

データアーカイブの現状と国内外の データセンター

大石 雅 寿

〈国立天文台 天文データセンター 〒181-8588 東京都三鷹市大沢2-21-1〉

e-mail: masatoshi.ohishi@nao.ac.jp



世界中で作成されている天文観測アーカイブは無数にある。これらの多くは、主として先進国に存在するデータセンターに収集され、世界中に公開されている。また約10年前より、世界のデータアーカイブやデータセンターを仮想的に統合するレイヤを提供するヴァーチャル天文台が構築された。本稿では、データアーカイブやデータセンターの現状をレビューするとともに、天文学におけるデータ共有機構が他の科学分野におけるデータ共有機構構築の参考とされていることを述べる。

1. 特集「データベース天文学」に あたって

「今回観測した銀河について他の波長での画像が必要だ。」「思いがけない観測結果が得られた。この天体の性質を調べるために観測プロポーザルを出さないか。」「星形成領域で水メーザーやメタノールメーザーが存在するかどうか、観測してみたい。」天文学研究の現場ではこのような会話が日々交わされ、観測プロポーザル作成のための議論が行われる。

もちろん観測データを得るために観測プロポーザルを作成するのは当然だ。だが、もし必要なデータがすでに取得されているとしたら？ 観測プロポーザルを提出し、審査結果を待ち、観測を実施しなくても、そのデータを二次利用するだけで実施したい観測的研究がすぐに推進できるのではないか！

天文データベースを活用した天文学研究手法は、データベース天文学あるいはデータ活用型天文学 (Data-Intensive Astronomy) と呼ばれ、観測、理論、シミュレーションに次ぐ、第4の研究手法と言われている。データベース天文学を推進

するためには、研究に必要となる、時には多種類また大量のデータを容易に発見し、入手することが、まず、必要となる。さらに、入手したデータを容易に統計的な観点も含めた解析ができる環境が提供されていることが望まれるようになる。

データベース天文学という言葉が発明される以前から世界の主要な天文台は、それぞれが保有する望遠鏡によって得られた観測データをアーカイブするとともに、アーカイブをデータベース化して公開してきた。特にスペースからの観測となる衛星プロジェクトの場合、衛星から送信された生データをプロジェクト側が較正処理し、science-readyとしたデータを公開することを通じてその科学的成果を最大にする努力が払われている。地上望遠鏡によるデータアーカイブの公開も行われるようになり、天文データリソースは格段に増加した。一方、リソースの増加は必要なリソースを探すことに困難をもたらすようになった。また、多波長のデータを扱おうとすると、それぞれの観測装置の特性を踏まえた較正処理を一人の研究者が行うことが難しくなる。この困難さの解決を支援する研究支援環境として、ヴァーチャル天文台 (Virtual Observatory; VO) は、世界の天文デー

タリソースへの共通手法によるアクセスを実現し、データベース天文学を推進するために、世界の主要国が共同して構築を進めている。

以上の観点をもった特集は、2002年に組まれた（天文月報2002年6月号、7月号、12月号）。それから10年以上経過し、天文データアーカイブの種類も増え、また、ヴァーチャル天文台を用いた研究論文も世界で多数出版されるようになった。しかしながら、海外に比べて日本では、徐々に増えてはいるものの、アーカイブデータを利用する頻度がまだ少ない。本特集は、太陽データを除く天文データアーカイブをめぐる最新情報や成果をコンパクトにまとめ、天文研究者の皆さんにさらにデータアーカイブなどを活用していただくことを目的としている。

今回の特集は2回の天文月報にわたる。本号が第1回目にあたり、本稿および白崎裕治氏による「公開天文データの活用方法」から構成される。第2回目は次号を予定しており、「データサービスの利用例と成果」に関しての複数の記事、そして「データベース天文学の将来像」についての記事から構成されることとなっている。

2. データアーカイブやデータベースの歴史と現状

2.1 歴史的アーカイブの例

後冷泉院・天喜二年四月〔五月〕中旬以後の丑の時、客星觜（し）・参（しん）の度に出づ。東方に見（あら）わる。天関星に孛（はい）す。大きさ歳星の如し。

これは藤原定家による有名な明月記の一節で、西暦1054年に出現した超新星SN1054を目撃した貴重な記録である。SN1054が、現在、カニ星雲として知られていることは本稿読者の方であれば誰でも分かるであろう。明月記は、望遠鏡発明前の超新星記録7件のうち、3件（1106年、1054年、1181年）が記録されている、歴史的に見て

