結合することで，我々は計算機の中にデジタル形式の仮想的な宇宙を作り，それを様々な角度から解析（観測）する「仮想天文台(Virtual Observatory)」を立ち上げることが 21 世紀の天文学の重要な一面となるであろうと考えるに至り，ここに Japanese Virtual Observatory (JVO) の開発計画を提示するものである。

本報告書では，このような開発計画にいたった背景についての詳細，JVO に求められる機能，その技術的な実現可能性，具体的な開発計画，開発体制等について述べる。

1. 天文データベースの現状

データベース天文学の必要性を理解するためには、まず、世界や日本の天文データベースの現状とその問題点を把握しておくことが必要である。ここでは、現状のデータベースについて世界と日本に分けて概観する。

1.1 世界の天文データベースの現状

1.1.1 衛星観測によるデータベース

世界では早くから衛星による観測データのデータベース化が進んできた。衛星による観測は、衛星打ち上げ費用が極めて高額であることを反映し、その多くが観測データ公開を前提として取得される。そしてこれらのデータ公開は、下記に述べるデータセンターにより電子的(Web や ftp を利用)に公開される。以下に主なものを紹介する。

1.1.1.1 国立宇宙科学データセンター(NASA National Space Science Data Center, NASA-NSSDC)では COBE, IUE のデータ公開を行うと共に、各種天文カタログの収集・頒布を後述する CDS と協力して進めている。

1.1.1.2 ゴダード宇宙飛行センター(NASA Goddard Space Flight Center, NASA-GSFC)では、高エネルギー天体物理研究所との協力のもとに、高エネルギー天体物理学アーカイブ研究センター(High Energy Astrophysics Science Archive Research Center, HEASARC)を運営しており、ほとんど全ての高エネルギー天文衛星のデータを保有・頒布している。

1.1.1.3 赤外線観測処理解析センター(Infrared Processing and Analysis Center, IPAC)では IRAS 衛星による観測データの収集・標準化・頒布を行っている。

1.1.1.4 スペーステレスコープ科学研究所(Space Telescope Science Institute, STScI)では、ハッブル宇宙望遠鏡(HST)に関する運営の一切を取り仕切っており、HST の観測データの取得・頒布の中心となっている。

1.1.1.5 カナダ天文学データセンター(Canadian Astronomy Data Center, CADC)はドミニオン天体物理観測所に所属し、カナダにおける HST 観測データの頒布センターとして設立された。その後カナダ・フランス・ハワイ望遠鏡のデータセンターとしての役割をも担っている。

1.1.1.6 ラザフォード・アップルトン研究所(Rutherford Appleton Laboratory, RAL)は英国における STARLINK データサービスの中心となっている。ここにある Space Data Center

は、IRAS、ROSAT、IUEなどのデータの提供、天文カタログ STADAT や SMM のデータセンターとなっている。また、最近観測を終えた ISO のセンターにもなっている。

1.1.2 地上観測によるデータベース

衛星データのデータベース化に較べて、地上観測データのデータベース化は遅れていると言えるだろう。しかし、古くから写真乾板で撮影した画像は、例えば、Palomar Chart として利用されてきた。そして最近では、これをデジタル化した DPOSS が公開され、さらに、全天を CCD でサーベイする SDSS 計画では当初から観測データをデータベース化することとしている。また、電波天文では古くから英国ケンブリッジ大学によるキューサーのカタログ(所謂 3C カタログなど)、水素原子の 21cm 線やホルムアルデヒドのスペクトル線を用いたサーベイデータ (Green Bank Surveys) が作成され、赤外線においても $2\mu\text{m}$ におけるサーベイを行った AFGL カタログが作成されている。しかし、電波サーベイや赤外線サーベイの多くは、オンラインで検索できるようなしくみを当初から備えているものではなかった。

1.1.3 文献データベース

文献データベースとして最も有名、かつ、役立っているものは ADS (Astrophysics Data System) であろう。上述した IPAC が運営しているものである。

1.1.4 世界のデータセンター

上で述べた各種データベースは、様々なデータセンターによって開発・運用されている。最も有名なものとしてストラスブール天文学データセンター (Centre de Données Astronomiques de Strasbourg, CDS) があるが、これは 1972 年以来国際的な天文学データセンターとして天体カタログの収集・頒布の仕事を行ってきている。また、恒星の位置等を与えるデータベースである SIMBAD を運用しているのもこのセンターである。また、上述のように NASA の配下には複数のデータセンターがある。

1.2 日本の天文データベースの現状

1.2.1 国立天文台の観測データベース

わが国における天文データベースとしては、まず、国立天文台がその所有する望遠鏡用のものが存在する。岡山天体物理観測所と木曾観測所の観測データを蓄積した MOKA (Mitaka-Okayama-Kiso Archive)、野辺山宇宙電波観測所の 45m 大型ミリ波望遠鏡と 6 素子ミリ波干渉計の観測データを蓄積した nrodb、そしてハワイ観測所のすばる望遠鏡の観測データを蓄積する STARS (Subaru Telescope ARchival System) および MOKA を発展させてすばる望遠鏡のデータにもアクセスできる機能をもたせた SMOKA である。また、太陽関係では、

野辺山ラジオヘリオグラフの観測データは野辺山太陽電波観測所に全て蓄積されて全世界の太陽研究者が利用しているし、太陽活動世界資料解析センターは太陽に関する様々なデータを世界に供給するセンターとして機能している。

1.2.2 宇宙科学研究所の観測データベース

宇宙科学研究所には、観測データやテレメトリデータのデータベース化及びその運用を行うための組織(宇宙科学企画情報解析センター：通称 PLAIN センター)があり、ASCA、Yohkoh、Geotail、Akebono や BeppoSAX、ROSAT といった衛星のデータ等をオンラインで公開すると共に、国立天文台と共同で開発した多波長画像表示システムである jMAISON を運用している。

1.2.3 海外望遠鏡群のデータベース

また、国立天文台天文学データ解析計算センターでは、海外の研究機関が開発・運用している観測データベースのミラーを行っている。その中には、HST、DSS、SDSS、IUE が含まれ、いずれもオンラインで検索することができる。

1.2.4 カタログデータベース

国立天文台天文学データ解析計算センターでは従来より、論文などで発表されたカタログの収集・公開を行ってきた。これらはいずれもテキスト形式のファイルとして web からダウンロードできるようになっている。

1.2.5 文献データベース

国立天文台天文学データ解析計算センターでは、また、電子化された天文関係の論文へのアクセスを容易にするため、NASA ADS や各種雑誌(ApJ、AJ、A&Ap)のミラーを行っている。

1.3 アーカイブデータを用いた天文学研究の現状

先に述べたように、衛星搭載の天文観測装置の場合、データを公開して広く天文コミュニティに利用してもらうことを前提としている。実際、X線やγ線天文学といった衛星からではないとデータが取れない波長領域の場合、観測データから装置の特性を除く一次較正処理まで観測運用チームが行い、また、データの品質保証をもしているために、アーカイブされたデータは広く利用され、多くの天文学的成果を生み出している。その良い例として、わが国が打ち上げた ASCA 衛星が挙げられるし、赤外線衛星(IRAS)のカタログ(IRAS Point Source Catalogue)が広く利用されていたことは記憶に新しい。また、ハッブル宇宙望遠鏡(HST)のデータがアーカイブされており、天文研究者はアーカイブの中から必要な観測データを探し出して研究を行っていることも良く知られている。

図1にHSTのアーカイブデータの利用率が年々上昇していることを表すデータを示す。特に1997年の後半以降は、望遠鏡によるデータ取得量の4-5倍にも及ぶ観測データがアーカイブからダウンロードされ、天文学の研究のために利用されていることが分る。

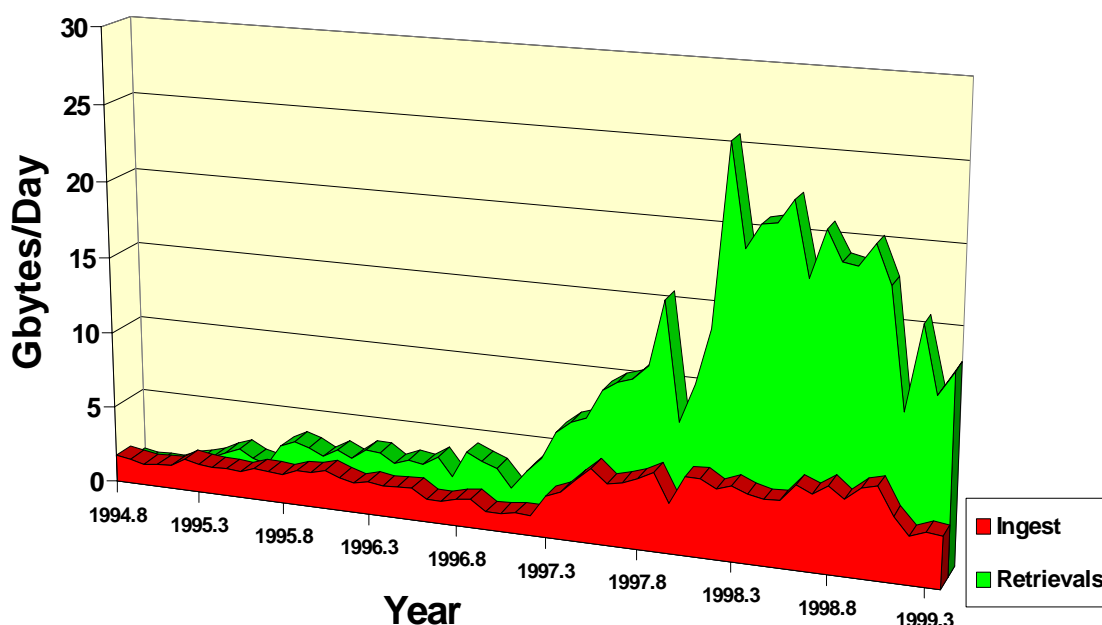


図1 HSTによる取得データ(赤)とアーカイブからダウンロードされたデータ(緑)の量の年度毎の変化

2. 既存の天文データベース・データアーカイブの分析

2.1 役割：何のために、誰が作り、誰が運用しているか

衛星を用いて取得した観測データは、衛星の運用チームが責任をもってDBを作成・運営しており、巨額の費用をかけて打ち上げの前提通り天文学研究者に公開している。地上データの場合でもDB作成等は観測所が行っているが、基本的には「その装置の利用者もしくはその装置からのデータ特性に詳しい研究者」による利用を想定している。

