

天文学連携データベースシステム (ヴァーチャル天文台)の開発・ 計算機資源の国際連携機構

A Mechanism in Federating Internationally
Distributed Databases and Computing
Resources to Realize Virtual Observatories

本田 敏志^{*1} 大石 雅寿^{*1} 白崎 裕治^{*2}
田中 昌宏^{*1} 川野元 聡^{*2} 水本 好彦^{*1}

Satoshi HONDA Masatoshi OHISHI
Yuji SHIRASAKI Masahiro TANAKA
Satoshi KAWANOMOTO Yoshihiko MIZUMOTO

天文学では、線から電波まで様々な波長の観測データベースを利用した研究が重要となってきた。そのため、世界中の天文台に分散配置された様々なデータベースを連携し、統合的に扱えるシステム(VO)の開発が世界中で進められている。我々はVOのプロトタイプを製作し、天文データベースを統合的に扱うための検索言語やXMLメタデータ検索機能などを設計、実装してきた。新たに開発した実用化を目指したプロトタイプ3では、他国VOとの連携を目標として標準プロトコルの制定を進め、すでに一部のVOとは接続に成功している。本論文ではプロトタイプ3の機能を紹介するとともに、実際に他国VOと接続しての天文研究への応用例を紹介する。

In astronomy it is crucial to utilize multi-wavelength databases to investigate astronomical objects. Since VO systems can provide seamless access to astronomical distributed databases, construction of VO (Virtual Observatory) is being done all over the world. We have built a prototype of VO, and have developed a language to query distributed databases, an XML metadata search function, and so on. In the prototype 3 we established standard protocols for the interoperation with other VOs, and we have already succeeded to interconnect with some other VOs. This paper describes of JVO prototype 3 in detail.

1. はじめに

近年の天文観測は様々な観測装置の開発によって、可視光の観測だけでなく、ガンマ線から電波まで広い波長域で観測を行うことが可能になり、対象となる天体の性質にさまざまな角度から迫ることができるようになった。さらに国立天文台のすばる望遠鏡をはじめとする大型望遠鏡の建設によって、これまで観測できなかった暗い天体のみならず、一度に

*1 正会員 国立天文台

honda.satoshi.masatoshi.ohishi.masahiro.tanaka.mizumoto.y@nao.ac.jp

*2 非会員 国立天文台

yuji.shirasaki.kawanomoto.satoshi@nao.ac.jp

多数の天体を観測することも可能になったため、非常に効率よく観測データが得られるようになった。

このようにして様々な観測装置や望遠鏡を使って得られた観測データは、観測者などによって画像処理などされた後に、科学的議論に乗せることができるデータとなる。処理された画像データやスペクトルデータは主に天文観測で標準とされているFITS[1]とよばれるバイナリ形式で各観測所などに保管され、引き出された科学的データについてはカタログデータとして論文発表される。各観測所に保管されている観測データは、他の波長域のデータや過去のデータと比較したりできるように、ほとんどがデータベース化され、ネットワーク上で公開されており、論文発表されたカタログデータもほとんどが世界中のデータセンターなどを通じて得ることができるようになってきている。

しかしながら、観測装置の多様化、大型化によって日々生産される観測データは膨大なものとなり、もはや従来の手法では処理できなくなってきている。例えば、すばる望遠鏡では毎年数十TBの観測データが生産され、建設中のALMAの場合は年に数PBにも及び、今後さらに加速して増加していくと考えられている。

このように、現在の天文学では世界中の観測所に分散配置されている膨大なデータベースから、いかにして必要な情報を迅速に検索し、取得するかということが重要となっている。バーチャル天文台(Virtual Observatory: VO)は近年著しい発展を見せているネットワーク技術やデータベース技術を使って、分散データベースをあたかも全天の、そして、全波長を包含する1つの巨大データベースであるかのように見せるためのシステムであり、今後の天文学の発展にとって必須のものと考えられている。そのためVOの開発は世界中で進められているが、世界のデータベースを連携するには各国の協力が必須である。このためにIVOA(International Virtual Observatory Alliance)[2]が各国のVO連携のために結成され、プロトコルの標準化などが議論されている。

我々は2002年度よりJVO(Japanese Virtual Observatory)の研究開発を進めてきた[3,4]。本稿ではこれまでに蓄積された技術を元に、世界の天文データベースの連携に成功したプロトタイプ3について述べる。

2. これまでのJVOの開発

JVOの開発は、限定した天文研究手法を実現するプロトタイプを開発することを通じて用いた技術の有効性や問題点などを洗い出し、実用システムに向けた基盤技術を蓄積していく手法で進めている。我々は過去に2つのプロトタイプを製作し、その評価を行った。