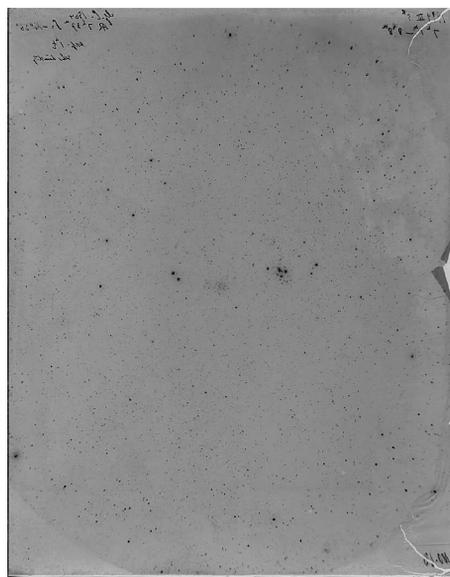


図1 国立天文台で発見された日本最古の星野写真乾板。1899年3月5日にとも座付近を1時間7分露出したもの¹⁾。

も、極めて貴重な天体现象のアーカイブ情報と言えよう。

写真術の発明が、天体现象について客観的な記録を可能にし、これにより天体物理学が成立したことはよく知られている。天文観測データは、いわば、観測時点での宇宙のあるスナップショットであり、実験によって再現することはほぼ不可能である。写真乾板は、こういった天体现象を記録するため、約100年前から天文学者たちに広く利用されるようになった。図1は、国立天文台天文情報センターが保管している日本で最古の星野写真の乾板である。天体现象が貴重であることを理解していた天文学者には、アーカイブを構築・保持していかうとする素地が元々あったのである。ある程度シニアな天文学者は、一度は“パロマチャート”にお世話になったものである。パロマチャートは、米国のパロマ天文台の1.2 mシュミット望遠鏡によって得られた北天での可視光写真乾板であり、そのコピーが主要な天文台には備え付けられてガイディング・チャートとして活用

されたものである。パロマチャートの乾板に、南天をカバーしたアングロオーストラリア天文台（オーストラリア）の1.2 m UKシュミット望遠鏡によって撮影された乾板を加え、共にスキャナーによってデジタル化したものが、DSS (Digitized Sky Survey)²⁾である。

2.2 現代天文学におけるアーカイブ

現代天文学では、可視光域のみならず、 γ 線から電波にわたるすべての電磁波を研究対象としている。目に見えない波長域では、必然的に検出器からの出力をコンピューターに取り込んでデジタルデータとして処理することとなった。特に衛星観測では、早くから観測データのデータベース(DB)化が進んできた。天文衛星プロジェクトでは、衛星打ち上げ費用が極めて高額であることを踏まえ、多くの場合、**研究者コミュニティによる科学研究の成果を最大にするために観測データ公開するという前提でプロジェクトがデザインされる**。そして研究者コミュニティへのデータ公開は、後述するデータセンターを介して電子的(Webやftpを利用)に公開されている。

一方、衛星データのDB化に比べて、地上観測データのDB化はだいぶ遅れていた。これは、地上望遠鏡に搭載された観測装置の安定性や天候にデータの品質が左右されることが主たる理由であった。しかし、先に述べたように古くから写真乾板で撮影した画像は、例えば、パロマチャートとして利用されてきた。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(いわゆる3Cカタログなど)、1.4 GHz帯³⁾や4.85 GHz帯⁴⁾での連続波源サーベイデータが作成された。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⁵⁾が作成された。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仕組みを当初から備えているものではなかった。検出器技術も年々進歩していたため、他の研究者が取得したデータを融通してもらうよりも、新規に観測を実

施したほうが良質のデータが得られていたという背景もある。

半導体技術とネットワークの発達がこの状況を大きく変えた。大容量のデータベースを容易に扱えるようになっただけでなく、高速に遠隔地から天文データベースにアクセスすることが可能となったのである。例えば全天をCCDでサーベイするSDSS計画⁶⁾では当初から観測データをDB化し、観測プロジェクトチームが品質チェックをしたうえで、データを公開している。検出器技術も成熟し、**二次利用しないのがもったいないほどの品質のデータがかなりの量、取得できるようになっている**。このため、白崎氏の記事⁷⁾で個別のデータアーカイブ/データベースが紹介されているように、今では、数え切れないほどの天文データベースにアクセスできる時代になっている。

これらに加え、論文から採録したのものも含めたカタログデータベースが多数作成されてオンライン公開されている。後述するCDS (Centre de Donnees Astronomiques de Strasbourg) が作成・運用しているVizieR⁸⁾は、代表的なカタログデータベースサービスである。CDSでは、構築したデータベース資産を活用し、天体名と座標の対応付けサービス、また、天体名と関連する文献情報を表示するSIMBAD (Set of Identifications, Measurements and Bibliography for Astronomical Data)⁹⁾ サービスも提供している。オンラインで天体名から座標を求めるアプリケーションの多くは、バックグラウンドでSIMBADを用いているため、読者の多くもSIMBADを知らず知らずのうちに使っているだろう。

さらに、文献DBとして最も有名、かつ、役立っているものはADS (Astrophysics Data System)¹⁰⁾であろう。今や、天文学者は、研究を進めるために必要な文献情報を、まずADSにアクセスして取得することが当たり前となっている。ADSは、NASAが運営資金を提供し、後述するIPAC (In-