2.2 形態：どんなデータがどんな形(カタログとか画像とか)で

衛星データの場合は運用チームが一次較正まで行ったデータ(画像が主)がDB化されている。IRASを始めとする多くのサーベイ観測データの場合は、カタログ化も運用チームが行って公開している。特に2MASS, POSS/DSS/DSS2といった全天サーベイの結果得られたデータは、finding chartとしての役割が非常に大きい。Hipparcos衛星のデータ(カタログ)が活用されていることは、多くの天文研究者の認めるところとなっている。電波天文

では周波数方向の分解能が高いのでスペクトルデータでの公開も行われているが、干渉計は基本的に画像（UV データでも）の DB となっている。

2.3 機能：何が出来るか（検索機能，データ取得機能など）

座標，観測時，観測装置，観測波長・周波数などによる検索が一般的に可能。多くはオンラインでダウンロード可能である。

2.4 現状の天文データアーカイブの問題点

一方，Web を通じて公開されている DB はその URL を知らないとアクセスできず，検索方法などのユーザーインターフェースも DB 毎に少しずつ異なる。また，波長や装置が異なると観測データの形式が異なる。一般に FITS (Flexible Image Transfer System) 形式が採用されているが，FITS 形式とは 1 レコードを 2880 バイトで構成し，ヘッダ部分に続いてデータ部分が続くということと，ヘッダ情報の書き方を規定しているだけである。従って，各 DB のヘッダ情報は装置ごとに大きく異なる。このため，それらを解釈できる解析ソフトがデータ毎に必要となる。アーカイブから取得したデータを解析する際に慣れ親しんだ解析ソフトを用いるのであれば問題ないが，他の波長の観測データを解析する場合などには必要な解析ソフトを入手するのに苦労する。さらに，解析ソフトやダウンロードしたデータをストアする計算機の管理も自分でやらないといけない。特に，キャリブレーションがきちんとできていない場合も多いので，観測装置の「くせ」を熟知したユーザーでなければ，データ解析に耐えるだけのデータの質を確保することが難しいなどの問題がある。

3. 高速ネットワーク上における仮想天文台の構築

3.1 天文学としての要請

天文学では常に見えないものを見ようと，新しい波長域の開拓，観測装置や解析方法の工夫を重ねてきた。そして，歴史ある可視域での観測に加えて，電波天文学，赤外線天文学，X 線天文学等の発展が促されて，低温の星間ガス，高温の星間ガス，そしてガスから星，惑星の形成とその輪廻が解明されてきた。

その一方で，宇宙の開闢直後の物理，その直後に起きたとされる銀河の形成，大規模構造の原因，さらに今後の宇宙の運命を定める宇宙の平均密度などについては分らない点が多い。これら現代天文学最前線の謎を解明するためには，従来のような観測装置の高感度化だけではなく，多くの天体の観測を行ってその統計的振る舞いを検討することが肝要である。特に，広い領域を均質に観測するサーベイは観測的宇宙論の進展にとって重要な鍵となる。

また，このような統計的扱いを行うと，性質が分っている天体とは異なる振る舞いを見せる天体データも数多く見えてくる。これらの「未知天体」は，多くの場合これまで知ら

れていなかった天体现象を反映しており，歴史を振り返ってみても天文学の発展のドライビングフォースとなってきた。

3.2 大望遠鏡や専用望遠鏡による大規模サーベイ観測による研究

高感度かつ均質な観測データが取得可能な時代となった。広く見ることによって初めて見えてくる世界がある（cf. 宇宙の大規模構造）。また，天体现象は一般に人間の生活時間に較べて非常にゆっくりとした変化しかしないので，様々な段階にある多くの天体の相互比較をすることにより，天体现象の正しい理解が可能となる。データの量が質に変化するのである。これを目指して天文学ではできるだけ多くのデータを取得しようと努力を重ねてきたのである。

一方，望遠鏡で観測する領域（波長方向も含めて）が広がると，研究目的以外の天体も観測されることになる。多くの場合は，観測提案をした研究者はこれらの天体データの解析は行わない。しかしながら，他の研究者にとっては「せっかく取得した観測データを活用したい」と思う研究対象であるかもしれないのである。即ち，**残存情報は不要なのではなく「宝の山」**なのである。1.3 で示したように，アーカイブを利用した研究が近年激増していることから，**天文コミュニティとして「宝の山」を共有したいという要望が強いこと**が理解されよう。

3.3 多波長観測データを利用する研究

天体は一般的に広範囲な電磁波スペクトルを持つことは良く知られた事実である。実際，恒星や銀河の研究では「色」によって表現される物理量を用いた様々な研究が行われてきた。これから容易に類推できるように，広い波長帯域のデータを見ないと天体の「正しい姿」が見えないのである。その例として図2にカニ星雲を様々な波長で撮影した画像を示す。



図2 カニ星雲の画像
(左上) X線，(右上) 可視光
(左下) 赤外，(右下) 電波

3.4 計算機性能の加速度的向上（微細加工技術の進展など）

天体物理学の歴史は、写真乾板以来、観測装置の高感度化や大型化を通して、できるだけ多くの観測データを取得しようとした歴史とも言える。これまでは増大したデータを、(1)ハードディスクやテープ装置の集積度が足りないために蓄積する場所の確保に苦勞、(2)解析用計算機的能力（CPU 速度やメモリ量）の制限のために大量のデータ処理を行うために時間がかかりすぎた、(3)ネットワークの転送速度の制限により大量のデータを輸送するためにはテープで送ることが主体、という状態であった。つまりその宝を取っておくだけの計算機資源が確保されないために、せっかくの宝も死蔵されてしまう運命を辿っていた。

最近ではデジタル技術の導入により、観測データの巨大化は驚くべき速さで進んでいる。半導体の稠密度の向上は、経験則である Moore の法則「半導体の集積密度は 18 ヶ月で倍増する」に沿って進んできた。その最新技術を用いた天体観測用検出器を構成する CCD 次元素子は、すばる望遠鏡の Suprime-Cam の場合、 $2k \times 4k (x 10)$ となっている。つい数年前はこのサイズの CCD 素子は「最先端」であったものが、現在では「普通」のものなのである（図 3 参照）。この結果、生み出される観測データの規模も巨大化し、先の Suprime-Cam の場合、毎夜最大 15 GByte にも達する。また現在国立天文台が欧米と協同して建設を進めている ALMA 望遠鏡の場合、60 台のアンテナからの相関データを超高速に処理するために、データ生産量は年間~PByte ($1\text{PByte} = 1 \times 10^{15} \text{Byte}$) にも達する。

一方、この Moore の法則は計算機の心臓部である CPU の性能向上についても成り立つ。さらに最近の PC やワークステーションに接続するハードディスクやテープ装置の容量の増加、低価格化には目を見張るものがある。即ち、半導体微細加工技術の発展の結果、観測データの取得装置、蓄積装置、処理装置のいずれも急速に巨大化 / 高性能化していることが分る。

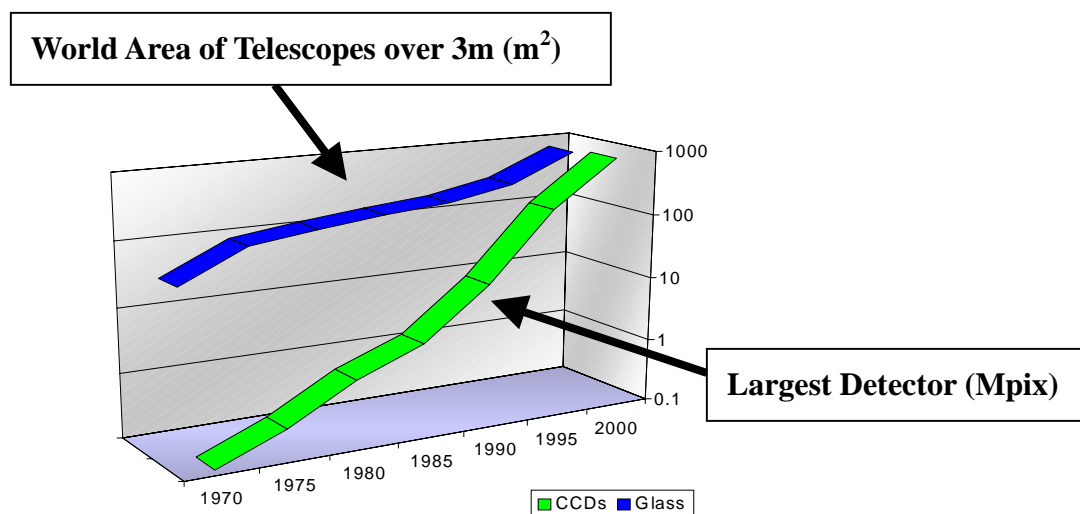


図 3 CCD 素子の発展と望遠鏡集光力の増大

3.5 観測データの爆発的増加と人間の処理能力の限界

一方、観測データを処理する際に最終判断を下す人間の能力には限界がある。人間の考え方や行動パターンは、過去の経験等に照らして決定されるため、大きな慣性を持つ。つまり、3.4 に述べたようなデータ生産率の爆発的増大が生じるにも関わらず、従来通りのデータ処理を行おうとする。

