

特集・データベース天文学 その(2) 世界の「仮想天文台」計画

洞口 俊博

〈国立科学博物館理工学研究部 〒169-0073 東京都新宿区百人町 3-23-1〉
e-mail: horaguti@kahaku.go.jp

地上やスペースの観測装置から得られるデータ量は、近年飛躍的な増加を示している。「仮想天文台」は、さまざまな波長域のそれらの観測データやカタログデータ等を統合し、あわせて比較、統計などの解析処理を行うことによって、これまでの方法では得られなかった天文学的成果を目指す開発プロジェクトである。アメリカ、ヨーロッパなどで立ち上がりつつあるそれらの計画について概観し、今後の展望、日本の果たすべき役割について述べる。

1. はじめに

1.1. 「仮想天文台」計画とは

最近 Virtual Observatory (仮想天文台) というプロジェクトが世界の注目を集めている¹⁾。この計画は、現在世界中に分散して存在する天体観測データのアーカイブを有機的に結合し、それらを解析するツールを開発することによって、これまで知られていない宇宙の新しい姿を明らかにすることをねらいとしており、膨大な観測データを収め管理する計算機やデータベースシステムを「望遠鏡」、そのデータを解析するツールを「観測装置」として、データアーカイブに記録された宇宙を観測するという意味から「仮想天文台」という名前がつけられている。

ハッブル宇宙望遠鏡 (HST) や X 線、赤外線などの人工衛星、さらに地上の大口径望遠鏡、サーベイ観測望遠鏡など、さまざまな大規模観測プロジェクトによって、われわれはこれまでにない大量の情報を得ることが可能となってきている。例えば、HST からは毎日 3-4GB、年間約 1TB のデータが生み出されており、さらに数 TB クラスのサーベイ計画もまた数多く立ち上がり始めている。これらの観測においては、観測者はもちろんそれぞれのターゲ

ットについて解析、研究を進め、プロジェクト主体はまたそれぞれカタログなど有用な情報の抽出を行ったりするわけであるが、その他にもこれらのデータの中には興味深い有用な情報が埋もれていることは明らかである。実際、HST では全観測データののべ3倍、赤外観測衛星 ISO では2倍の量のアーカイブデータがこれまで利用に供されており、さまざまな成果をあげている。計算機の中のデータは、もう1つの宇宙といっても過言ではないのである。

これらの観測データは一般に、プロジェクトごとに分散して管理され、それぞれ個別の方法で公開されているのが現状である。その理由の1つとして、天文の観測情報は得られるデータの形がさまざまであり、装置の特性やデータの内容、意味を熟知した者がそれぞれ固有の処理を行った上でないと利用できず、その処理手順もしばしば更新されていくということがあげられる。これらの観測データを組み合わせ、例えば異なる波長のデータと比較したり異なる時期のデータと比較したりすることが自在にできるようになれば、天体の物理的性質や活動現象に対する理解は大きく進展すると期待される。また個別の天体の探求だけでなく、観測データ全体をサーベイすることが可能になれば、100万天体に1つといった貴重な未知の天体の発見や大

規模サンプルの統計に基づくような研究についても大きな道が開かれることは確実である。このような研究の実現こそが Virtual Observatory, 「仮想天文台」建設の目的なのである。

1.2. 「仮想天文台」の目標

「仮想天文台」計画の最終目標は、以下のよう
なデータ、計算機などの資源：

- ・ TB クラスのデータアーカイブ
- ・ 100 万個クラスの γ 線から電波まであらゆる波長域における天体画像ライブラリ
- ・ データ解析・マイニング用ソフトウェアツール
- ・ PB クラスの外部記憶装置をそなえたテラスケールの超高速計算機
- ・ 主要な天文センターを結ぶ超高速ネットワークを結合し、以下のような機能を実現することである。

- ・ TB クラスアーカイブの高速検索
- ・ 大規模カタログ・画像データベースの可視化
- ・ 特異天体・現象やパターンの抽出
- ・ 大規模な統計的研究, 理論計算との比較