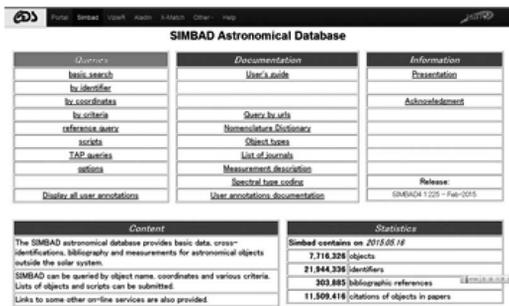


図2 天体同定, カタログ, 文献ツールであるSIMBAD.

frared Processing and Analysis Center) が運営しているものである。ADSの開発スタッフと何度か話をしたことがあるが、「ADSを使うことがあまりにも当たり前過ぎるので誰も謝辞を書いてくれない」と苦笑していた。データベースの開発や運用は縁の下の力持ち的存在であるので、利用者がたくさんの謝辞を書くことが、そのサービスをいつまでも利用できるようにするための重要な手段となる。ADSを利用して執筆した論文には、是非とも、“*This research has made use of NASA’s Astrophysics Data System*” と記載したいものである。

3. 国内外のデータセンター

3.1 日本のデータセンター

日本では、主として地上観測を担う国立天文台と科学衛星ミッションを推進する宇宙科学研究所が大きなデータセンターを擁し、日本のみならず世界の天文学コミュニティに観測データを提供している。また、いくつかの大学でもユニークなデータベースを構築して公開しており、世界から高く評価されている。

3.1.1 国立天文台

まず、日本の天文学のナショナルセンターである国立天文台が提供するデータサービスからみていこう。

国立天文台では、従来、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; 現在はバージョンアップしたSTARS2), すばる望遠鏡による公開生データ等へのアクセスを提供するSMOKA (Subaru-Mitaka-Okayama-Kiso Archive), 野辺山宇宙電波観測所の45 m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などが運用されてきた。これらに加え、今回は詳しくは触れないが、太陽物理学のために、ひのでサイエンスセンターと太陽観測所が太陽データを世界に公開してきた。

最近では、南米チリで観測を開始したアルマ望遠鏡の観測アーカイブが、アルマ地域センターとJVO (Japanese Virtual Observatory) から公開されている。JVOからはアルマ望遠鏡の公開データをFITS形式でQuick-Lookしつつダウンロードできるようになっている。この機能は、アルマ望遠鏡を共同運用する欧米でもまだ提供されていない先進的なデータサービスとなっている。

すばる望遠鏡に搭載された次世代主焦点広視野カメラであるHyper Suprime-Cam (HSC) の観測が始まっている。その共同利用観測生データは、ハワイ観測所のデータ公開方針に従い、2015年9月から順次SMOKAから公開される予定となっている。

国立天文台が公開しているデータリソースのURLを表1にまとめた。

SMOKAは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) を原型としてすばる望遠鏡のデータをも包含するものである。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これまで野辺山太陽電波観測所にすべて蓄積されて全世界の太陽研究者が利用していた。野辺山太陽電波観測所は2015年3月末で閉鎖されたが、ラジオヘリオグラフのデータは天文データセンターを経由して公開され続けることとなっている。

国立天文台天文データセンターでは、海外の研

表1 国立天文台のデータリソース.

リソース	URL
JVO	http://jvo.nao.ac.jp/
STARS	https://stars2.naoj.hawaii.edu/
SMOKA	http://smoka.nao.ac.jp/
nrodb	http://nrodb.nro.nao.ac.jp/
ALMA Archive	http://almascience.nao.ac.jp/alma-data
VERA	http://veraserver.mtk.nao.ac.jp/index-J.html
Hinode	http://hinode.nao.ac.jp/index.shtml
太陽観測所	http://solarwww.mtk.nao.ac.jp/jp/solarobs.html
RISE	http://www.miz.nao.ac.jp/rise-pub/

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, などが含まれ、いずれもオンラインで検索することができる。また同センターでは、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ウェブからダウンロードできるようになっている。なお、天文データセンターが提供する、データベースやデータセンターを仮想的に統合して天文データ利用を容易にするVOについては、章を改めて紹介することとする。

さて、これらのデータリソースへのURLを一つ一つ記憶するのはたいへんである。そこで天文データセンターに、データサービス/データ解析ポータルが設置された(図3)。ここにアクセスすることによって、国立天文台が提供するすべての公開データサービスに到達することができる。このように、国立天文台は、それ自身が一つの大きなデータセンターとして機能し、また、さらに発展させようとしている。

3.1.2 宇宙航空研究開発機構宇宙科学研究所

ISASには、観測データやテレメトリデータのDB化およびその運用を行うための組織(科学衛星運用・データ利用センター: Center for Science-Satellite Operation and Data Archive (C-SODA))がある(<http://csoda.isas.jaxa.jp/>)。C-SODAでは、Suzaku, ASCA, Hinode, AKARIと



図3 国立天文台データサービス/データ解析ポータル(日本語版). <http://www.adc.nao.ac.jp/Portal/naojportal.html>同様のデザインの英語版も用意されている。

いった天文衛星、Geotail, Akebonoといった太陽地球惑星間物理のための衛星、HAYABUSAやKAGUYAといった惑星物理学のための衛星が生み出したデータをオンラインで公開している(DARTS; <http://darts.jaxa.jp/>)。AKARIは豊富なデータを公開しており、ポインティング観測、カタログアーカイブ(CAS)、マップデータカタログが公開されている。図4に、2015年3月に公開を開始したマップデータカタログの検索画面を示す。