容易に理解できるように、従来通りのインタラクティブなデータ処理方法を取っていたのでは、「データを取得する時間よりも、データを処理する時間のほうが圧倒的にかかる」という事態に陥ってしまう。新世代望遠鏡による爆発的データ生産に対応するためには、人間の情報処理能力で対応できる量までデータを reduce しなければならない。そこで、従来の方法論とは異なるデータ処理方式を導入せざるを得ない。

3.6 高速ネットワークの普及

わが国におけるネットワークの高速化は他国に較べて遅れている。しかし 2002 年初めには国立天文台も 10Gbps の速度を持つ SuperSINET に接続され、それまでに較べて対外接続速度は 1000 倍にもなった。

インターネットの帯域幅の拡大については、半導体の集積度の向上を示す Moore の法則に対応する Nielsen の法則　ハイエンド・ユーザの接続速度は毎年 50% 増大する　で記述される。

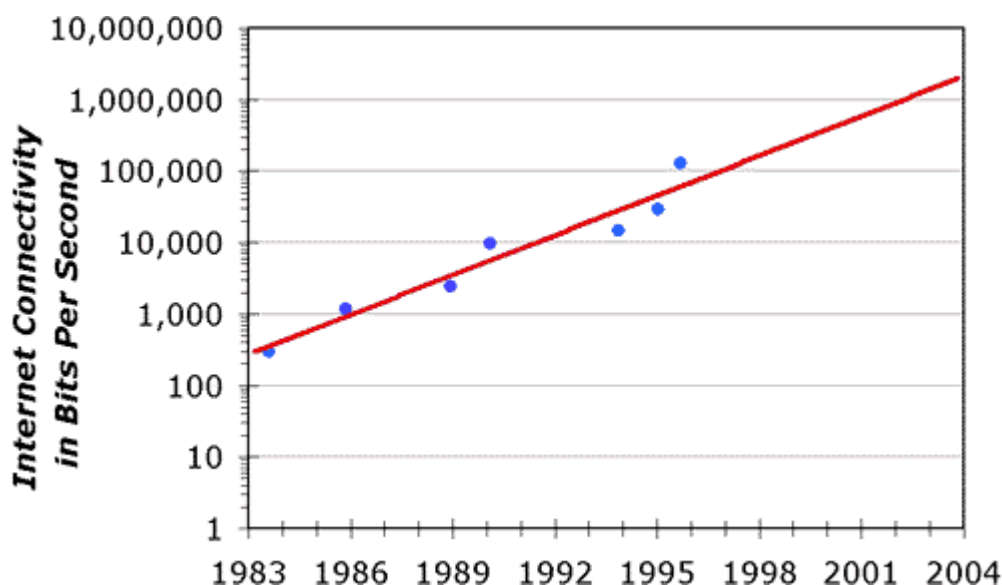


図4 Nielsen の法則。横軸は年（西暦）である。

Nielsen の法則が示すネットワーク転送速度の急激な高速化（同時に低価格化）は、大量

のデータの転送が極めて容易になること、また、遠隔地の計算機資源をあたかも local な計算機資源として利用することが容易になる可能性を示している。

即ち、**計算機利用に関するパラダイム転換が必然**であることを意味している。

3.7 情報学研究，計算機科学，特にデータベース技術の進展

前節で述べたパラダイム転換のためには，ハードウェアとして利用可能な計算機資源の持つポテンシャルを十分に活用するための技術（ソフトウェア，ミドルウェア等）が必須である。情報学，計算機科学は最近の計算機性能の向上にも貢献すると同時に，その利用技術の研究も進めている。

Web の導入がインターネット利用を学術目的から商用目的に拡大した事実は記憶に新しい。Web 上に散在する各種情報を自動的に収集してデータベース化する技術により，私たちはインターネット空間を，図書館代わりに利用できるようになった。これが，電子図書館や電子出版を進める原動力になったとも言える。ネットワーク（LAN）上に分散した計算機資源を透過的に利用するための CORBA (Common Object Request Broker Architecture) 技術はすばる望遠鏡システムにも導入され，DASH システムとして稼動を開始している。この考えをさらに広域ネットワークにも広げた GRID 技術が注目されている。遠隔地にあるスーパーコンピュータを GRID で結合して巨大スーパーコンピュータとして使おうという Grid Computing，遠隔地にある巨大データ資源を透過的に使おうという Data Grid の研究が世界的に広がっている。例えば，CERN（欧州加速器機構）における高エネルギー加速器実験である ATLAS 計画の中核を担うミドルウェアとして研究されている GriPhyN や英国の天文 Data Grid である AstroGrid などがある。

さらにデータベース技術の研究においても OODB (Object Oriented Data Base) 技術が開発され，これまで多用されていた RDB (Relational Data Base) よりもデータベース項目の拡張が容易であるという点から，利用が拡大している。また，これらを用いた VLDB (Very Large Data Base) の構築においては毎年世界的規模の研究会が開催されている。多量のデータの中から「重要な知見」を見出すための技術であるデータマイニングも情報学における研究フロンティアの一つである。決定木(回帰木)法，ニューラルネット法，Memory-Based Reasoning 法など，様々な手法が研究・開発されており，これらをビジネスに応用している例も多い。

3.8 JVO (Japanese Virtual Observatory)の構築：望遠鏡，観測装置の抽象化

さて観測の手順を思い起こしてみる。「実観測」は計算機コンソールに向かって，観測天体の座標，観測装置の設定パラメータ，開始時間などを入力して実行する。装置が取得したデータはモニターなどで監視でき，観測が正常に終了すると，ヘッダのついた観測データファイルが（観測データベースシステムを経由して）解析用計算機に送られる。研究者はこれを解析して論文とする。

一方「データベース観測」も計算機コンソールに向かって，観測天体の座標，観測装置

の設定パラメータ, などの検索パラメータを入力して DB 検索を実行する。データアーカイブから取得したデータは早見画像などで大雑把なチェックができ, 検索が正常に終了すると, ヘッダのついた観測データファイルが (検索データベースシステムを経由して) 解析用計算機に送られる。研究者はこれを解析して論文とする。

この手順を思い起こしてみれば, 「実観測」でも「データベース観測」でも得られるデータは本質的には変わらない。そして, 「DB/DA」, 「解析計算機」は必ずしも研究者の手元にある必要はなく, 高速ネットワークを通じて遠隔地にあるものを利用することで十分である。つまり, 「データベース観測」の場合は, 高速ネットワークに接続された研究室の端末が先に述べた「コンソール」となり, 様々な命令を入力することで「観測」が可能となるのである。言い換えると, これは高速ネットワーク上に建設された「仮想望遠鏡」 Virtual Observatory (VO) を利用することに等しく, 観測そのものをこのように抽象化することができるのである。

