

1. シリコンとファイバ上の天文学

1. 世界中の天文データベース連携を実現するバーチャル天文台

白崎 裕治

国立天文台 天文学データ解析計算センター
yuji.shirasaki@nao.ac.jp

天文学の分野では、近年の望遠鏡や観測装置の高機能化、CCDの普及、データストレージ装置の大容量化といったさまざまな技術的進歩により、良質で大規模な天文データが生産・蓄積されています。データは毎年倍増する勢いであり、計算機の演算能力やネットワーク帯域の向上率に比べて急ピッチに増えています。こうした大量のデータから自分の研究に必要なデータを効率的に探し出すためには、観測所ごとに構築されているデータベースを相互運用可能とするシステムの構築が必要となります。そうしたシステムを構築することを目的として、世界各国の天文研究者や情報技術者が集まりデータベース連携のための標準仕様を定めるバーチャル天文台計画がスタートしました。

■ 現代天文学と情報技術 ■

現代天文学は多くの情報技術に支えられています。現在研究目的で利用されている望遠鏡の観測データはほぼ例外なくCCD等を利用したデジタルデータとして取得されており、データの保存方法や研究者間でのやりとりの仕方などにおいてデータベース技術やインターネットを利用した電子情報の交換技術等が応用されています。最近では望遠鏡の操作もコンピュータによる自動制御が可能となり、インターネットを介して遠く離れた場所にある望遠鏡でリアルタイムに観測を行うことが可能となっています。筆者も先日、東京三鷹市の国立天文台からハワイにあるすばる望遠鏡のリモート観測を行いました。あたかも現地で観測しているかのような感覚で行うことができました。

天体现象というものは一定不変のものだけではなく、ある時突然発生することが非常に多くあり、そうした現象を捉えることが重要となる研究テーマもあります。ガンマ線バーストと呼ばれる現象は、ほんの数秒程度の間X線からガンマ線の波長域で突発的に明るく輝く現象で、可視光や電波の波長域でも数日～数カ月に渡って観測可能な残光があることが分かっています。こうした現象を捉えるためには迅速な情報交換が必要であり、世界中のさまざまな望遠鏡がインターネットを介して情報を自動もしくは人間の判断が介在する半自動でやりとりできるシステムが作られています。筆者はガンマ線バーストを最初に捉えるX線衛星の運用チームのメンバーですが、ガンマ線バーストの発生から数秒後には自分の

携帯電話に情報が飛びこんできます。また、必要であれば自宅のパソコンからでも、衛星から送られてくるたった今発生したばかりのガンマ線バーストのデータを解析することも可能で、その結果を世界中の観測者に報告することができます。こうした即時性を要する観測は情報技術、特にインターネット技術なくしてはできないことです。

データベース技術も現代天文学では必要不可欠なものとなってきています。天文学研究に利用される望遠鏡はおしなべて大規模化が進んでいます。すばる望遠鏡は口径8.2mと世界最大級の望遠鏡です。さらに巨大な、口径30mを超える望遠鏡の計画も進められています。そのような望遠鏡は国家レベルまたは国際協力により建設されるもので、巨額の前算が必要となりますから、それによって得られるデータも非常に価値の高いものであり、人類の共有資産としてアーカイブされ多くの人が利用できることが求められています。したがって、観測データは観測終了と同時にデータベース化されるシステムを持つことが常識となっています。

望遠鏡の利用はだれもが自由に行えるわけではなく、観測を許可してもらうためにはその科学的意義を示す提案書をつくり、それが何人かのレフリーにより審査され合格点を得られてはじめて観測が可能となります。世界中の研究者が応募してくるため、大型望遠鏡の利用は競争率が高く、なかなか自分の観測を行うことはできません。また、天体现象の背後に隠された物理を導き出すことが天文学の目的ですが、特定の波長域の観測だけでは不十分なことが多くあり、可視光以外に電波、X線、ガ