また、より具体的な目標としては

- ・ 観測データのパイプライン処理, アーカイブ格納・取り出しに関する共通システムの確立
- ・ さまざまな波長における既存の大規模データベースの結合
- ・ 新たな大規模データベースを組み込むための枠組みの提供
- ・ カタログデータや画像のピクセルレベルデータの検索・可視化・相関・分類・統計解析ツールおよび理論計算との比較解析ツールの開発
- ・ ユーザーによるプログラム開発および計算サーバへのアップロードのための枠組みの提供
- ・ 既存および新規の文献情報等との結合
- ・ データ解析用超高速並列計算機の開発
- ・ アーカイブデータや超高速計算機を結ぶ高速ネットワークの確立
- ・ 主要な天文センター・研究機関との国際的な協力の推進

- ・ 同様な課題（大規模アーカイブの統計解析, 分散データ統合利用・処理など）をもつ他分野（高エネルギー物理学, 遺伝子科学, 地球科学など）との連携協力の促進があげられている。

2. 進行中の「仮想天文台」計画

実際の Virtual Observatory (VO) 計画は1つではなく、現状では表1に掲げたような国や地域においてそれぞれ進められている。どの計画も立ち上がって間がなく、目標やスケジュールなど今後の計画の進め方が定められた段階で、具体的な開発はいずれもこれからというところである。以下、それぞれの計画について概要を紹介する。

2.1. National Virtual Observatory (米)

米国 National Academy of Science の天文天体物理調査委員会は、新世紀を迎えるにあたって次の10年の研究活動に対する評価を行い、今後優先して進めるべきプロジェクトについて勧告を行った⁶⁾。National Virtual Observatory はその中でトップに掲げられた計画の1つで、全米科学財団 (NSF) から5年間で1000万ドル (約13億円) の研究費を得て、NASA や Caltech, SDSS のグループなどの研究者をメンバーに、2002年から本格的にプロジェクトを開始している。

主な開発要素としては

- ・ メタデータスタンダード：データの相互利用に必要なデータの構造や内容, 意味, 提供サービスなどを記述したメタデータの標準規格
- ・ データアクセスレイヤ：すべてのデータ, メタデータ, サービスのための共通インターフェース・プロトコルを提供する下位ソフトウェア。これによって上位ソフト開発者はデータの所在や格納方法を意識する必要がなくなる。データの部分転送などのためのサーバ側ソフトや画像のキャッシュ, ユーザーのソフト開発などのためのクライアント側インターフェース等を含む

表1 各国(地域)のVO計画およびそれらに関連した主な研究会(参考)

計画(国・地域, 関係機関)	関連 URL
National Virtual Observatory (米, NASA/NSF/...)	http://www.voforum.org/
Astrophysical Virtual Observatory (EU, ESO/ST-ECF/CDS/...)	http://www.eso.org/projects/avo/
AstroGrid (英, Edinburgh 他4大学/Jodrell Bank/MSSL/RAL)	http://www.astrogrid.org/
Australian Virtual Observatory (豪, CSIRO)	http://www.atnf.csiro.au/projects/avo/
(参考) 研究会(開催地, 開催年)	
Virtual Observatories in the Future ²⁾ (Pasadena, 2000 June)	http://www.astro.caltech.edu/nvoconf/
Mining the Sky ³⁾ (Garching, 2000 July-Aug.)	http://www.mpa-garching.mpg.de/~cosmo/
Astronomical Data Analysis Software and Systems X ⁴⁾ (Boston, 2000 Nov.)	http://hea-www.harvard.edu/ADASS/
Astronomical Data Analysis ⁵⁾ (San Diego, 2001 Aug.)	http://spie.org/conferences/Programs/01/am/confs/4477.html
Astronomical Data Analysis Software and Systems XI (Victoria, 2001 Sep.-Oct.)	http://cadwww.hia.nrc.ca/adass_2001/
Toward an International Virtual Observatory: Scientific Motivation, Roadmap for Development and Current Status (Garching, 2002 June)	http://www.eso.org/gen-fac/meetings/vo2002/