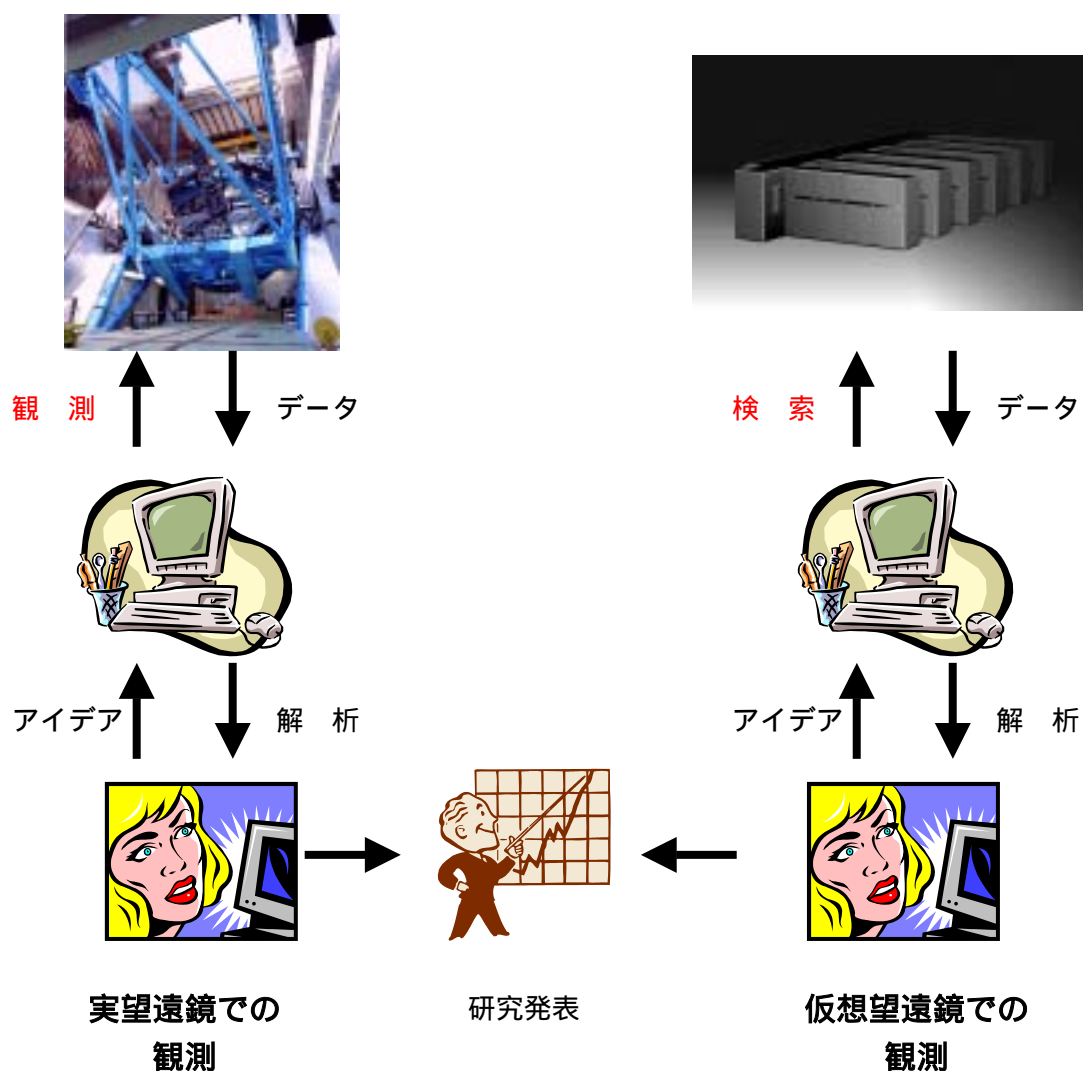


図 5. 実望遠鏡と仮想望遠鏡による観測の比較。「観測」と「検索」だけが異なる。

「実観測」の場合、これまでに観測が行われていないデータを望遠鏡で取得すると考えることもできる。「データベース観測」の場合に VO にデータがない場合は、自動的に観測手順（実望遠鏡の操作命令列）を生成して VO が「実観測」を行えばよい。研究者は、研究目的をクリアにしたプロポーザルを作成し、それに対応するデータを得ればよいのである。

3.9 世界における仮想観測所（Virtual Observatory）建設に向けての流れ

世界では欧米を中心として、1.1～1.3 に述べたことを背景として、仮想観測所「建設」計画が進みだしている。ここでは代表的なものとして米国の NVO (National Virtual Observatory) と ESO の AVO (Astrophysical Virtual Observatory) について概観する。他にも、英国の AstroGrid、オーストラリアの AVO (Australian Virtual Observatory) 計画などがある。

3.9.1 NVO (National Virtual Observatory)

NVO は California 工科大学や Johns Hopkins 大学が中心となり、米国が設置してきたデータセンター (HEASARC, IPAC, STScI など)、スパコンセンター (イリノイ, サンディエゴ)、文献サービス (ADS, NED など)、データ解析ソフトウェア (IRAF, AIPS, AIPS++ など) を高速ネットワークで結合し、観測データ検索や計算サービスを提供することにより他波長の数値宇宙を構築し新しい天文学の研究スタイルを構築しようとするものである。NVO は天文学者のみならず計算機科学研究者と協同で開発することとなり、2001 年に NSF は NVO に 5 年にわたった 1000 万ドルの予算をつけた。

3.9.2 AVO (Astrophysical Virtual Observatory)

AVO は ESO と ESA が中心となって VLT (ESO), ISO, ST-ECF (ESA), SIMBAD 等のカタログ DB (CDS), CFHT + MEGACAM (TEREPHX), 及び MERLIN (Jodrell Bank) を高速ネットワークで結合するためのインフラストラクチャーを構築し、観測データの再利用をするだけでなく多波長、多観測装置データの比較処理を行うことを通じて新たな天文学的知見を見出そうというものである。AVO の特徴としては、この「仮想望遠鏡」に対するプロポーザルを公募することにより、焦点が絞られた研究をバックアップしようという点、また、天文データセンターの重要拠点である CDS が参加しているのでカタログデータベースが充実している点が挙げられる。EU は AVO に 400 万 Euro の予算をつけることを決定している。

3.9.3 次世代望遠鏡 (ALMA, JASMINE など) と VO との関連

ALMA 望遠鏡は、日米欧共同でチリ・アタカマ高原に建設するミリ波サブミリ波の大型干渉計である。そのデータ生産量は年間~PByte にも及ぶ。観測データはチリから日米欧に配信され、3 ヲ所の RSC (Regional Support Center) を経由して研究者が利用することとなる。

そして、米国の RSC、欧州の RSC もそれぞれの地域で構築を進めている VO (NVO, AVO) に接続していくことを前提に、RSC の機能設計を進めている。

日本にも ALMA-J の RSC を設置することとなるが、欧米との接続性を考えると、ALMA-J RSC も日本版 VO (JVO) に接続することを前提として設計を進めることになる。JVO には、国立天文台のすばる望遠鏡の観測データが接続され、また、国立天文台が計画しているアストロメトリ計画である JASMINE チームも JVO に接続することを希望している。このことは、JVO が保有する数値宇宙に「非常に精度の高い reference frame」を与えることが可能になるという、他の VO には見られない特徴を与えることを意味する。

3.10 JVO の他にない特色

6.3 章に開発体制を述べているが、開発の中心となる水本・大石は、これまで国立天文台の大型望遠鏡の制御、データ解析システム、DB システムの開発に携わってきた。野辺山の 45m 望遠鏡やミリ波干渉計の制御システム (COSMOS) はワークステーションによる分散制御システムとなっているがこれは大石の発案である。すばる望遠鏡の制御システムは、野辺山の COSMOS をベースにさらに改良・発展させた形態となっている。また、これにヒントを得て水本は分散解析システム DASH (Distributed Analysis System Hierarchy)を開発し、解析システムと制御システム、及びデータベースシステムがネットワークにより透過的に接続された。これとほぼ同時期に、大石は野辺山の観測データベースシステムを構築し、すばると同様に、制御、解析、DB システムをネットワークで接続することに成功した。

これから容易に想像されるように、**JVO はこれまでの望遠鏡システム構築の流れの自然な延長の上に成り立つ構想である。**3.8 章や 7.1 章にも述べているように JVO では、「検索できない場合は VO が観測手順書を用意して、実観測システムに観測要求を出す」ことまで構想している、天文学者が主導する VO システムである。これは単一組織内に多波長に渡る世界レベルの観測装置を持つ国立天文台でなければできない計画であり、天文学者が情報学研究者と共同し衛星 DB の統合化から始まった米国の NVO や実効的な小プロジェクトの集積からなる欧州の AVO は、そこまで踏み込んだ計画とはなっていない。即ち、**JVO では将来的には VO が全ての観測のインターフェースとなり、観測という概念そのものを大きく変えることすら想定している**のである。

4. 我々が目指すもの：JVO に必要な機能

4.1 データフォーマットの共通化

JVO では、様々な望遠鏡や観測装置で取得した観測データを統一的に扱う。このためには、分散して存在する DB (それぞれのデータフォーマットは異なる) への検索方式や検索結果の表示法の共通化、及び、各 DB から VO にデータを読み込む際に、コンバーターを通して VO 内部での標準データフォーマットにそろえる必要がある。また、解析プログラムの

データ入出力方法の共通化が必要となる。

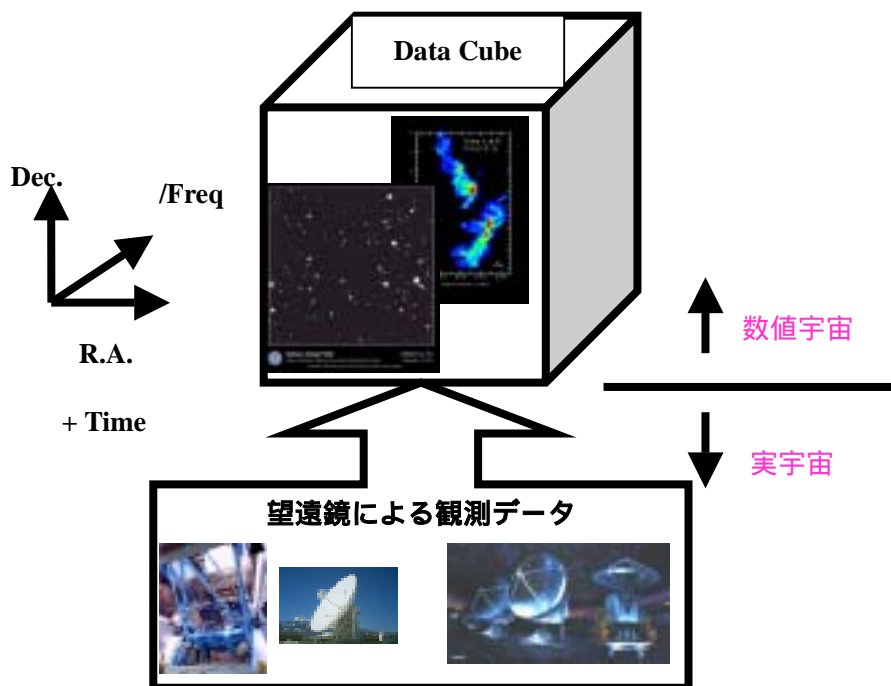


図 6 . デジタル宇宙 (数値宇宙) の概念

4.2 論理的数値宇宙の構築機能

VO に取り込まれた観測データは、OODB (Object Oriented Data Base) の機能により論理的に結合されてデジタル宇宙 (数値宇宙) の構成要素となる。観測的宇宙論や銀河形成など統計的手法を用いる研究には稠密な数値宇宙を構築することが望ましく、サーベイ観測データが中心となる。また狭視野の観測装置によるポインティング観測データも VO に取り込む機能が必要である。数値宇宙の構成要素となる観測データは、その質が保証されていなければ科学的研究に使用することが困難である。そのため観測データの質を示す指標を DB 情報の一つとして保持する機能が必須である。