プロトタイプの第1版では、分散天文データベースを連携検索するためにSQLを天文学用に拡張したJVOQL(JVO Query Language)[5]を定義し、LAN上に分散配置したデータベースに対しGRID機能を用いて実際に検索実行することを目指して製作した。JVOQLは、SQLに複数のデータベースへのアクセスを同時に書けること、天体の位置を示す天球座標を扱えること、異なるデータベースに格納されている同一の天体の異なる波長情報を引き出す機能(クロスマッチ検索)、といったことができるように拡張をおこなったものである。これらは十分動作することが実証されたが、GRID機能を実現するために用いたGlobus Toolkit 2[6]は長時間ジョブの発行といった利用を想定しているため、結果を問い合わせる間隔が長く、VOにとって実用的ではないこと

が明らかになった[4]。

そこで、第2版ではGlobus Toolkit 3のGridServicesを活用することによって、ジョブ起動の待ち時間をほとんど無くすることに成功した。さらに、増加する天文データベースから必要なもののみを効率的に探し出すために、アクセスできる計算機資源のメタデータをXML形式で記述し、さらにこれらをXMLデータベース（カレアラ[7]）を用いたレジストリに登録する機能を持たせた。ユーザーインターフェースも改良し、SQLの知識が無くても検索が行えるJVOQLエディタを実装した。実際にすばるの観測データを使って、天文学研究の検索を行ったところ、非常に有効であり実用的なことが証明された[4]。

3. プロトタイプ第3版の開発

これまでの2つのプロトタイプの評価などを踏まえて、プロトタイプの第3版では海外VOとの連携を主な目標として開発を行った。海外VOとの接続が可能となれば、VOの目指す「いつでもどこからでもあらゆるデータを得ることができる」システムに大きく近づくことになる。

3.1 プロトタイプの構成

JVOシステムは主に以下の4つから構成される（図1）。

- ユーザからの検索や解析の指令を受け、結果を表示するユーザーインターフェース部
- 検索、解析指令を受け付け、メタデータを管理するレジストリサーバやデータサーバの処理を呼び出し、結果を返すコントローラ部（JVOポータル）
- JVOポータルからの検索サービスや解析サービス要求に対して処理を実行し結果を返すデータサーバ部
- メタデータの管理検索を行うレジストリサーバ部

システムの開発はマシン環境に依存しないJavaを用いており、分散配置されたデータベースなどを統合的に扱うためにWebサービスを利用している。

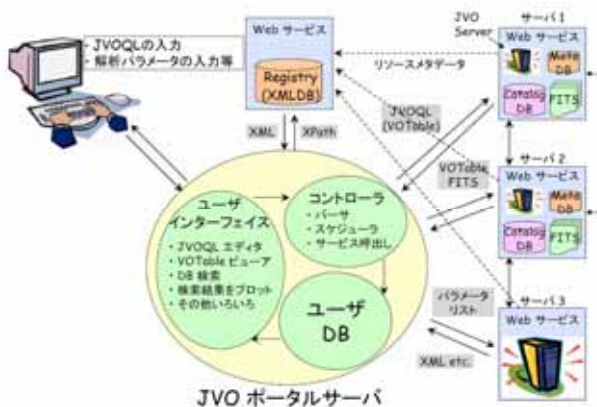


図1 JVOの構成

Fig.1 Architecture of the JVO System

ユーザーインターフェース部はTomcat, Java ServletとJava Server Pageで実装されており、JVOQLエディタの他に、検索結果の表示機能、特定のデータ解析用画面、ステータス表示機能などがある。

3.2 JVOにおける処理の流れ

JVOの利用者はWebブラウザを利用してユーザーインタ

ーフェースからポータルサーバにアクセスし、認証を行った後にJVOQLを記述するかJVOQLエディタから検索条件を入力し検索指令を送る。入力されたJVOQLはポータルサーバ内のパーサによって個々のデータサービスに対するSQLに分解され、これを元に一連の検索や解析の処理を記したWorkflowを生成する。天文データの検索や解析作業は、1度の検索で終了することはなく、様々な検索や解析の組み合わせが必要となるため、実行手順は動的なものでなければならない。クロスマッチ検索を行うときには検索条件に合う天体の個数が少ない方を先に検索することが求められ、1つのサーバに対する検索が失敗した場合には別の似たサーバへ検索の再試行といったことに対応するためである。

生成されたWorkflowはVOSpaceと呼ばれるユーザごとの個人用データ保管領域で管理され、再検索を行う場合や、解析処理の追加などに利用することができる。将来的にVOSpaceには各ユーザ用のデータベースを備え、他のVOと共有することも可能にすることを目標としている。また、利用者管理や認証、登録情報についてはLDAPサーバ[8]を用いて管理される。