- ・ 検索・解析エンジン：画像のピクセルレベルでの扱いも含めたデータ検索・処理手法および分散・並列計算システム
 - ・ データマイニングソフトウェア：多変量の統計解析や可視化を含むアプリケーションやインターフェース, ツールキット
 - ・ 計算機システム：大容量記憶装置, データ管理, 高速ネットワーク技術
 - ・ 学校・市民向け教育普及プログラムなどをあげ, 5年間の計画としては
 - 2002-2003：統合されたサービスの確立
 - ・ メタデータスタンダード・通信プロトコル・表示サービスの確立, データ検索・配信システムの標準化
 - ・ 小規模な相関・可視化ツールの実装
 - ・ 大規模な相関処理プログラムのプロトタイプの開発
 - ・ 最終的に必要なネットワーク・計算能力の評価
 - ・ 教育普及プログラムの確立
 - ・ 国際的な情報基盤の設計, 確立
 - 2003-2004：本格的運用開始
 - ・ 関連計算機, ネットワークの本格導入
 - ・ データアクセスレイヤの開発, 実装
 - ・ データ検索・比較ツールの完成, 定常運用
 - ・ 大規模相関プログラムの実装
 - ・ 複雑なデータセットに対する可視化ツールの開発, 実装
 - ・ 海外機関との相互運用のための連携の確立
 - 2005-2006：運用体制の確立
 - ・ データアクセスレイヤの本格実装, 定常運用
 - ・ ユーザー定義による処理エージェントのサポート
 - ・ 海外協力機関へのサービスの拡張
- のような3段階で進める計画となっている。

2.2. Astrophysical Virtual Observatory (EU)

Astrophysical Virtual Observatory は欧州で立ち上げ

られた VO のプロジェクトで、3 年間で 400 万ユーロ（約 4.6 億円）という予算で今年から Phase A の開発研究を開始している。本格運用の Phase B を目指す Phase A の主な目標は以下のとおりである。

- ・分散的でスケーラブルな計算基盤を提供する Grid 技術を含めた、設計・実装・運用に必要な要素の洗い出し
 - ・データアーカイブの連携に必要なスタンダード、インターフェースの定義
 - ・既に運用中のさまざまな波長のアーカイブへの技術の適用と検証
 - ・一定のデータの質を保証するための既存アーカイブのレビューと標準的なクォリティーコントロールに必要な要素の洗い出し
 - ・他分野の Grid 計画との協力も含めたスケーラブルな計算機、記憶システムの評価、開発、配置
 - ・リモート資源利用のための Grid 技術テストシステムの開発、配置
 - ・国際的な拡張を視野に入れた米、加、豪との協体制の確立
- Grid について多く言及されているが、それについては 2.3 で触れる。参加している主な組織は、
- ・ESO：VLT, NTT など多くの地上観測データアーカイブを有する。VLT の本格運用によって 50TB / 年、また現在進行中の VST 計画では 100GB / 夜のデータ産出が予想されている。
 - ・ESA：HST, IUE, ISO, XMM-Newton など多くのスペース観測データアーカイブを有する。ま