4.3 数値宇宙に対する多様な検索機能

先の数値宇宙に対し多様な検索キーで検索する：座標 (赤経・赤緯の各元期、銀経・銀緯)、波長 (周波数)、観測時期、など。また一歩進めて、例えば「似た特徴を持つ天体」を探させる—4.4 の機能を応用すれば可能となろう。ユーザーは (例えば) 画像を指定し、それから VO が検索キーを自動生成するというのも面白いであろう。

4.4 大量データの解析機能

同じ視野の異なる観測時期のデータを比較するなど多様な画像処理，天体の自動認識・自動抽出およびそのパラメータ抽出とカタログ生成，空間分解能を考慮したクロスマッチ等と統計処理，wavelet，curvelet，ridgelet等を用いた天体の特徴抽出，カタログを参照しつつ未測定のパラメータを追加する機能，各種プロット機能，カタログや測定パラメータのモデルフィッティング，等。

4.5 大規模多次元データの可視化機能

検索や解析を行った結果は基本的に多次元空間に分布する物理量パラメータの組であるため，これを可視化することで研究者による結果解釈をしやすくする機能。これは，人間が持つ高度な画像認識能力をうまく活用（Visual Data-mining）して研究成果を挙げるために必須な機能であると考えられる。2D，3Dプロット（軸はユーザー指定），あるいは，可視化画像の自由な回転などもできることが求められる。アニメーション化や，あるいは，カラーと半透明表示をうまく用いた4Dプロットも有効であろう。

4.6 より高度な解析機能

データマイニング技術を活用することにより，多次元パラメータ空間の特徴抽出，クラスタリング，分類，パラメータ間の相関ルール抽出，等を可能にする。これを活用することにより，例えば，大量の銀河データのなかから自動的にこれまでの分類に当てはめると同時に，当てはまらない「未知天体」を探し出すことが可能になる。

4.7 観測所（実望遠鏡）との連携

これまでに述べたVOに必要な機能を実現するには，**観測所が観測データの一次校正まで行ってデータの質を保証し，観測DBに登録しておくことが必須である**。VOの利用者は，様々な観測装置の詳細を知らないと考えるべきであり，VOに取り込まれたデータから装置の特性を除去することを各利用者が行うことは期待できない。従って，「観測データの一次校正を行ったデータのデータベース化」までが観測所の責任で，それ以降はVO側の責任であると，明確に切り分けることで，観測所とVOの連携を行う。

5. JVOとしての必要技術

5.1 基盤技術

VOは成長するDBシステムと捉えることもでき，また，天候に左右されずにいつでもどこからでも「観測できる」ことが特徴である。従って，高い拡張性能，容易な保守性能，セキュリティも含めた高い安定性が必要とされる。そのためには分散処理（GRID）及びデータベース（Object Oriented Data Base）技術を活用しなくてはならない。

3.7 でも述べたが，GRID 技術はネットワーク上に分散した計算資源を結合し，活用するための基盤ミドルウェアとして注目されている。CERN では GRID 技術を用いて加速器データを世界中に配信することを考えており，国内では高エネルギー加速器機構と産業技術総合研究所が ALTAS の GriPhyN 計画の国内窓口となっている。また，SuperSINET を活用して遠隔地のスーパーコンピュータを連携運転させる (Grid Computing) ための基盤技術としても用いられることとなっている。

GRID 技術を実装したツールキットはいくつかあるが，機能が豊富で最も広く用いられているのが globus tool kit (<http://www.globus.org/>) である。Globus tool kit は無料で利用でき，Solaris，IRIX，及び Linux で動作する。

Globus には，遠隔地の計算機資源を用いるための，計算資源管理ツール，ローカルな認証を行うことで登録された遠隔地の計算機資源を利用できるようにするための認証サービス，セキュリティ機能などが搭載されているため，VO を構築するためには最適のミドルウェアである。また，このツールを用いて既存の解析ソフト等をラップすることにより，容易に新しいソフトウェアを追加できる。さらにネットワーク上のあるマシンが停止しても，計算機資源管理ツールによって他の同等の機能を持つマシンを自動的に選択することができるので利用者はどのマシンが使えるかを気にすることなく VO を利用することが可能となる。

一方データベースシステムであるが，項目の拡張性を確保するために，現在多く利用されている RDB ではなく，OODB の利用を検討する。

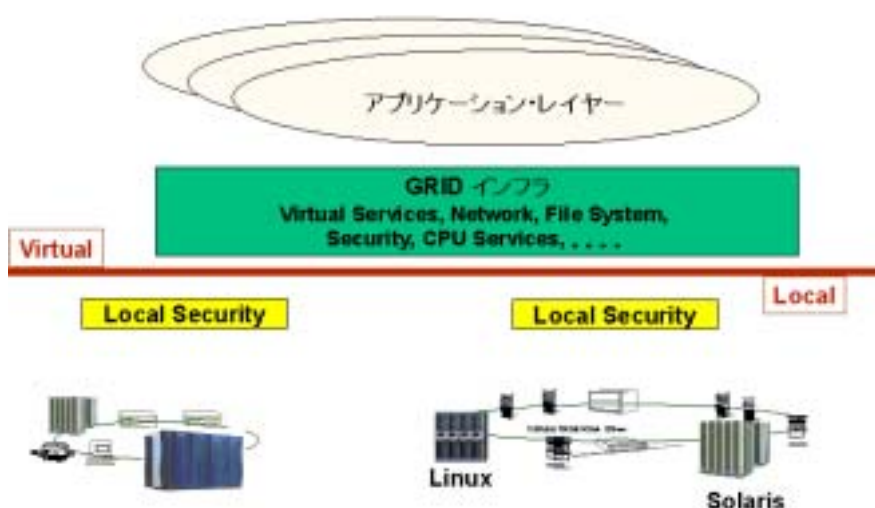


図7 GRID を利用した遠隔地計算機の結合概念

5.2 解析の自動化

VO では基本的に大量のデータを客観的に取り扱う。このためにはパイプラインによる自動処理が必要となる。パイプラインには、数値宇宙を観測する望遠鏡としての検索と観測装置としてのデータ解析（パラメータ測定やカタログ作成等）の2種類がある。また、データ解析処理ソフトを自由に組み合わせてデータ解析パイプラインを作るためのパイプラインビルダーの機能も必要である。

5.3 Move Not the Data But the Programs and the Results

さて、Grid を用いることにより、従来のように「観測データを解析マシンのハードディスクに落とす」必要がなくなる。JVO ではネットワーク上で移動するのは観測データだけではなく、解析プログラムとその結果も移動可能である。例えば SuperSINET 等の高速ネットワークを用いても数 10GByte のデータを転送するためには、それなりの時間がかかってしまい、転送したデータをローカルに蓄積しなければならないという従来からの問題に遭遇する。JVO ではネットワーク上で移動するのは、基本的には、はるかにサイズが小さい解析プログラムとその結果である。