国内		
SMOKA	http://smoka.nao.ac.jp	国立天文台
DARTS	http://www.darts.isas.ac.jp	宇宙科学研究所
NRO DB	http://nrodb.nro.nao.ac.jp/	国立天文台野辺山観測所
JMAISON	http://maison.nao.ac.jp	国立天文台 + 宇宙科学研究所
海外		
MAST	http://archive.stsci.edu/	宇宙望遠鏡科学研究所 (米)
HEASARC	http://heasarc.gsfc.nasa.gov/docs/archive.html	高エネルギー宇宙物理研究所 (米)
ROSAT	http://wave.xray.mpe.mpg.de/rosat	マックスプランク研究所 (独)
VizieR	http://vizier.u-strasbg.fr/	ストラスブルグ天文台 (仏)
SDSS	http://cas.sdss.org/dr2/en/	スローンデジタルスカイサーベイプロジェクト (米・日・独)
IRSA	http://irsa.ipac.caltech.edu/	NASA (米)

表-1 主要な天文データベース公開サービス

ンマ線といった多波長観測が必要となってきます。そのような広い波長域を単一の望遠鏡で観測することはできないため、それぞれの波長域に特化した望遠鏡が必要ですが、そう簡単に観測が行えるわけではないのです。そうしたことを背景として、自分で観測を行うかわりにデータベースに登録されているデータを利用した研究が盛んになってきています。

望遠鏡の大規模化は鏡の巨大化という方向とは別に、受光装置である CCD の大面積化という方向へも進んでいます。デジタルカメラの普及等の影響により CCD の製造単価が安くなってくると、多数の CCD をモザイク状に配置し、一度に広い範囲の天空を観測する望遠鏡が作られるようになりました。米国 Sloan 財団の援助を受けて建設された SDSS 望遠鏡は、2048 × 2048 画素の CCD 30 枚を並べた検出装置を持ち、画素数というと 1 億画素を超えます。データの生成レートは一晩で 100GByte にもなります。こうした観測装置が作られるようになった背景には記録媒体の高密度化により大量のデータを保存できるようになったことも挙げられます。こうして産み出される大量のデータの中から自分の研究に必要なデータを効率良く取り出すためにも観測データのデータベース化は必須の要件となってきています。

■ バーチャル天文台 ■

以上のように良質で大規模な天文データが続々と産み出されるようになり、それらのデータベース化も進められてきました。たとえば、国内では国立天文台内にすばる望遠鏡によるデータのほか、岡山・木曾天文台で得られたデータのアーカイブが構築され、それらは公開されています。論文発表されたカタログデータの収集、他

国データサービスのミラーサイトの提供も行っています。宇宙科学研究所ではおもに衛星観測によるデータのアーカイブが構築されています。そこからは X 線衛星 Ginga, ASCA, Yohko 等のデータを取得することができます。海外ではフランスの天文台によるカタログデータサービス VizieR, ハッブル宇宙望遠鏡 HST のアーカイブ, NASA の高エネルギー天文衛星データアーカイブ, Harvard-Smithsonian Center for Astrophysics (CfA) のデータアーカイブなどがあります(表-1 参照)。

これらのデータアーカイブ・カタログ検索サービスから多種多様なデータを入手することが可能となりましたが、そのデータベース化は各観測所単位で独立に行われたため利用する側からのインタフェースは千差万別となってしまっており、多数のデータベースを利用した研究を行う上で大きな障害となってきました。また、公開されるデータベースサービスも増えてきており、どこにどういったデータが存在するのかを知るためのシステムが必要となってきました。データベースから得られるデータは一次処理が必要な生データが大部分であり、そうした処理を行うためにはデータごとに異なる解析ツールを利用しなくてはならない場合が多くあります。したがって、データをとってくるだけではなく、その解析ツールも同時に取得し、自分が利用する計算機上にインストールする必要があります。当然使い方も学ばなくてはなりません。

バーチャル天文台 (Virtual Observatory: VO) は世界中のデータベースを連携し、天文学の研究に必要なデータを容易に取得できる環境を実現するシステムです(図-1 参照)。また、データサービスと解析サービスも連携することにより、汎用的な解析ならば自分の計算機にプログラムをインストールすることなく行えるような