3.1.3 国内大学によるDB

東京学芸大学ではDSSによって得られた画像を処理して暗黒星雲の全天アトラスの作成を行い、オンラインで公開している。図5に、暗黒星雲データベースのトップ画面を示す(<http://darkclouds.u-gakugei.ac.jp/index.html>)。この暗黒星雲アトラスは研究者だけではなく研究者ではない人々にも利用されるような工夫がなされてい

All-Sky Map Search

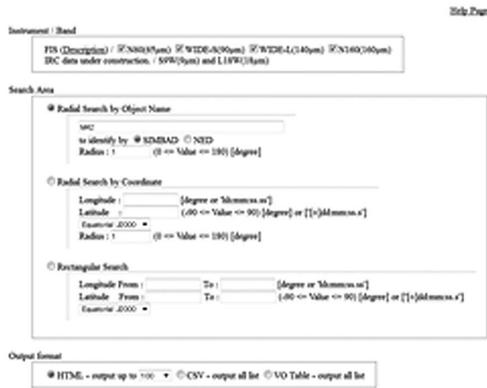


図4 2015年3月に公開を開始したマップデータカタログの検索画面。

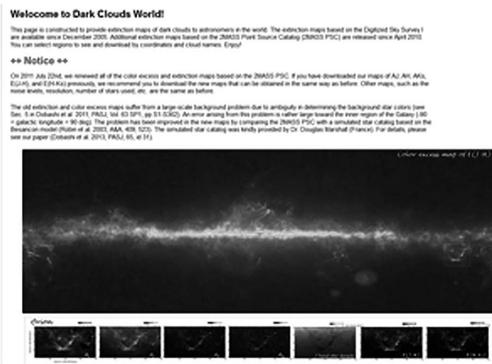


図5 東京学芸大学が制作、運営している暗黒星雲データベース。暗黒星雲博物館や暗黒星雲アトラスへのリンクもある。

る。東京学芸大学ではこの暗黒星雲アトラスに基づき、スターカウント法によって計2,448個の暗黒星雲カタログを作成した¹¹⁾。

北海道大学と東京大学では、ユニークな金属欠乏星データベース SAGA (Stellar Abundances for Galactic Archeology Database) を構築、運用している (<http://saga.sci.hokudai.ac.jp/wiki/doku.php>)。SAGAについては、本特集の第2回で須田拓馬氏が紹介する予定となっているので、本稿ではその詳細は割愛する。

3.2 世界のデータセンター

各種DBは、さまざまなデータセンターによっ

て開発・運用されている。

3.2.1 ストラスブール天文学データセンター

ストラスブール天文学データセンター (Centre de Donnees Astronomiques de Strasbourg; CDS) 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、恒星の位置などを与えるDBである SIMBAD, 論文から抽出したカタログサービスである VizieR を運用しているのもこのセンターである。さらにCDSでは、電子化されたジャーナルに掲載できないほどの大きな表の公開業務を引き受けている。本稿読者の方々の多くが、電子ジャーナルに誘導される形でCDSにアクセスした経験があると思われる。

3.2.2 NASA 国立宇宙科学データセンター

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ともに、各種天文カタログの収集・頒布を上記のCDSと協力して進めている。

3.2.3 NASAゴダード宇宙飛行センター

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、ほとんどすべ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。

3.2.4 赤外線観測処理解析センター

赤外線観測処理解析センター (Infrared Processing and Analysis Center; IPAC) は、元来、IRAS衛星による観測データの収集・標準化・頒布を行うためのデータセンターである。しかし現在では、GALEX, Herschel, Planck, Spitzer, Keplerをはじめとするさまざまなミッションによるデータを公開している。将来的には、開発中あるいは建設中である Euclid, LSST, TMT のデータ

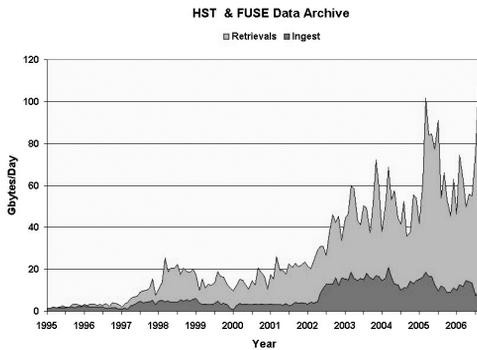


図6 2007年までのハッブル宇宙望遠鏡とFUSEデータの利用統計。

公開をも担う予定である。

3.2.5 宇宙望遠鏡科学研究所

宇宙望遠鏡科学研究所 (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。HSTのデータは、世界中から利用されている。やや古いグラフだが図6にHSTとFUSEの利用統計を示す。アーカイブされるデータ量の4-5倍のデータがダウンロード・利用されていることが明瞭に分かる。STScIの年次報告によれば、年間ダウンロード量は約32 TBである¹²⁾。