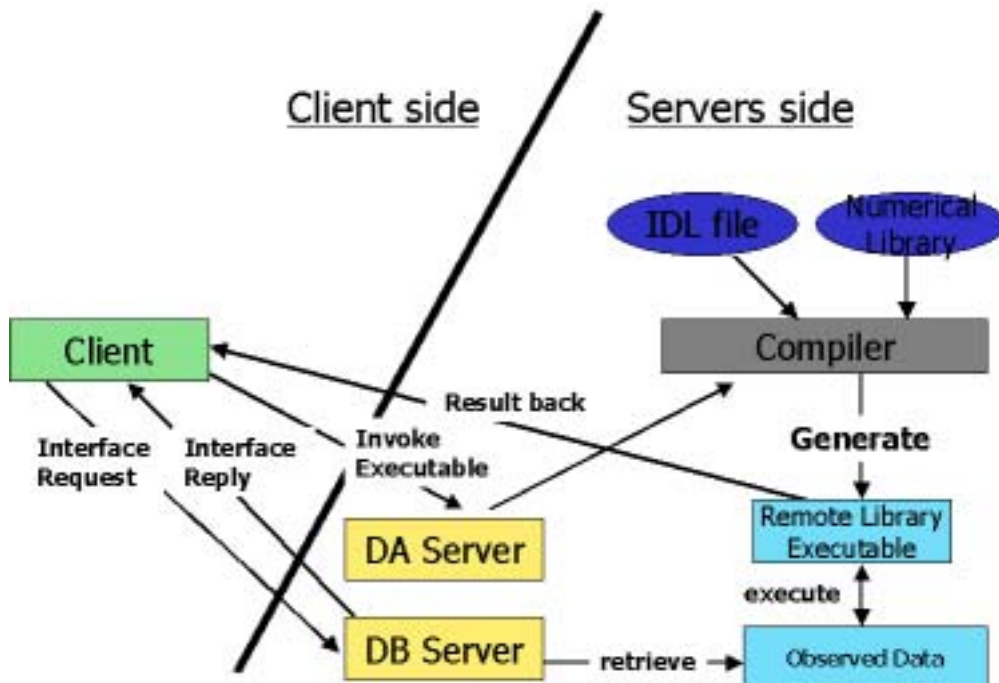


図 8 GRID 上における遠隔解析のイメージ

もちろん JVO では従来のように、分散 DB から必要な観測データを特定の計算機に移し、その特定の計算機上でしか動作しない解析プログラムを用いることも可能である。この場合でも、解析結果はネットワークを通してユーザーターミナル上で表示させることができる。

5.4 データの質の保証：望遠鏡，観測システムとの連携

すぐには難しいが、まずはサーベイ観測では data quality を保証する観測を行うことが必須である。第 7 章に述べるように VO を通じて「全ての観測」を行う時代になれば、VO が適切な観測手法や観測時間を自動的に指定することにより観測データの質を保証する。ユーザーは観測時間を指定する必要がなくなる。

5.5 多様な解析

既存解析ソフトの組み込み利用，自作解析モジュールの組み込みを容易にするために、globus tool kit によって構成される標準入出力モジュールや解析ソフトを自作するための「雛型」を提供する。自作モジュールが有用でかつ不特定多数の利用に耐えるだけのエラー処理等を組み込んであれば、これをシステムに登録して広く利用できるようにする。

5.6 可視化：データ cave

数値宇宙を観測し、新しく物理パラメータを測定して得られるカタログは多次元パラメータリストである。このパラメータ間の相関や分類には n 次元パラメータ空間の 2 次元 / 3 次元投影による可視化が有効である。立体視によってパラメータ空間中に入り込みクラスタリングの状態など人間の空間認識能力を生かして調べることができる可視化のための装置としてデータ cave の利用が考えられる。

6. JVO システムの構築

6.1 JVO の機能概要

VO のこれまでの天文データアーカイブにはない特徴として interoperability と scalability が考えられる。Interoperability は多波長にわたり様々な観測装置で得られたデータアーカイブを相互に連携させるために必要な機能であり、VO 標準データモデルの策定により実現を図る。Scalability は扱うデータのサイズに関わらず各種の解析方法が同じパフォーマンスで実行可能であるシステムとするために必要であり、解析システムを GRID 化されたプラットフォームの上に構築することにより実現を図ることを計画している。この 2 つの軸を実現するために以下のような階層的なシステムを考える。下に行くほど一般ユーザに近くなる。

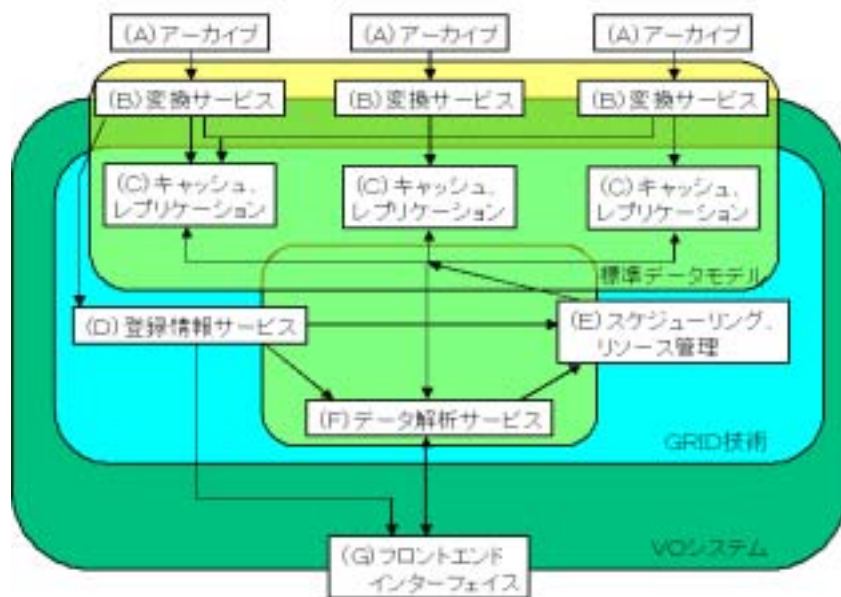


図9 Virtual Observatory プロトタイプ概念図

- (A) カタログ，アーカイブ
- (B) VO 標準データモデルへの変換サービス
- (C) データキャッシュ，レプリケーション，ユーザデータベース
- (D) 登録情報検索サービス
- (E) スケジューリング，リソース管理
- (F) データ解析サービス
- (G) フロントエンドユーザインターフェース

- (A) カタログ，アーカイブ

ここは実際のカタログ，画像データなどが格納されている場所である。実際のデータがどのような形で格納されているかについての規定はなく任意のデータアーカイブが VO へ参加できることを可能にする。ただし，格納されているデータが大量の場合には検索要求，I/Oなどを並列化してパフォーマンスをあげるように設計されている必要がある。

(B) VO 標準データモデルへの変換サービス

VO 内で各種のデータアーカイブを連携して利用するためにはそれぞれのデータアーカイブに対するアクセス方法を標準化しておく必要がある。そのために、VO 標準データモデルを策定する必要がある。この標準データモデルには各データアーカイブが提供しているデータの表現方法、必要なデータを検索するためのアクセス方法、データ処理を行うための手順、検索、データ処理の結果として返されるデータの内容の記述方法、返されるデータのデータ構造などが含まれる。これらの標準データモデルが確定していれば各データアーカイブにそれぞれのデータをこの標準データモデルに変換するサービスを実装することで内部的な構造には関わりなく VO に参加することができるようになる。逆に VO 内では個々のデータアーカイブ特有の構造には気を使うことなく相互の連携が可能になり interoperability が実現される。

(C) データキャッシュ、レプリケーション、ユーザデータベース

VO で扱うデータの中には非常に大規模なものがあり、また、それらを組み合わせて扱うことが考えられる。たとえば、光での天体カタログと赤外での天体カタログを座標でマッチングを取って光と赤外のカラーを調べるなどという解析がありえる。VO では各データアーカイブは独自にデータを修正、追加、削除できるので、各処理は動的に行う必要があるが、データベース全体を処理するような要求のあるたびに行うには非常にコストがかかるものが存在する。このような処理の結果をキャッシュしておく、または、ネットワーク上の別の場所にレプリケーションし、負荷を分散してパフォーマンスを上げることが考えられる。

また、ユーザが自分の解析ソフトで測定したパラメータなど独自に作成したデータベースを保存しておきユーザが再利用できるようにユーザデータベースを構築する。これを他のデータアーカイブと組み合わせて処理したり、他のユーザにも公開して活用してもらうことも可能である。

(D) 登録情報検索サービス

(A)、(B) によってデータアーカイブに対する標準的なアクセス方法が確立するが、それらを利用するためには、現在利用可能なデータアーカイブにはどのようなものがあり、それらからどのようなデータを取得することができるのかというような登録情報を検索するサービスが必要となる。また、VO で利用可能な解析手法なども含まれる。これらの情報をもとにしてユーザからの検索、解析要求をより細かい実際の検索、解析手順へと翻訳するのに利用する。

(E) スケジューリング，リソース管理

ユーザからの検索，解析要求を分析し，それらを (D) で得られる情報をもとにして細かい実際の手順にするとともに，それらのスケジューリング，必要なリソースの割り当てなどを行う。一部は globus などの GRID 用のミドルウェアを利用して実現することになる。

(F) データ解析サービス

実際の解析処理を行う部分である。ユーザからの要求に基づき各データサーバーへのアクセス，取得データの処理，解析結果のユーザへの表示などを行う。この部分には，通常のデータベース検索，座標などをキーにして複数のデータベースを組み合わせた検索，カタログデータを基にした相関関数などの統計的処理，クラスタリング解析，自動分類などのデータマイニング処理，画像データに対する天体検出，パラメータ測定などの処理が実装される。カタログ，画像にアクセスするための API を提供することによりユーザが独自のプログラムを作成し，それをシステムに組み入れることも可能にする。API 側で scalability を実現し，解析プログラムの開発者はそのプログラムの実行される環境にとらわれないことなく開発を進めることが可能となる。

また，処理の結果をユーザインターフェイスに向けて送り出すための加工なども行う。多次元パラメータ空間内でのサンプルの分布を見る際にクラスタリング解析などを行って表示する処理が考えられる。

(G) フロントエンドユーザインターフェイス

ユーザが VO にアクセスし，処理要求を発行し，結果を得るためのインターフェイスである。VO で利用可能なデータ，リソースなどを参考にしながら処理要求を組み立てデータ解析サービスへ処理を依頼する。解析結果は適当に処理されて返される。