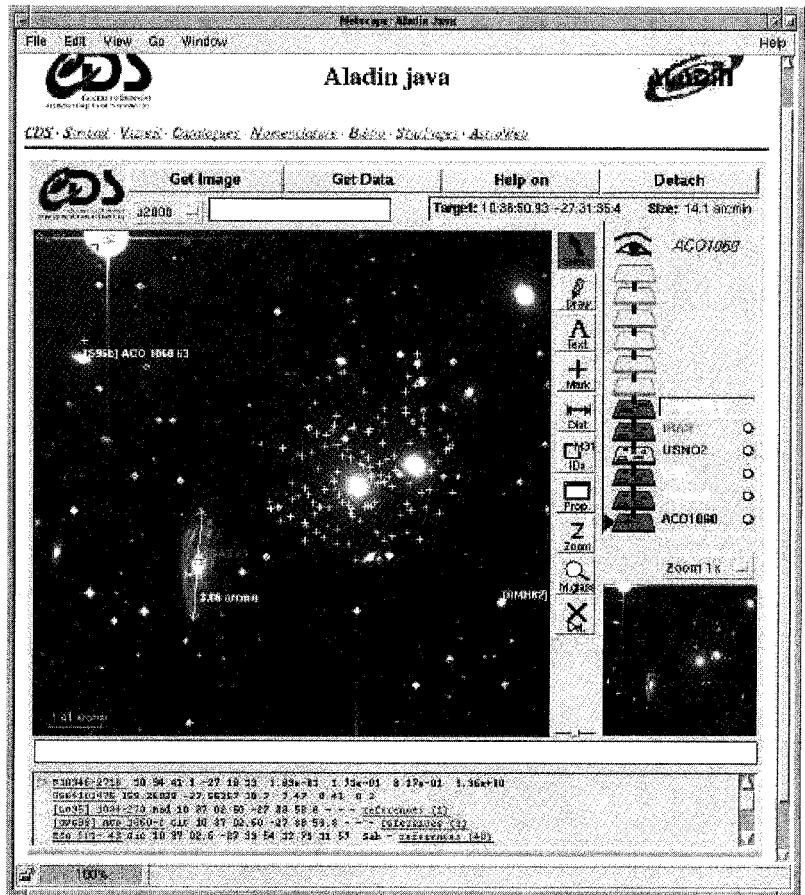


図 1 ALADIN の画面の例。DSS の画像の上に SIMBAD や GSC などのカタログの天体位置がプロットされている。下の方にはクリックした天体についての基本的な情報およびより詳しい情報 (SIMBAD や VizieR) へのリンクが示されている。

た ESO と協力して後述の ASTROVIRTEL プログラムを実施している。

- ・CDS：天体・文献情報サービス SIMBAD⁷⁾、カタログデータサービス VizieR⁸⁾、スカイアトラス ALADIN⁹⁾ (図 1) 等の開発・運用を行っているストラスブールの天文データセンター。天文情報の標準化・相互利用に関して大きな実績を有する。
- ・TERAPIX：CFHT 主焦点で計画されている広視野撮像装置 MEGACAM を主とした大画像のデータ解析センター。Institut d'Astrophysique de Paris におかれている。

- ・ Jodrell Bank 天文台：MERLIN (Multi-Element Radio Linked Interferometer Network) 電波アーカイブを運用している。較正済データで 10 TB / 年になる e-MERLIN 計画が進行している。
- ・ AstroGrid：Grid 技術の天文分野における利用について研究している英国のプロジェクト。詳しくは 2.3 を参照。

となっている。ASTROVIRTEL は 2000 年から始まったプログラムで、ESO や ST-ECF などのアーカイブを用いて、将来は計算機を用いて実現される VO の機能をまずは人手でシミュレーションしてみようという試みである。年間 5 つ程度のプロポーザルを受け入れて、必要要素の洗い出しや機能の開発が進められている。この他にも多波長のデータを利用した ESO の高赤方偏移宇宙サーベイプロジェクト GOODS でも必要要素の洗い出しが行われ、また市販の PC をネットワークで複数結合した高速データ処理用 Beowulf クラスタ計算機の開発・運用も各所で開始されている。

2.3. AstroGrid (英)

AstroGrid は英国の e-science 計画で提案された Grid プロジェクトの 1 つで、Grid 技術の天文分野への適用を目指して、3 年間で 500 万ポンド (約 9.4 億円) の予算で 2001 年からスタートしている。Grid は 1990 年代後半に生まれた概念で、コンセントにプラグを差し込めばどこの発電所の電気が利用者に意識させることなく最適な方法で電力を供給してくれる電力網 (electrical power grid) のように、一度ネットワークに接続すればデータや計算機がどこにあるか意識しなくても利用できる分散的で動的協調的な計算機環境の実現を目標としている。e-science 計画では天文学の他に生命科学や素粒子科学などの Grid プロジェクトが進められているが、国際的にも米国の大型加速器や LIGO (レーザー干渉重力波天文台)、SDSS 計画が参加している GriPhyN プロジェクトや CERN を中心とした EU Data Grid プロジェクトなどが活動を開始している。

AstroGrid の実現において必要な技術としては、データやメタデータ、プロトコルの標準化、データベースの検索手法、内容解析、可視化などのほかに、計算資源の割り当て、負荷分散、セキュリティ・アクセス管理、分散データ管理、検索エージェントなどがあげられている。プロジェクトは国内のキーとなるデータベースについて実動システムを作成すること、そして国際的な VO を構築するにあたって重要な寄与と認められるようなツール、手法を開発することを目標としており、手始めとしてまず SOHO や Yohkoh の太陽データで地理的に離れたデータの結合、SuperCOSMOS や SDSS などの光赤外データで巨大データの結合、XMM-Newton や Chandra の X 線データでフォトンイベントデータの結合、CLUSTER や EISCAT の宇宙プラズマデータで非常に複雑で異なる種類のデータの可視化、MERLIN や VLA の電波データで実時間再処理の実現を目指している。