Workflowに記された様々なデータサービスへの検索要求を実行するため、JVOポータルはアクセス先URLを検索用レジストリサーバから取得する。データサーバは世界中に存在している天文データベースや解析サービスである。通常天文データベースはRDBであるが、ハードウェア、OS、DBMS、問い合わせの形式が個々に異なる場合が多い。この問題に対し、後で述べるようにインターフェースの標準化が進められている。

検索結果はVOにおけるデータ交換の標準フォーマットとされているXML形式のVOTable[9]で受け取る。結果を表示するユーザーインターフェースにはVOTableをユーザの見やすい表示に変えることのできる機能も実装した。

3.3 レジストリ

VOでは世界の分散データベースを扱う。アクセスするサーバは様々であり、かつ、ダイナミックに変化する。そのため、VOではそれぞれのデータベースについてURLやテーブル情報などのメタデータをレジストリとして登録し公開する必要がある。プロトタイプ1ではレジストリをUDDI(Universal Description, Discovery and Integration)によって構築したが、天文データを記述するさまざまなメタデータを登録するにはUDDIでは不十分であることが分かり、プロトタイプ2ではレジストリをXMLデータベースによって構築した。プロトタイプ3でもこの方式を踏襲した。

天文データベースでは、メタデータも変更される可能性があるため、常にメタデータの交換を行わなければならない。そのためIVOAの取り決めにより、OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)[10]プロトコルを利用してメタデータの交換を行うこととした。OAI-PMHは電子図書館間の相互運用のために策定されたメタデータ交換プロトコルである。

レジストリには、データサービスの管理者がメタデータを登録するPublishing Registryと、Publishing Registryからメタデータを取得し登録されているメタデータを検索することができるSearchable Registryの2種類が存在する(図2)。IVOAではメタデータに登録される項目も決め、データ構造を定義するスキーマも用意している。レジストリサービスは、Publishing Registryに対して、起動時、及び定期的にサービス情報の更新を問い合わせ、差分情報を取得し、自身の持つ

サービス情報レジストリに登録して管理する。差分情報の取得は、問い合わせ時にタイムスタンプを指定した選択的ハブスティングにより実現した。

レジストリに登録されたサービスのメタ情報は、ポータルサーバからのカタログ検索要求やデータサービスの呼び出し時に利用される。レジストリサービスは IVOA で決められたインターフェースを Web Service により実装した。

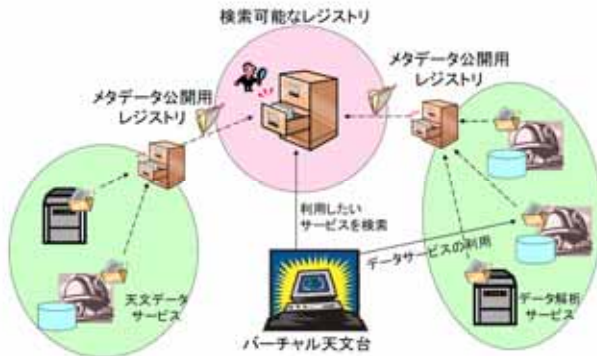


図2 レジストリ情報交換の概念図

Fig.2 A diagram in Exchanging Registry Data among VO's.

3.4 他国 VO との連携

3.4.1 国際標準の策定

既に述べたように、世界各地の天文データベースは独立に開発されているため、そのインターフェースは多種多様なものとなっている。そのため、データベースの連携には、インターフェースの国際的な標準化が必要である。IVOA では VO の標準プロトコルの制定を行っており、Simple Image Access Protocol (SIAP) と Astronomical Data Query Language (ADQL) 等がこれまでに策定され、さらに、スペクトルデータの検索を記述する Simple Spectrum Access Protocol (SSAP) が定められつつある。VO ではこれらのプロトコルを採用することで世界中の天文データベースや解析サーバが利用可能となる。

SIAP は天体撮像データの取得を目的としたもので、天球上の領域等を指定する複数のパラメータを HTTP の GET メソッド等で送信することにより、取得可能な画像データへのアクセス方法やそのメタデータを VOTable として取得することができる。ADQL は SQL 言語を拡張したもので、天文のカタログデータ検索が容易に記述できるよう定義されている。ADQL は我々が定義した JVOQL を基盤として制定されたものである。ADQL 自体は XML 文書として表現され、Web サービスを使って渡される。現状の ADQL の仕様には画像検索を行う機能は含まれていないが、画像検索も行うための仕様拡張が検討されている。