VO へのアクセスはネットワークの太いところも細いところも考えられるので，それらに対応できるように結果の表示の仕方を変化させる。たとえば，データ分布図のような場合は最初はコントラで表示し，要求に応じて徐々に一部を拡大していくなど。

6.2 開発計画

JVO の構築は段階的に進めてゆく。まずプロトタイプを構築して VO という概念が成り立つことを検証した後に，本システムの構築を行う。その後 ALMA 等のデータ解析が可能になるようにコンポーネントの充実を図る。

JVO構築全体計画案

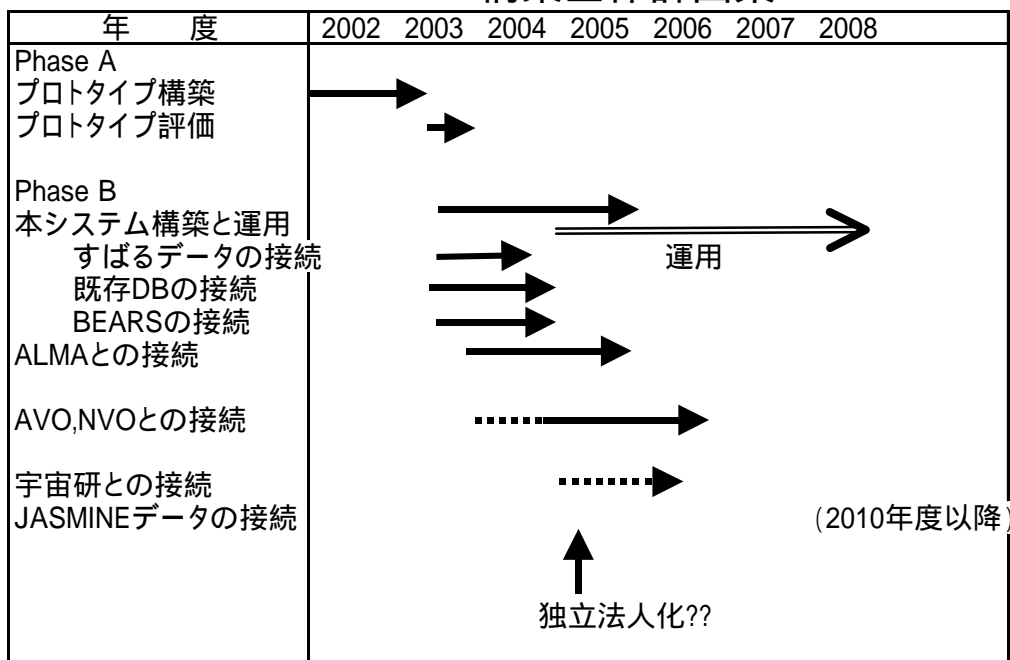


表 1 . JVO 構築の全体計画案

6.2.1 プロトタイプ開発 (2002 - 2003) : Phase A

6.4 で述べるプロトタイプは、VO の概念に沿った開発研究を行うことにより、想定したVOの機能が実現できることを確認することを目的とする。従って、国立天文台が保有あるいは保有するであろう全ての観測データを対象とするのではなく、まず、大規模サーベイであるSDSS, 2MASS, すばるサーベイプロジェクトのデータ、及び、数種の天文カタログを対象とする。

6.2.2 本システム構築(2004 - 2005) : Phase B

プロトタイプ of 構築を通じて基本的機能を実現させた暁には、これまで国立天文台天文学データ解析計算センターが中心となって構築した既存の各種DB(HST, DSS, DSS2, IUE, NRODB, MOKA等)を包含するためのフォーマット・コンバーターを用意して、これまでのDBをVOに統一する。これと並行して、すばる望遠鏡や45m望遠鏡のBEARSを始めとする他の観測装置による、ポインティングデータも含めた観測データをVOに取り込むための開発を行う。これによりJVO初版が完成する。

6.2.3 近未来 (2006年頃まで)

1.4.3にも述べたがALMA-J RSCはVOの概念に沿って構築することとなっている。よって、ALMAチームではALMAのデータをVOに取り込むためのフォーマット・コンバーターとデータ解析モジュールを開発すれば、JVO初版に組み込むことができる。また宇宙科

学研究所の IRIS , (2010 年以降の作業となるが) 国立天文台の JASMINE も接続できるように共同で開発を進める必要がある。

6.3 開発体制

6.3.1 情報学 , 計算科学的開発

遠隔地 DB の結合や超大規模 DB の設計には情報学 , 計算科学の研究者の参加が必須である。これまでの検討・議論・研究予算獲得の過程で , 下記の方々の協力が得られることとなった。

超大規模 DB 研究の第一人者であるお茶の水大学理学部の増永教授を平成 14 年度の国立天文台客員教授に迎えることとなった。

九州大学情報基盤センターの松尾教授及び竹田助教授は , SuperSINET を利用して遠隔地 DB の論理的結合において協力していただけることになっている。

Grid 関連では , 特定領域研究 C 「情報学」の公募研究として我々の研究計画が採択されており , その研究班の中で東京工業大学の松岡教授 (Grid を用いたクラスター構築の第一人者) と協働できることとなった。

大阪大学サイバーメディアセンターの下條教授や大崎助手はネットワークを効率的に利用する研究を行っており , 我々の VO 構築においてネットワークの性能を十分に出す面で協力いただけることになっている。

6.3.2 天文学的開発

天文学特有のソフトウェアやデータフォーマット等に関しては , 国立天文台 DB 天文学推進室スタッフ (水本 , 安田 , 大石) に 3 名程度の研究員を加えて開発研究を進める。幸いにして天文学データ解析センター内の計算機資源 , 及び , 科学研究費補助金が多少あるので , これらを利用することで Phase A の段階では特段の費用はかからないと考える。但し , 実際の開発の際には 3 名の研究員を必要とする。研究員の役割としては ,

新研究員 A : VO の標準データモデルやその関連ツールの開発

新研究員 B : GRID 上における計算機資源の登録・モニター機能・データ解析ツールの開発

新研究員 C : 大規模データに適した DB 及び DB 連携方法の開発

を考えている。

またデータベース天文学推進室における開発と同期を取りながら , すばるや野辺山のソフトグループと協力し , それぞれの観測データを JVO に接続するための機構作りを進める。

6.4 プロトタイプ作成計画

プロトタイプでは、データアーカイブとして SDSS の公開データと 2MASS の公開データを用いて、効率的なデータベースの構築、データ配置方法などについて検討を行う。そして、VO 標準データモデルのプロトタイプを策定し、SDSS、2MASS のデータに対するデータ変換サービスを実装する。プロトタイプの策定には、将来データプロバイダとして参加することになると予想されるすばる望遠鏡、ALMA のデータアーカイブ担当者の参加も求め、考えうるできるだけ広範なデータ形式を網羅できるものを策定することを目指す。

VO 標準データモデルは将来の国際 VO との連携を考えると米国、欧州で策定されるものと大きく異なるものにはしたくないが、開発時点での状況に応じて必要であれば独自にプロトタイプとして策定していく。

プロトタイプの段階ではデータキャッシュ、レプリケーション、スケジューリング、リソース管理などは GRID 関連技術として実現される可能性が大きいので当面後回しにし、infrastructure としてはユーザデータベース、登録情報検索サービス、解析要求の解釈機能の実現を目指す。データ解析サービスとしては、標準データモデルに基づいて解析プログラムを開発し、2 つのデータベースを連携させた検索機能、ユーザ作成プログラムのための API の策定とその実行機能の実現を目指す。

これらの VO 内部での interoperability の実現のための仕様の策定を主にしたプロトタイプの作成に 2 年程度を費やす予定である。その間の、他国での進捗状況、GRID など関連技術の進歩などを踏まえて、プロトタイプの評価を行い、本格システムの作成に取り掛かる。また、このシステムがうまく動くようであれば、すばる望遠鏡によるサーベイデータや 野辺山 45m 望遠鏡の BEARS データ、ASTRO-F のサーベイデータ、ALMA のデータなども、標準データモデルへの変換サービスを実装することで容易に VO の中に参加させて行くことができると考えられる。