3.2.6 欧州スペース天文センター

欧州スペース天文センター (European Space Astronomy Centre; ESAC) は、欧州宇宙機関 (European Space Agency; ESA) の地上セグメントに対応する。ESACではESAが運用する各ミッションの運用を行うとともに、各ミッションが送ってくるデータを受信し、評価をし、必要な処理を施し、science-readyなデータとして研究者に提供している。ESAは天文ミッションだけではなく惑星科学ミッションも実施しているため、ESACでは天文学、惑星科学、これに加えて情報科学の専門家が多数集結している。ESACは、後述するヴァーチャル天文台のためのソフトウェア開発においても大きな貢献をしており、天文VO

で得た技術を惑星科学データを共有するための国際組織 (International Planetary Data Alliance; IPDA¹³⁾) における標準作成に応用している。

3.2.7 カナダ天文学データセンター

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。このため、カナダも参加していたハワイ島にあるサブミリ波望遠鏡であるJCMTのためのデータセンターとしても機能している。さらに最近では、カナダ国内に張り巡らされた天文学用高速ネットワーク網であるCanadian Advanced Network for Astronomical Research (CANFAR)¹⁴⁾のハブの一つとして、巨大なストレージサービスを提供している。

3.2.8 欧州南天天文台

欧州南天天文台 (European Southern Observatory; ESO) は南米にあるVLTなどの大型観測装置のホスト研究所であり、ESOに所属する観測装置が生み出すデータはESO Science Archive Facility (SAF) として広く世界に公開されている。SAFにはハッブル宇宙望遠鏡の観測データも含まれている。またSAFの特徴の一つとしてヴァーチャル天文台用のツール類も公開されていることが挙げられる。

3.2.9 ラザフォード・アップルトン研究所

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、ISOデータ頒布のセンターにもなっている。

3.2.10 LAMOST

中国国家天文台が建設を進めてきたLarge Sky

表2 世界の主なデータセンター。

データセンター	URL
CDS	http://cdsweb.u-strasbg.fr/
NASA NSSDC	http://nssdc.gsfc.nasa.gov/
NASA GSFC	http://www.nasa.gov/centers/goddard/home/index.html
IPAC	http://www.ipac.caltech.edu/
STScI	http://www.stsci.edu/resources/
ESAC	http://www.esa.int/About_Us/ESAC
CADC	http://cadwww.dao.nrc.ca/
ESO SAF	http://archive.eso.org/
RAL	http://star-www.rl.ac.uk/star/docs/sun162.htx/sun162.html
LAMOST	http://www.lamost.org/

Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) が、2015年3月に最初の data release (DR1) をアナウンスした。LAMOSTはデータセンターとは異なるものの、東アジアにおけるパートナーである国家天文台が公開データをリリースしたため記載することとした。本稿執筆時点で DR2 までアナウンスされている。

これらのデータセンターの URL を表2にまとめた。

4. 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う。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、前述したように、ハッブル宇宙望遠鏡 (HST) アーカイブで

は、データ取得量の4-5倍にも及ぶ量の観測データがダウンロードされ、天文学の研究のために利用されていることがよく知られている。

4.1 現状の天文データアーカイブの問題点

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イスもDBごとに少しずつ異なる。出力データ形式として一般に FITS (Flexible Image Transfer System)¹⁵⁾ 形式が採用されている。FITS形式では、1レコードを2,880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。このため、各DBではそれぞれの天文台や装置開発者が、FITSのヘッダ情報を比較的自由に定めることができた。その結果、各DBのメタデータ (ヘッダ情報のように、あるデータが付随してもつそのデータ自身についての付加的なデータ) 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ごとに必要となった。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勞する。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、場合によっては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。

5. データセンターやデータベースを仮想的に統合するヴァーチャル天文台

前章で述べたようにデータセンターや天文データアーカイブが世界の主要天文台等で構築されているにもかかわらずその活用のための環境が整っ

ていたとは言い難かった。一方、天文観測装置やシミュレーション用のスーパーコンピュータも次第に規模が大きくなり、必然的に巨額の費用が必要となってきた。予算規模が大型化するにつれて予算を捻出した社会（国家）への説明責任度も高くなり、取得した観測データやシミュレーションからより多くの研究成果を生み出す必要が出ている。一般的に天体は多波長で放射をしているため、各種天体现象の本質を知るためには多波長データの活用が必要であることも周知の事実であった。天文学が社会に対して果たすべき責任という点からも、これらの難点を早急に解決する必要があった。

1990年代後半からの情報通信技術の急激な発展により、高速ネットワーク環境が容易に利用できたり高機能な計算機が安価に購入できるようになった。これらの**情報通信技術を利用すれば世界中の天文アーカイブを連携し研究に必要な観測データを容易に収集し解析することが可能になる**だろうという自然な発想が世界各地で独立に沸き上がった。これが**ヴァーチャル天文台（Virtual Observatory; VO）**構想である。VOはデータベースやデータセンターの存在の下に、それらを見かけ上統合するレイヤーを被せ、研究者等が必要なデータを容易に発見・ダウンロードし、また、各種VO対応ツールを用いて研究するためのインフラストラクチャーなのである。よく誤解されるのであるが、VOにアクセスするだけで夢のような研究ができるのではなく、VOを利用する研究者のアイデア次第で素晴らしい成果に結びつくのである。