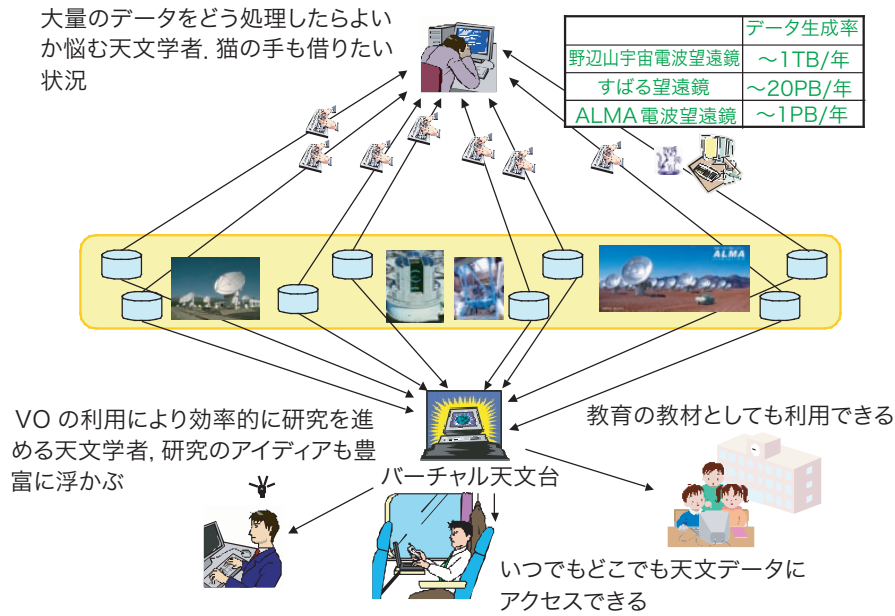


図-1 バーチャル天文台の概念図

	実天文台	バーチャル天文台
観測対象	空、天体からの光	データベース、観測データ、天体カタログ
観測装置	望遠鏡、CCD、光ファイバ	インターネット、コンピュータ
天候の影響	大	なし
観測時間	暗い天体程長時間	明るさに関係なく短時間で終了
観測可能領域	望遠鏡の視野内	全天の観測

表-2 主要な天文データベース公開サービス

環境づくりを目指しています。研究者はバーチャル天文台に必要なデータの要件と解析手順を入力すれば、あとはバーチャル天文台のシステムが世界中のデータベースの中から要件にあったデータベースを探し出し、検索を実行し、結果が返ってくるのを待てばよいのです。これにより研究の効率化が図られ、これまでの手法ではできなかったような研究の実現により、新たな天文学上の発見が期待されています。研究者の利用のみならず、一般の人々の利用も可能になります。たとえば、小中学校の理科教育の教材としても利用できることをバーチャル天文台は目指しています。バーチャル天文台と実天文台は表-2 に示したような違いがあり、それぞれメリット・デメリットがありますが、「利用者の必要なデータを取得する場」という意味では同じであり、バーチャル天文台は「実天文台で得られたデータの公開の場」という見方に立てば、相互補間的と言えます。

■ データベース連携のためのプロトコルの標準化 ■

バーチャル天文台では利用者はデータがどこにあるかを意識することなくさまざまなデータサービスにアクセスできることを目指していますから、利用したいデータがどこにあるのかを発見するための手続き、そしてそこからどのように取得するのかといった取り決めが標準化されている必要があります。そうした標準化を行うために国際バーチャル天文台連合 (International Virtual Observatory Alliance: IVOA)^{☆1} が結成され、さまざまな規格作りを行っています。