2.4. Australian Virtual Observatory (豪)

オーストラリアでは、CSIRO の Norris らを中心に計画の議論が開始されている。国際的な VO の一部として他の VO 計画と密接に協力を進めることによって、オーストラリアのデータを世界に提供すること、オーストラリアの研究者に国際的な VO を通した研究環境を提供することを目標とし、さらに一定の質のデータを確保するための観測手法や観測システムのありかたについても議論を行っている。

3. まとめ 一国際協力に向けて一

観測によって得られるデータ量の増大は、CPU の演算能力や記憶装置の容量の伸びをはるかに上回る速度を示している。一方、これらのデータに対する一般研究者の期待は非常に高く、既存のデータアーカイブの利用実績から考えて、1 つの天文台建設あるいは 1 つの科学衛星打ち上げに匹敵するものと考えられている。VO 計画は、これらの

莫大な新規データを個別に管理し処理するのではなく、他の観測データやカタログデータと結合できるような共通の枠組みを整備し、まったく新しい科学的成果を目指して比較、統計などの解析処理を行うことが可能な統合的な環境を構築することを目標としたプロジェクトである。その際、ユーザーにデータや処理計算機・ソフトウェアの所在を意識させず、また適切な時間内に命令を処理し結果を返すために、Grid技術を用いた分散処理や並列処理が重要になると考えられている。

VO計画は現在世界各地でそれぞれ立ち上がりつつあるところで、実際的な開発はまだ始まったばかりである。米国では、既に行っているサービスとしてMAST（スペースを中心としたデータアーカイブ）、ADC（カタログ）、ADS（文献）等の実績があり、またSDSSでは実際にVOを意識して天体座標の階層的表現や検索の並列化等に関する研究が行われている。欧州では、これまで多数の地上・スペース観測アーカイブがESO/ST-ECFによって開発、運用され、またカタログ・文献サービスについてはCDSが非常に積極的な開発・運用を行ってきている。CDSはこれまでの経験と実績を生かし、VOの開発においても、データや情報の交換方法の標準化について世界的な中心となりつつある。

印象としては、米国は計算機関係を中心に豊富な人的資源を有し、高い技術志向をもってVOの開発を進めようとしているようすが感じられる。ただし、汎用技術では扱えない天文の特殊要素も意

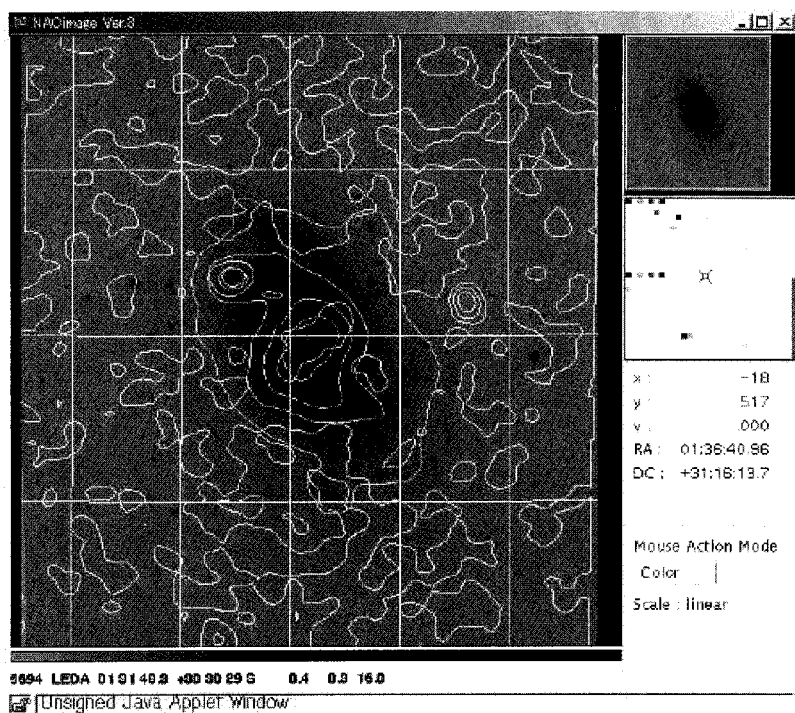


図2 jMAISONの画面の例。DSSの画像の上にIRAS (100 μm) や Greenbank (4.9GHz) のデータが等輝度線表示されている。

識して、技術的にしっかりした議論が行われている反面、実際のデータとの関係がまだ薄く理念が先行した議論となっているところも感じられる。一方、欧州については、さまざまなアーカイブ、カタログ等のサービス実績を基礎として、可能なところから着実に必要な要素の開発、組み込みを行い、実用的なシステムを構築していこうという姿勢が感じられる。その反面、分散・並列処理といった計算機・ネットワーク技術的な面に関しては、英国のAstroGridにかなりの部分を依存しているようにも見受けられる。