ヴァーチャル天文台はまさに発展中の天文研究基盤であり、それを支える技術などは日々進化している。本章ではヴァーチャル天文台の黎明を振り返り、それを実現するための技術を概観し、世界各国の天文研究資源を共有するための標準化活動に触れ、どのような新しい研究基盤が構築されるかについて述べる。

5.1 ヴァーチャル天文台構想の黎明

World Wide Web（WWW）はいまや多くの人々の生活を支える基本的な情報技術となっている。WWW技術はまさにPCの中に世界のすべてが入っているかのような錯覚を覚えさせる。WWWの基本はその透明さ（transparency）にあると言って良い。httpプロトコルという標準を使用している世界のどこにでも一瞬にしてアクセスすることができる。

ヴァーチャル天文台は天文データや天文計算資源に関してWWWと同様の姿を構築しようとするものである。2001年から2002年にかけて世界の複数の研究グループが

- ・国際規模の天文研究プロジェクトはどうやったらデータ量の増大と計算量の増大に対応できるのか？
- ・今後国際協力研究を進めるためには各種観測装置で取得した多波長データや解析施設を研究チーム内で共有する必要があるが、それをどうやって実現するのか？
- ・巨費を投じた国際プロジェクトでは、プロジェクト関係者だけではなく世界の天文学研究者による研究を通じて科学的成果を最大化する社会的責任を負っているが、どうやれば実現できるか？

といった問題について独立に検討をしていた。彼らはWWWが標準技術の確立によって構築されていることを参考にし“世界中のすべてのデータを研究者のPCに格納”するために各国の同様のグループ間で共通のインターフェイス、アクセスプロトコルを策定することによってその世界を実現することを目指した。世界中の天文台などが共通の検索データアクセス言語を受け入れれば世界中のどこにしようとも各種データに同じ方法でアクセスできる。同様に、データ解析サーバーが同じインターフェイスをもっていれば、世界のどこからでも同じ方法によってデータ解析依頼をすることができる。データサーバーやデータ解析サー

バーからの出力結果が標準形式で記述されていれば、結果の解釈やその後の処理も容易となる。観測データだけではなく、遠隔地の記憶装置（ハードディスクなど）や解析に用いた処理手順（ワークフロー）も共有できれば、解析結果の再現が容易となる。

ヴァーチャル天文台で共有するものは主としてデータであるため、情報学の研究者は、ヴァーチャル天文台はデータグリッドの例と呼ぶこともある。しかしヴァーチャル天文台では単にデータ共有するだけでなく、その処理や、結果の共有も目指しているため、いわゆるデータグリッドよりも未来志向のインフラストラクチャーでもある。現在はデータ解析は研究者の手元の計算機もしくは所属大学等の基盤センターが用意している計算資源を用いるのが主体であるが、ヴァーチャル天文台でデータを保有している“どこの解析サーバー”でも利用できることを目指している。

5.2 ヴァーチャル天文台の概念

図7にヴァーチャル天文台を構成する要素間の関係図を示す。

利用者はVOに対応したポータルサイトやツールを用いる。利用者の要求はポータルやツールによって解釈され、図の中程にあるVOインターフェース層を介して、図の最も下に書かれているデータを格納しているHDDや解析のためのCPUにその要求が渡される。CPUやHDDといった計算資源に届いた要求に合致するデータや解析サービスの結果も、やはりVOインターフェイスを介して利用者ポータルやツールに返される。もちろん利用者が予めリソースのURLなどを知っている場合（従来のデータベースや解析サーバー利用法）は、従来どおり直接アクセスすれば良い（図の右側にある垂直線に対応）。

こういったインフラストラクチャーは、各国の主要天文学研究機関がそれぞれが資源を供出して構成される。機能としては重複する部分が発生するが、それによりヴァーチャル天文台全体

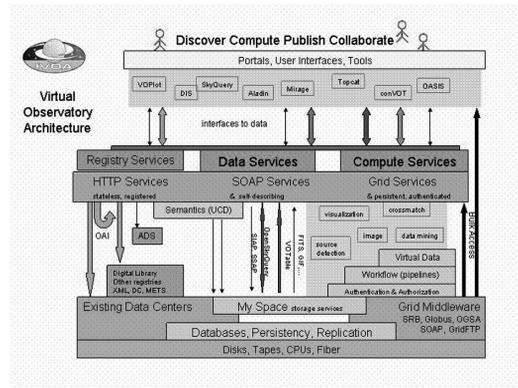


図7 ヴァーチャル天文台を構成する要素間の関係図（研究者などの利用者を最上位に据えた）。

としての安定稼働が保証される。すなわち、各研究機関は独立に計算資源の保守・維持・機能向上作業を行うことが可能となる。もしこのインフラストラクチャーが世界に一つしか存在しないとしたら、それが一時的にせよ機能停止した場合の影響は大きい。したがって、ヴァーチャル天文台を実現するためには標準プロトコル群を定め、それらの活用による相互運用性の確保が極めて重要となる。