まず、データのありかを捜し出すためにはデータのメタデータを保持するデータベースが必要です。メタ

☆1 <http://www.ivoa.net/>

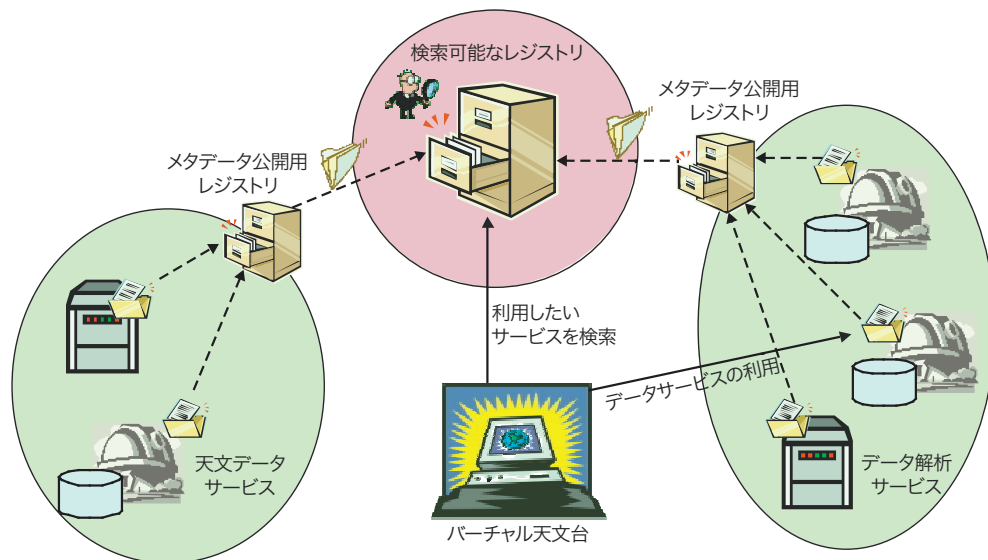


図-2 天文データ・解析サービスの公開方式

データを保持するデータベースは「レジストリ」と呼ばれ、その機能によって大きく2つに分けられます(図-2参照)。1つは「公開用レジストリ」で、これはメタデータ登録専用のレジストリであり、データサービスや解析サービスがバーチャル天文台から利用できるようにするための第一歩がここから始まります。公開用レジストリは通常データセンタごとに作られ、センタが管理しているデータについてのメタ情報が登録されますが、たとえば研究者が個人で作ったデータベースを世界に公開する場合にセンタが管理する公開用レジストリに自分のメタデータを登録するという利用方法も考えられます。もちろん、個人で公開用レジストリを立ち上げることも可能です。このレジストリは検索機能は持たず、もう1つ別の種類の「検索用レジストリ」によってデータを吸い上げられます。「検索用レジストリ」はその名のごとく登録されているメタデータの検索を行うことができ、公開用レジストリに登録されているメタデータを自分のデータベースに取り込む機能を持ちます。したがって、バーチャル天文台から利用できるすべてのリソース情報を保持します。

公開用レジストリからのメタデータの収集はOAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)^{☆2}と呼ばれるプロトコルに基づき行われます。このプロトコルはあらゆる電子情報を効率的に共有

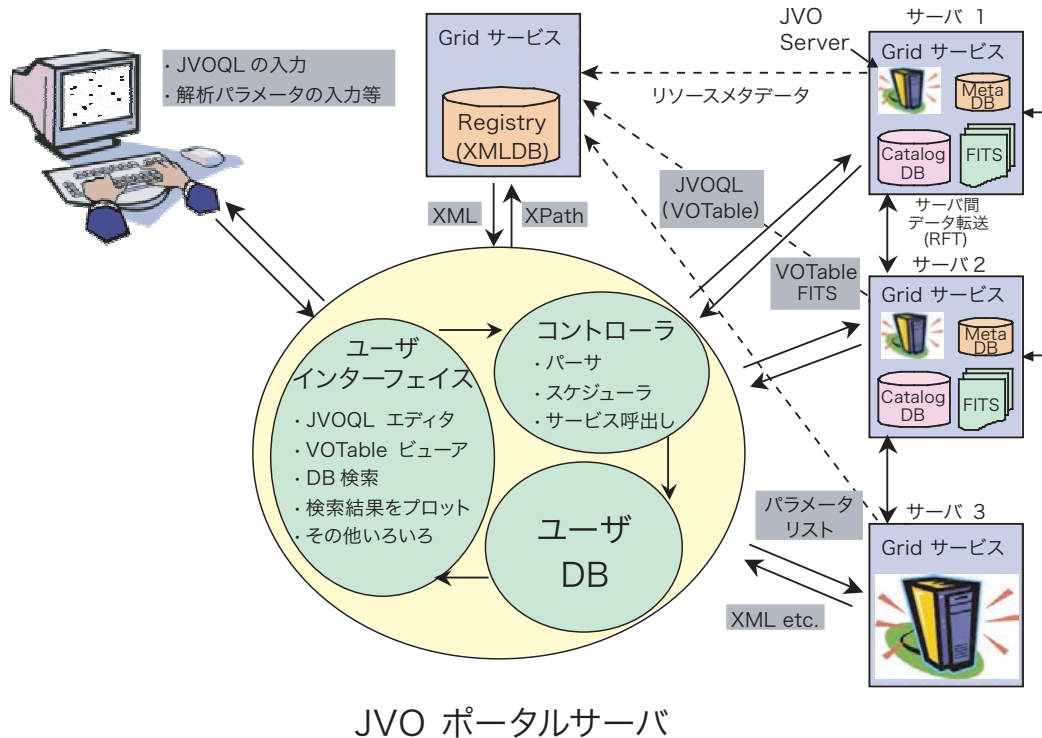
するためのシステムを作ることを目的として作られたものであり、電子図書館サービスなどで利用されています。