このようにそれぞれ特徴をもって進められているVO計画であるが、いずれも国際的な連携を重視しており、うまく相補的に協力して開発を進めることができれば大きな進展が期待される。実際、メタデータの標準化については、CDSのフォーマット

を元に欧米の VO 関係者が協力して VOTable という規格が作成され提案が行われている。

日本は世界的にも第一級のすばる望遠鏡や X 線衛星などの観測データを保有している。また、ALMA や赤外線衛星などの有力な計画も多数控えている。これらのデータは世界的にも大きな財産であり、他のデータと組み合わせることによってさらに大きな成果が期待される。日本はソフトウェアの面でも、他の類似システムにはないユニークな機能を有する SMOKA (すばるアーカイブ) や jMAISON (多波長天文画像検索・閲覧システム, 図 2) を開発している。さらに、どの VO 計画も言及している学校・市民に対する教育普及の有効な手段という意味でも、日本が VO に取り組む意義は大きいといえる。

国際的な VO 構築の流れの中で注意しておかなければならないことの一つに、データ利用権の問題がある。米国はまだ態度を明確にしていないが、欧州はそのパートナーシップポリシーの中でヨーロッパ諸国およびこの計画の遂行において関係の深いアメリカ、カナダ、オーストラリアに成果を公開するとうたっている。実際、ESO のデータアーカイブは現在 ESO の加盟国およびチリのみ公開が行われている。欧州はただし、アーカイブへのデータ提供も含めて、直接間接を問わず欧州の VO 計画への寄与があれば、その国をパートナー国として同様にアクセス権を与えるとも述べている。

日本の天文学的成果を世界に還元し、さらに世界のデータを最大限に活用して研究を進めていくためには、世界の VO 計画に積極的に参加していくことは必須と考えられる。VO は一朝一夕にしてできるものではなく、日本版 VO の構築にあたっては、アーカイブのための基本的なデータの整備はもちろん、関係する技術のサーベイなど事前の取り組みが必要不可欠である。早急な対応が望まれるところである。

参考文献

- 1) Szalay A., Gray J., 2001, *Sci* 293, 2037
- 2) Proc. of Virtual Observatories in the Future, 2001, ASP Conf. Ser. Vol. 225, eds. Brunner R.J., Djorgovski S.G., Szalay A.S.
- 3) Proc. of MPA/ESO/MPE workshop, Mining the Sky, 2001, eds. Banday A.J., Zaroubi S., Bartelmann M. (Springer, Berlin)
- 4) Proc. of Astronomical Data Analysis Software and Systems X, 2001, ASP Conf. Ser. Vol. 238, eds. Harnden Jr. F.R., Primi F.A., Payne H.E.
- 5) Proc. of SPIE 4477, Astronomical Data Analysis, 2001, eds. Starck J., Murtagh F.D.
- 6) Astronomy and Astrophysics in the New Millennium, 2001 (National Academy Press, Washington D.C.)
<http://www.nap.edu/books/0309070317/html/>
- 7) Wenger M., et al., 2000, *A&AS* 143, 9
- 8) Ochsenbein F., et al., 2000, *A&AS* 143, 23
- 9) Bonnarel F., et al., 2000, *A&AS* 143, 33

Virtual Observatory Projects

Toshihiro HORAGUCHI

Department of Science and Engineering, National Science Museum, Tokyo, 169-0073

Abstract: "Virtual Observatory" (VO) is a project which aims to federate data archives of various wavelength along with catalogue data and bibliographic data, etc. The project also intends to develop software tools and transparent interfaces for data query and analysis in order to establish new style of astronomy that mines the digital sky. VO initiatives that are planned at several nations are reviewed, and the contribution of Japan to the forthcoming international VO is discussed.