こういったVO Frameworkを構築するために2002年に結成されたのが国際ヴァーチャル天文台連合（International Virtual Observatory Alliance <http://www.ivoa.net/>）である。日本もその当初から参加し、積極的な活動を展開している。IVOAには、2015年現在、21の国と地域（EU）を代表するVOプロジェクトが参加して相互運用が可能な標準プロトコルを制定している。

その標準は、以下の観点を満足させようとするものである。

- 国際的な合意に基づく、標準プロトコルの制定
- 各VOプロジェクトの「糊」となるソフトウェアの構築：ポータル、レジストリー（各リソース情報データベース）、ワークフロー、利用者認証機構、遠隔ストレージなど

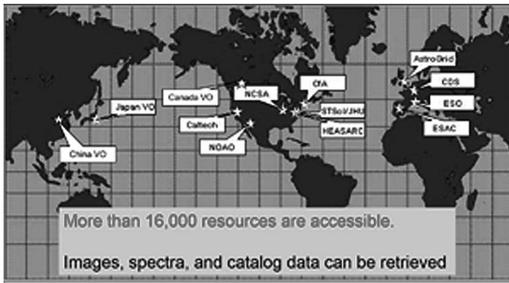


図8 ヴァーチャル天文台により仮想的に統合された世界のデータセンター。

- ・各国に存在するデータセンターが世界にデータ発信したい場合、すなわち、VO対応サービス立ち上げの助言
- ・データ解析ツールの開発
- ・レジストリおよび利用者支援システムの立ち上げと保守

なお、本稿では触れていないVOプロトコルやVO技術の詳細を知りたい方は、IVOAのサイトを訪問することをお勧めする。

これらの標準化活動の結果、2015年現在、日米欧の主要な天文台やデータセンター（日本：国立天文台、米国：CfA, STScI, HEASARC, NCSA, NOAO, CalTech, 欧州：AstroGrid, CDS, ESO, ESACなど）がVOインターフェイスを通じて相互に接続され、16,000以上の天文リソースが利用できるようになっている（図8）。このVOを用いた研究成果については、本特集第2回目において紹介する予定である。

6. オープンデータとデータシェアリングを巡る世界の動向

研究データをアーカイブ、データベース化し、研究コミュニティ全体で活用しようという考えは、当然、他の分野にもある。一例としてタンパク質研究分野のデータベースを紹介する。タンパク質の研究分野では、タンパク質を構成するアミノ酸配列に関するデータをProtein Data Base (PDB)¹⁶⁾（蛋白質構造データベース）に登録する



図9 大阪大学に設置されている日本のPDB。

ことが半ば義務づけられている。そして驚くべき事に、タンパク質研究者は、PDBへの登録後でなければ実験によって得たアミノ酸配列に関する研究論文を受理されないのである。タンパク質のアミノ酸配列情報は、創薬研究に結びつくことがあり、社会への貢献が大きく期待される。さらに、創薬研究には膨大な費用がかかるため、既に分かっているタンパク質構造を独立に調査することは経費と時間の無駄につながる。このためPDBは、米国、欧州と日本の3カ所に登録ノードをもち、天文VOと同様なアイデアの下、相互のデータを共有する仕組みを構築し、運用されている。図9に日本で運営されているPDB (PDBj)のトップ画面を示す。

国際科学会議（International Council for Science; ICSU)¹⁷⁾は、IAUもその構成メンバーとなっている世界の国際的科学組織体の連合である。ICSUは従来、“データをオープンに使う”との信念の下、科学データを世界に公開するためにWDC (World Data Centers)を運営してきた。WDCに参加していた大学や研究機関は、ほとんどが地球物理分野からであり、これにいくつかの天文分野の研究機関が加わっていた。国立天文台の太陽関係のデータや、フランス・ストラスブールのCDSもWDCのメンバーであった。一方、WDCは、個々のデータセンターの緩い集合体であったため、その運営方針などは個々のデータセンターにお任せ状態であった。このためICSUは、一つの運営方針の下でさまざまなデータを世



図10 World Data Systemのトップ画面. その本部は、日本の小金井市に設置されている。

界で共有可能にするために、2013年にWDCを発展的に改組してWorld Data System (WDS)¹⁸⁾を設置した(図10)。2009年から2011年にかけて、WDSの運営方針を定めるために科学的な観点から勧告¹⁹⁾を作成するad hoc委員会に私も参加することとなった。そのad hoc委員会の中では、天文学(IVOA)、タンパク質(PDB)、海洋データのデータベースやIPCCが先進例として認識され、勧告作成の基盤となったことが特筆される。