JVO Phase A スケジュール案

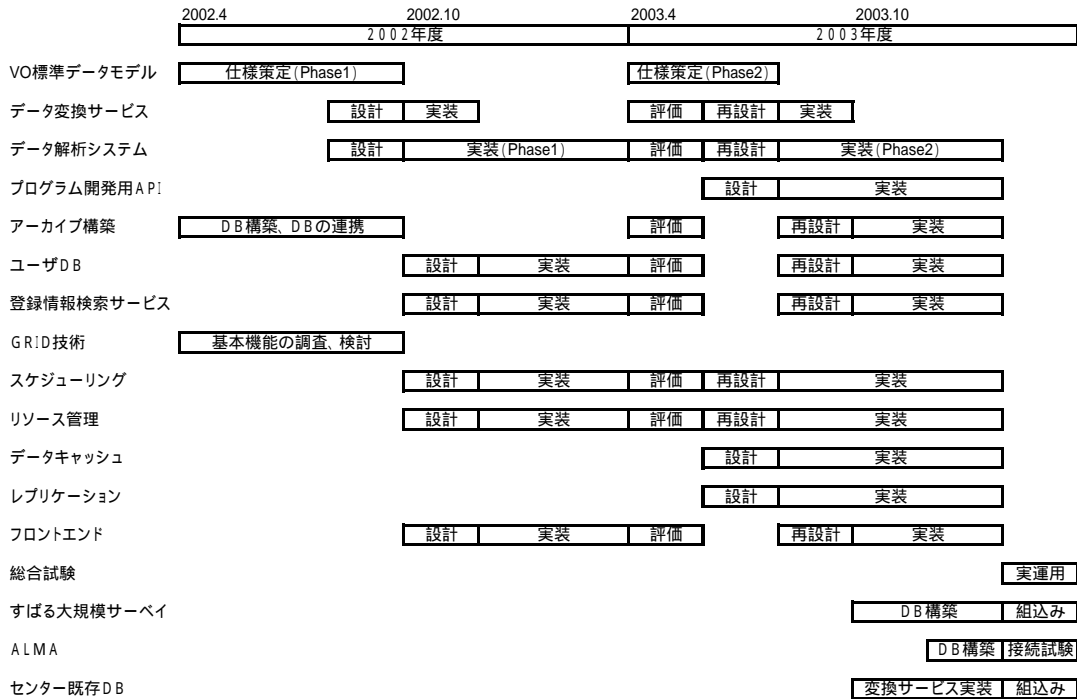


表 2 . Phase A の工程表案

6.5 それで何が出来るか：利用法のイメージ

VO の基本的な利用パターンとしては、以下のものが考えられる。

- すでにカタログ化されているデータを各パラメータ空間で制限をつけて自分の興味のある天体を選び出す。この時に任意のカタログをまたぐ検索も可能。検索結果はユーザデータベースに保存される。
- カatalogなどにリストされているデータに対応するいろいろな波長の画像に対して画像解析プログラムで新しいパラメータを測定して結果を保存する。
- 保存されているすべての画像データに対して独自の画像解析プログラムを走らせて新しいカタログを作る。
- カatalog化されているデータについて統計的解析，クラスタリング解析などを行う。

プロトタイプ (Phase A) での SDSS と 2MASS を使ったデータからは以下のような科学的研究が可能になると考えられる。

- 可視光と赤外線の合わせて 8 (5 + 3) バンドを使った photometric redshift から分光限界よりも暗い銀河まで redshift を求め、銀河の光度関数の redshift ごとの変化を調べる。

- ・ 可視光と赤外線での銀河計数観測を利用して銀河系による星間吸収の場所依存性を遠赤外線によるダストの観測とは独立に定量化する。
- ・ 可視光と赤外線のデータから星の分類を行い、それらの空間的な分布から銀河系のハローの構造と起源を調べる。
- ・ 多次元のパラメータ空間の中からこれまでに知られていない性質を持つ天体を発見する。
- ・ 同じ領域を別の時間に撮ったデータを比較することで、超新星などの変光天体、小惑星などの移動天体を発見しそれらの統計的な研究を行う。

7. その先にあるもの

7.1 VOは全ての「実観測」のインターフェース

Grid ベースの VO は、様々な計算機資源上にある各種リソースを容易に接続することができる。従って、「実観測」のための望遠鏡制御システム群を GRID 化することにより、VO をインターフェースとしたリモート観測機能が実現可能である。野辺山宇宙電波観測所の 45m 電波望遠鏡やミリ波干渉計のための観測制御システムである COSMOS は LAN 上での分散制御システムとなっており、これらを GRID 上に移植する作業は比較的容易であると予想される。またすばる望遠鏡の制御システム SOSS も野辺山の COSMOS をベースに構築されており、GRID 化は可能であろう。このように VO に観測制御システムを接続すれば、VO にデータがない場合に VO が観測要求を作成し、それを「実望遠鏡」に実行させることが可能となる。

一方、ここで述べた機能は、観測データの質を制御するために必須となろう。即ち、VO は「全ての観測」のインターフェースとなり、いつでもどこでも観測できる究極の「観測装置」へと進化するのである。

7.2 VOはひとつ：世界のVOとの連携

さて、3.9 で述べたように NVO や AVO の構築が欧米で進みつつある。それぞれが包含する観測データは異なるため、JVO、NVO、AVO を相互接続し、かつ、それぞれの観測データを透過的に利用できるようにすることは、世界の天文学研究の加速度的進展をもたらす。そのためには、標準データフォーマット（観測データの VO 内部形式）の共通化を行うか、もしくは、それぞれが他の VO とのデータフォーマットコンバーターを用意さえすれば複数 VO の容易な連携が可能となる。

7.3 大学・大学院教育への波及効果

VO が実現すると、大学や大学院における若手教育、特に観測による研究者を育てる際の望遠鏡観測時間の確保問題が半減する。全ての研究者は、VO を「自分の望遠鏡」であるかのように利用すればよいのである。

また、VO に登録されていない新たな解析ツールが必要になり自作することは十分にあり得る。この場合は GRID インターフェース仕様 (VO 仕様) に従ってコーディングしたプログラムは容易に VO で利用できるばかりでなく、システムに登録することにより他の研究者にも利用してもらうこともでき、大学や大学院におけるソフトウェア開発を行う大きな原動力となる。

7.4 小中高における理科教育への波及効果

近年「理科離れ」、「学力低下」が叫ばれている。これには様々な要因はあると思われるが、第一線の天体画像が教室から容易に「観察」することができるようになれば、自然の不思議さ・おもしろさから自ら勉強しようという意欲にも繋がっていくであろう。VO は、まさにこれを実現するものであり、VO に非専門家向けの検索機能を追加することによって小中高校における理科教育を大きくバックアップすることが可能なものであると同時に、将来天文学を志す子供達に対し強い動機付けを与えることもできる。

7.5 Public Outreach

日本はアマチュア天文家の数が非常に多いだけでなく、天文ファンも一般社会に多い。また、最近では国民の税金で運営されている公的機関の説明責任が求められている。国立天文台には情報公開センターがあり、最新の天文学の成果を分かりやすい形で国民に還元する努力を行っているが、VO に非専門家向けの検索機能を追加することによって成果の国民への還元はこれまで以上に容易になるものと期待される。ブロードバンドインターネット接続は一般家庭にも普及しているため、家庭から VO に接続することによってすばるなどが撮像した画像を楽しむことが容易に可能となる。また、可視光の画像だけでなく、他の波長における画像等にも容易に接することができ、可視光以外の天文学の理解をも促進できると考えられる。

8. データベース天文学の世界

抽象化望遠鏡と抽象化観測装置は観測波長を限らないし、数値宇宙もあらゆる波長の観測データを含む。そこで、数値宇宙を観測すれば、赤外線天文学とか電波天文学とかはもちろんのこと「光学赤外線電波天文学」とか「電波 X 線天文学」とか複数の波長に渡った天文学の研究が容易にできるようになる。こうなると多波長のデータを使った研究方法が当たり前になり、それをただの「天文学」と呼び、狭い波長領域のデータだけを使った研究を例えば「可視天文学」とか呼ぶ時代が来るかもしれない。その中間の時期にあるのが「データベース天文学」だろう。

新しい波長で宇宙を観測すると何が見えるか予想がつかない。VO を用いたデータベース天文学も同じで、どんな発見があるのか分からないから「わくわく」する。データベース

天文学の研究は

異なった波長で、
異なった装置で、
異なった時間に

観測された異なったデータの膨大な塊(数値宇宙)の中からあるパターンを探すという方法
を採る。これは金や石油の鉱脈探し(探鉱)になぞらえて「データマイニング」と呼ばれる。
その一つの方法は、星の分類に用いられる HR 図(星の色(温度)と明るさの 2 パラメータの分
布図)を拡張したものである。多数の天体について様々なパラメータ(n 種類)を測定し、その
 n 次元パラメータ空間の中での天体の分布を調べる。分布図の中でデータ点が密集した塊(ク
ラスタ)の形や種類、塊にならないデータ点などを調べ、天体の分類を行う。例えば、同じ
領域の長期間にわたる観測データの中から明るさや位置の時間変化を調べることにより、
高赤方偏移の Hypernova やガンマ線バースト天体の探索、クエーサの時間変化の研究、マイ
クロレンジングや系内惑星、太陽系近傍天体の探索などができる。天体の 2 つの色の分布
図からは高赤方偏移のクエーサや、冷たい白色矮星、褐色矮星など稀な天体の探索が可
能。複数の色での天体位置の相関から cosmic string (宇宙紐) 探索などもできる。

このような研究では大量データの統計処理を駆使するとともに、 n 次元パラメータ空間内
のデータ点分布の特徴を分析するための道具が必要となる。人間の高度な空間認識能力を
生かすための、 n 次元パラメータ空間を 2 次元または 3 次元に投影して見せる可視化技術が
威力を発揮するだろう。実際に n 次元パラメータ空間の 3 次元投影の中に入り込んでデー
タ点の分布を詳細に観察することができる。子供の頃に屋根の上に寝ころんで夜空を見上
げると星空の中に自分が浮かんでいる様な感じを経験された方も多いただろう。データベ
ース天文学者は数値宇宙という仮想空間に入り込んで観察するとき、ふと懐かしい想いが甦
るのではないだろうか。

VO 上のデータベース天文学によって、宇宙年齢のような宇宙論パラメータを膨大なデー
タの統計処理で正確に決定するような天文学の精密化 天文学が精密科学になる! と、
未知天体の発見ラッシュが期待されるのである。