我々は先に述べた JVOQL をさらに改良し JVOQL2 として SIAP と ADQL を統合した仕様策定も進めている。VO の利用者の観点で考えると、単一の検索言語であらゆる種類の天文データにアクセスすることが望ましく、IVOA でも我々の提案する JVOQL2 に基づいて統一検索言語についての議論を開始している。

3.4.2 標準プロトコルの実装

JVO では ADQL によるデータサービスを実現する

SkyNode を導入した。SkyNode は IVOA で標準化されているカタログアクセス機能を提供するサーバである。しかし上述したように、我々は各種アクセスプロトコルを統一的に扱うことが可能な JVOQL2 を提案しており、プロトタイプ 3 に JVOQL2 を実装してその有用性を世界に提示することが今後の VO の発展にとって極めて重要であると考えた。そこで我々は SkyNode を独自に拡張し ADQL のみならず SIAP や SSAP 等も処理できるようにした SkyNode-J を実装することとした。図 3 にブロック図を示す。

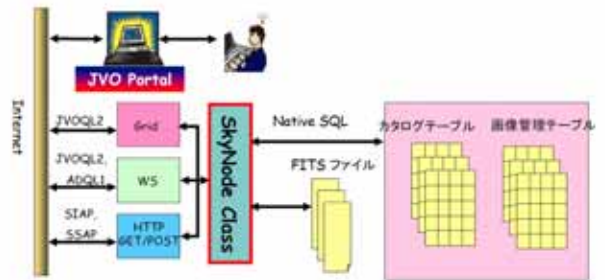


図3 JVO 内外とのゲートウェイになる SkyNode-J のブロック図。複数のプロトコルを受信できるように内部にプロトコル変換機構を実装している

Fig.3 A Schematic Diagram of SkyNode-J. SkyNode-J has a protocol translator to accept multiple protocols.

SkyNode-J は JVO への各種検索要求の窓口となる。一方 JVO 内部では検索言語として JVOQL2 を用いる。JVO ポータルに直接アクセスする利用者は JVOQL2 によって検索要求を記述することとなるので問題はない。しかし IVOA 標準を実装した海外の VO からのアクセスは、当然 ADQL, SIAP, SSAP によって行われる。そこで我々は ADQL, SIAP, SSAP を JVOQL2 に変換するプロトコル変換モジュールを SkyNode-J 内に実装した。JVOQL2 は ADQL, SIAP, SSAP の上位互換に当たるので、この変換は困難ではない。

SkyNode-J のプロトコル変換モジュールを通過したアクセス要求は JDBC を介して適切なデータベースサーバに伝えられる。各データベースサーバは、検索結果を返し、さらにその結果は VOTable に変換されて各 VO の利用者へ返される。

このような機能は既存のデータベースを大幅に変更することなく追加させることが望ましいので、我々は VO に対応したデータサービスとして公開するためのツールキットの開発を行っている。これは JDBC でアクセス可能な RDB を VO 化するためのソフトウェアコンポーネント群であり、このツールキットを配布することによって、世界中の天文データベースを連携することが可能になる。

4. 連携とサイエンスへの応用

はじめに述べたように、近年の天文観測は世界中さまざまな場所のみならず、宇宙空間でも行われ、そのデータは各観測所に保管されている。VO を利用することによって、これらのデータを必要に応じて容易に取得することができるようになる。

JVO のプロトタイプ 3 は 2004 年 12 月から動作し始め、

2005年4月の時点で、アメリカ、フランス、カナダ、スペインなどの86サーバへのアクセスが可能となっている。逆に、イギリス、アメリカのVOにJVOのリソースが登録され、海外からJVOへのアクセスも可能となっていることも確認された。さらに、JAXA/ISASのDARTS[11]にはSkyNode-Jが導入され、VOデータサービスとして実際に連携が実現されている。

JVOを実際に使ったサイエンスユースケースとして、すばる望遠鏡で得られたデータから重力レンズ天体を探す作業を行い、JVOの実用性を検証した。これはカタログ検索、データの解析、再検索、画像検索、画像の切り出しと表示等の作業が必要で、従来の手法では数時間かかる作業であった。しかし、JVOを使うと、数分ですべての処理が終了した。このことから、JVOが天文研究にきわめて有効であることが実証された。プロトタイプ3では海外のデータベースも利用したQSO周辺のイメージ検索による母銀河の検索や、宇宙紐の探査[12]と言った天文学的にインパクトの大きなテーマに取り組む予定である。