レジストリへ登録される項目はIVOA Registry WGにより決められています。データ構造を定義するスキーマも用意されています。たとえば、データベースの識別子とタイトル名は必須の項目となっています。その他、データベースが保持するデータの種別、すなわち天体カタログなのかそれとも画像提供サービスなのかといったことや、データの観測波長域といった情報、天球のどの範囲のデータを含むのかといったことを登録しておくことができます。

たとえばクエーサの周辺銀河の分布を調べたいとします。銀河はクエーサに比べて暗いためより高感度な望遠鏡のデータを調べる必要があります。そのような場合に、そういったデータを保持するデータサービスがどこにあるかを調べるのにレジストリを利用でき、すべてのデータサービスに逐一問い合わせる手間が省けます。

レジストリにより自分に必要なデータがあるデータサービスを見つけることができたとして、次に行うことはそのデータサービスに検索命令を発行することです。データベースへのアクセス方法は従来からも利用されてきたHTTPプロトコルのGET/POSTメソッドが標準として採用されており、パラメータ渡しによるデータ検索のほか、近年利用が高まってきているXMLを利用したWebサービスによるアクセス方法もIVOAで標準とすることに決められました。標準とはされていませんが、Gridサービスを利用するアクセス方法についても

☆2 <http://www.openarchives.org/>



JVO ポータルサーバ

図-3 JVO システム図

検討が行われています。

検索条件を記述する方法は、現段階ではデータ種別によって異なります。画像やスペクトルデータの検索にはパラメータ渡しによる方法を用い、カタログデータの検索にはSQLをXMLで表した文書を渡すWebサービスが基本となっています。画像検索用の検索言語はSIAP (Simple Image Access Protocol) と呼び、カタログ検索用言語はADQL (Astronomical Data Query Language) と呼ばれます。データ種別により異なる検索文法を持つのは煩雑なので、私たち日本のVOグループは、2004年5月に行われたIVOA Interoperability meetingにおいて、それらの言語の統合案を発表し、将来の言語規格として検討を行うことを了承されました。私たちの提案は画像データ検索についてもSQLを基本とした検索言語仕様とするもので、SIAPで指定されるパラメータをカラム名にマッピングし、取得される画像データや、天球上の画像を取得したい領域の指定文もカラムを構成する要素と位置付けます。領域指定文は無制限のパターン存在するので、その実テーブルが存在するわけではなく、そのような仮想的なテーブルに対して検索を行うわけです。

SQLによるデータ検索では検索対象テーブルのカラムを指定する必要があります。しかしながら、どの

テーブルにどういう名前のカラムがあるのかをあらかじめ知っていることはまれです。Unified Column Description (UCD) がデータ内容を意味付ける目的で定義されました。たとえば天球座標である赤経、赤緯はそれぞれPOS_EQ_RA_MAIN, POS_EQ_DEC_MAINと記述されます。

検索結果の標準フォーマットはXML形式のVOTableです。VOTableはデータを表す部分とそのメタデータを表す部分に分けられます。データ部分はシンプルなテーブル構造となっており、行は<TR>タグで表され、値は<TD>タグで表されます。メタデータ部にはカラムの意味内容を記述するメタデータが<FILED>タグにより記述されます。以上の標準規格の文書はIVOAのWebサイトから入手することができます。

■ 国内における開発状況 ■

筆者を含む国立天文台を中心としたJapanese Virtual Observatory (JVO)^{☆3}開発グループは、IVOAで決められた標準プロトコルにもとづく分散データベースの連携システムの構築を行っています。図-3に昨年度構築したシステムの構成図を示します。システムの基本開発言語はJavaとなっています。ユーザインタフェース層はユーザからの検索要求・解析処理ワークフローの入力を受け付け結果を返します。この層はJakarta Projectによ