このような世界の動きを受けて、日本を含むG8の科学担当大臣会合が、2013年6月にロンドンで開催された。この会合の目的は、**主要国政府が音頭を取って各種データをオープンに活用する政策を進めよう**という点に関する合意を形成することであった。G8科学担当大臣会合の結果、G8 science ministers statement: London, 12 June 2013²⁰⁾が採択され、主要各国はそれぞれがこの合意に基づく**オープンデータ政策**を推進することとなった。G8合意を受けて欧州では、Research Data Alliance (RDA)²¹⁾をさっそく立ち上げ、科学関係者のみならず、産業関係者も交えてオープンなデータ利用を促進するための検討を急ピッチで進めている。

欧州の動きに押され、日本政府も2014年度後半から動き出した。日本政府は、内閣府総合科学

技術会議の下に、2014年12月に「国際的動向を踏まえたオープンサイエンスに関する検討会」²²⁾を設置し、オープンデータやそれを活用するオープンサイエンスを推進するための基本方針が、「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書として取りまとめられた。報告では、研究成果(論文や研究過程で生成されたデータ等)を、特定の研究領域だけに閉じることなく、広く一般へも公開し、これらを効果的に活用し、科学研究を進めていくスタイル(オープンサイエンス)を進めていくことを期待している。これを受けて、文部科学省をはじめとする各省庁でも、その傘下の研究機関も巻き込む形で、オープンデータ/オープンサイエンスを推進するための政策を取りまとめる検討が急ピッチで始まっている。この動きは、確実に天文学研究にも影響してくると思われ、注視が必要である。

7. ま と め

本稿では、天文学におけるデータアーカイブやデータベースの利用に関する歴史や現状を概観した。ある地上の望遠鏡から観測できる天域が限定されるため、他の天文台で観測する必要性があったため、天文学では早くからデータ形式の統一(FITS)がなされた。やむにやまれぬ事情のためではあったが、これが幸いし、天文学はオープンデータ利用の先進分野と言われている。最新の望遠鏡によって得られる観測データは高品質かつ大量であり、観測を実施した研究者とは異なる目で見ることによりさまざまな知見を引き出すことが可能である。即ち、天文アーカイブは宝の山なのである。

しかし、欧米諸国に比べ、日本をはじめとする東アジア諸国における天文データアーカイブ利用度はまだそれほど高くない。最近公開されているアルマ望遠鏡アーカイブへのアクセス頻度統計によれば、東アジア諸国からのアクセスは全体の10%にも満たない。すばる望遠鏡に設置された

HSCのデータも2015年度後半から公開が始まることとなっているし、近い将来には、Data Machineと称されるLSST²³⁾のデータも公開されるようになる。多様なデータがどんどん出てくる時代にはデータを活用した天文学研究が盛んになることは明らかである。

本特集がきっかけとなって、皆さんの天文データ利用度が高まり、アーカイブを活用した研究成果がたくさん出ることを期待し、筆を置くこととする。

参考文献

- 1) 佐々木五郎ほか“国立天文台アーカイブ室の活動(3): 日本最古の星夜写真の発見”, 日本天文学会2012年春季年会Y32a
- 2) Lasker B. M., 1994, American Astronomical Society Meeting 184, 3501
- 3) Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ 115, 1693
- 4) Gregory P. C., Scott W. K., Douglas K., Condon J. J., 1996, ApJS 103, 427
- 5) Price S. D., Walker R. G., 1976, Interim Report Air Force Geophysics Lab., Hanscom AFB, MA. Optical Physics Div.
- 6) SDSS Project <http://www.sdss.org/>
- 7) 白崎裕治, 2015, 天文月報108, 498
- 8) Ochsenbein F., Bauer P., Marcout J., 2000, A&Ap 143, 23
- 9) Wenger M., et al., 2000, A&Ap 143, 9
- 10) Dobashi K., et al., 2005, PASJ 57, S1
- 11) NASA Astrophysical Data System <http://adswww.harvard.edu/>
- 12) MAST Annual Reports https://archive.stsci.edu/reports/annual_report.html
- 13) The International Planetary Data Alliance <https://planetarydata.org/>
- 14) CANFAR Project <http://www.canfar.net/>
- 15) Wells D. C., Greisen E. W., Harten R. H., 1981, A&Ap (Supplement Series) 44, 363
- 16) Protein Database <http://pdb.org/>
- 17) International Council for Science <http://www.icsu.org/>
- 18) World Data System <https://www.icsu-wds.org/>
- 19) <http://www.icsu.org/publications/reports-andreviews/strategic-coordinating-committee-on-information-and-data-report>
- 20) [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8 Science Meeting Statement 12 June_2013.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf)
- 21) <https://rd-alliance.org/>
- 22) 国際的動向を踏まえたオープンサイエンスに関する検討会 <http://www8.cao.go.jp/cstp/sonota/openscience/>
- 23) http://www.lsst.org/lsst/science/concept_data

Current Status of Data Archives, and Data Centers in Japan and the World

Masatoshi OHISHI

Astronomy Data Center, National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

Abstract: There are numerous astronomical observational archives in the world. Most of them are collected in Data Centers which are build and operated by developed countries, and are published to the world. This article reviews the current status of astronomical data archives and data centers worldwide. Further I will stress that the astronomical “data-sharing” infrastructure has been regarded one of successful cases, which can be referred to as a model for other science areas in building their own data-sharing mechanism.