5. まとめと今後の課題

プロトタイプ3ではこれまでの開発で構築したJVOQLを始めとする成果だけでなく、国際連携に基づき他国VOプロジェクトと共同開発した各種プロトコルをも取り込み、海外VOとの連携に成功した。海外VOへの検索時間は長くて1分以内と十分に実用に耐える待ち時間であることが分かった。これにより世界のVO間との相互の各種計算機資源が可能になった。

しかしながら課題もある。現状での計算資源共有は公開データベースに限定されている。VOの利用が普及すると、公開前の自分の観測データもVOで処理したいという要望が出てくると予想される。従って今後データの学術的な価値や観測提案者のアイデアを保護するためにもセキュリティ面に配慮したシステムに進化させる必要がある。また、研究を進めるためには天文研究者が普段利用している各種解析ツールもVO上で利用できるようにするための共通インターフェースの開発が必要である。今後は実際にユーザが使用して様々な要求に対応できるか調査し、必要な機能を追加しながら運用システムを構築し、データベース天文学の進歩を先導したい。JVOによる天文データベースへの容易なアクセスにより限られた天文研究者だけでなく多くのユーザが最新データを使うことが可能になり、教育現場での活用や家庭からのアクセスを通じ、天文学の底上げにもつながると期待される。

最後に、この膨大なデータを統一的に扱う手法は天文学に限らず、さまざまな分野への応用や連携といったことが可能であることを指摘しておく。

[謝辞]

本研究は安田直樹氏(東京大学)、増永良文氏(お茶の水女子大学)、大江将史氏(国立天文台)、ならびに中本啓之氏、小林佑介氏、吉田徳夫氏(株式会社セック)、石原康秀氏、山崎昭一氏、瓦井健二氏(富士通株式会社)との共同でおこなわれた。これまでの研究開発を支えていただいた国立天文台スタッフの方々に深く感謝致します。本研究は、文部科学省科学研究費補助金特定領域研究「情報学」(15017289,16016292)及び独立行政法人日本学術振興会先端研究拠点事業による支援を得た。

[文献]

- [1] FITS Support Office <http://fits.gsfc.nasa.gov/>
- [2] IVOA <http://www.ivoa.net/>
- [3] 白崎裕治: “世界の天文データベース連携を実現するバーチャル天文台” 情報処理, Vol.45, No.12, pp.1219-1224 (2004).
- [4] 田中昌宏, 白崎裕治, 本田敏志, 大石雅寿, 水本好彦, 安田直樹, 増永良文: “バーチャル天文台 JVO プロトタイプシステムの開発” DBSJ Letters, Vol.3, No.1, pp. 81-84 (2004).
- [5] 白崎裕治, 田中昌宏, 本田敏志, 大石雅寿, 水本好彦, 安田直樹, 増永良文: “天文学 DB 用検索言語仕様の考察” DEWS2004, 6-C-03 (2004).
- [6] <http://globus.org/>
- [7] <http://www.sec.co.jp/products/karearea/>
- [8] <http://www.openldap.org/>
- [9] <http://vizier.u-strasbg.fr/doc/VOTable/>
- [10] <http://www.openarchives.org/>
- [11] <http://www.darts.isas.ac.jp/>
- [12] Shirasaki, Y. et al “Searching for a cosmic string through the gravitational lens effect: Japanese Virtual Observatory science use case” ADASS XIII, Vol.13, 46, (2004).

本田 敏志 Satoshi HONDA

国立天文台光赤外研究部研究員。2002年 総合研究大学院大学数物科学研究科博士課程修了。博士(理学)。日本天文学会。日本データベース学会正会員。

大石 雅寿 Masatoshi OHISHI

国立天文台天文学データ解析計算センター助教授。1985年 東京大学大学院理学系研究科博士課程修了。理学博士。日本天文学会。日本データベース学会正会員。

白崎 裕治 Yuji SHIRASAKI

国立天文台天文学データ解析計算センター上級研究員。1997年 東京工業大学大学院理工学研究科博士課程修了。博士(理学)。日本物理学会。日本天文学会正会員。

田中 昌宏 Masahiro TANAKA

国立天文台天文学データ解析計算センター研究員。1997年 名古屋大学大学院理学研究科博士課程修了。博士(理学)。日本天文学会。日本データベース学会正会員。

川野元 聡 Satoshi KAWANOMOTO

国立天文台天文学データ解析センター研究員。2002年 東京大学大学院理学系研究科博士課程修了。博士(理学)。日本天文学会正会員。

水本 好彦 Yoshihiko MIZUMOTO

国立天文台光赤外研究部教授。1979年 東京工業大学大学院理工学研究科博士課程修了。理学博士。日本物理学会。日本天文学会。情報処理学会正会員。