☆3 <http://jvo.nao.ac.jp/>

る Tomcat を利用し、Java Servlet と JavaServer Page で構築されており、Struts フレームワークが利用されています。

JVO システム層 (図中のコントローラ) では分散したデータベースをあたかも1つのデータベースと見せかけるための処理が行われます。ユーザからの分散DB 対応 SQL 文を受け付け、それをパースし個々のデータベースごとの SQL 文に分解します。そして、実際にアクセスすべきデータベースの URL を検索用レジストリに問い合わせることにより解決し、データサービス側がサポートするプロトコルに合わせた検索命令文の変換がなされデータサービスへ転送します。

データサービスへのアクセス順序はユーザが指定する検索条件等によって変わってくるため、動的にスケジューリングを行うスケジューラが組み込まれています。検索用レジストリには XML データベースであるセック社の Karearea を利用しました。昨年度は Globus TK3 により Grid サービスを中心とした連携システムを構築し、さまざまな問題点の洗い出しを行い、実用システムをつくるノウハウの蓄積を行いました。今年度の開発では、他国との連携を視野にいれ、HTTP GET によるパラメータ渡しによる検索や Web サービスを中心とするシステム作りを優先して開発を進めています。

データ・解析サービス層 (図中の右端サーバ1~3) は実際にデータや解析のサービスを行うサーバ群からなり、それらは実際には世界各地に分散して存在するものですが、接続試験のため現在は LAN 上に構築しています。SQL のパーサは JavaCC により作成し、データベースマネージメントシステムは PostgreSQL と Oracle を利用しました。

天文データベースは天球座標による検索が主に行われます。大量にあるデータの中から、ごく限られた領域のデータを効率よく検索するためには天球座標値でインデックシングされている必要があります。また、天球座標は極と経度の0度で特異点を持ちます。アメリカ Johns Hopkins 大学の研究グループは Hierarchical Triangular Mesh (HTM) ^{☆4} による天球面のインデックシング法を提案しています。今回この提案にもとづき、HTM によるインデックシング法を実装しています。

以上のようにバーチャル天文台の開発には最新の情報技術が利用されています。こうして構築したシステムを利用することにより、これまでの方法では大変な手間を要するような研究を行えるようになります。私たちは、そうした研究の中から特に重力レンズ現象を利用し

たコスミックストリングの探査を取り上げシステムの有効性について検証しました。コスミックストリングは宇宙がビックバンにより生まれた直後の冷却時期において生成されたとされる空間の位相欠陥の1つであり、その存在の検証は宇宙論のみならず、素粒子論や最高エネルギー宇宙線の起源とも関連し重要な研究課題の1つです。コスミックストリングを発見する確率は大変低く、すばる望遠鏡で1,000 視野分の観測データが必要です。手作業でそれだけのデータをダウンロードして解析を行うのはきわめて困難です。今回開発したシステムを使えば、1視野を数分でデータ取得から解析まで完了してしまうので、データの用意ができれば約2日間でコスミックストリングを検出できるかもしれません。このような、データ取得と取得したデータの解析といった一連の手続きを、ワークフローによって実行できるシステムは、他国の VO にはまだない特徴となっています。

■ 今後の展望 ■

日本版バーチャル天文台 (JVO) の開発は2004 年中に IVOA 標準プロトコルの実装を終え、2005 年1月には他国の VO との連携を行うことを目指しています。来年度からは部分的に一般公開を始める予定です。また、VO 対応のデータサービス・解析サービスを簡単に立ち上げることができるツールキットを今年度の開発で完成させ、日本の全天文データベースの VO 対応化を来年からスタートしたいと考えています。

現段階では Web サービスを中心としてシステムが構築されていますが、大量のデータの移動には HTTP プロトコルによるデータ転送は簡単ではありますが、転送速度の面からは十分であるとはいえません。したがって、情報学の分野で現在盛んに研究が進められているグリッド技術は天文データベースや解析サービスを連携する上で重要なものと位置付けており、グリッド技術を利用したシステム作りの研究と実運用システムの構築が今後の計画となっています。

また、天文学研究の応用といった面での計画も持っており、昨年度の重力レンズ探しに加え、クエーサの進化と周辺銀河密度の関係を探るといった研究テーマや赤外データを利用した晩期型星の自動分類法の研究、第一世代星の探査といった研究テーマを JVO を利用して実現する予定です。

(平成16年11月5日受付)

☆4 <http://www.sdss.jhu.edu/htm/>