

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

ーワークフロー機能の実装ー

川野元 聡[†] 大石 雅寿[†] 白崎 裕治[†] 田中 昌宏[†] 本田 敏志[†] 水本 好彦[†]
 大江 将史[†] 安田 直樹^{††} 増永 良文^{†††} 石原 康秀^{††††} 堤 純平^{††††}
 中本 啓之^{†††††} 小林 佑介^{†††††} 坂本 道人^{†††††}

[†]国立天文台 〒181-8588 東京都三鷹市大沢 2-21-1

^{††}東京大学 宇宙線研究所 〒277-8582 千葉県柏市柏の葉 5-1-5

^{†††}お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

^{††††}富士通株式会社 〒261-8588 千葉県千葉市美浜区中瀬 1-9-3

^{†††††}株式会社セック 〒150-0031 東京都渋谷区桜丘町 22-14

E-mail: †{kawanomoto.satoshi, masatoshi.ohishi, yuji.shirasaki, masahiro.tanaka, honda.satoshi, mizumoto.y, masafumi.oe}@nao.ac.jp, ††yasuda@icrr.u-tokyo.ac.jp, †††masunaga@is.ocha.ac.jp, ††††{ishi, tsutsumi}@ssd.ssg.fujitsu.com, †††††{nakamoto, kobayashi, m-sakamoto}@sec.co.jp

あらまし 観測天文学ではあらゆる波長の電磁波でデータが取得され、DB化されている。これらのDBを連携して天文学発展の新しい局面を切り開くものとしてヴァーチャル天文台(VO)計画が世界各国で進められている。また各地のVOを結合するための国際的な連携も進んでおり、日本ではすばる望遠鏡などの画像・カタログデータを活用すべく我々のグループが国際連携に基づいてJVOの開発を進めている。JVOでは昨年までに各国のVO計画と連携したデータの相互利用が可能になっているが、今回新たにBPEL4WSのサブセットに相当するワークフロー(WF)言語を定義し、WF処理機能を実装した。ここでは、VOを使ってサイエンスを行なう際に必要な手続きの流れをXML形式のファイルに記述したものをWFと呼んでいるが、この実装によってJVOは単なる天文データベース検索ツールではなく、データと解析サービスを結合した実用的なサイエンスのツールとしてより有効になった。

キーワード 科学DB、XML、ワークフロー、天文学DB

Development of Astronomical Database System

—Implementation of workflow function—

Satoshi KAWANOMOTO[†], Masatoshi OHISHI[†], Yuji SHIRASAKI[†], Masahiro TANAKA[†],
Satoshi HONDA[†], Yoshihiko MIZUMOTO[†], Masafumi OE[†], Naoki YASUDA^{††}, Yoshifumi
MASUNAGA^{†††}, Yasuhide ISHIHARA^{††††}, Jumpei TSUTSUMI^{††††}, Hiroyuki NAKAMOTO^{†††††},
Yusuke KOBAYASHI^{†††††}, and Michito SAKAMOTO^{†††††}

[†]National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka-shi, Tokyo, 181-8588 Japan

^{††}Institute for Cosmic Ray Research, University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi,
Chiba, 277-8582 Japan

^{†††}Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

^{††††}Fujitsu Limited, 1-9-3 Nakase, Mihama-ku, Chiba-shi, Chiba, 261-8588 Japan

^{†††††}System Engineering Consultants Co., Limited, 22-14 Sakuragaoka-cho, Shibuya-ku, Tokyo,
150-0031 Japan

E-mail: [†]{kawanomoto.satoshi, masatoshi.ohishi, yuji.shirasaki, masahiro.tanaka, honda.satoshi,
mizumoto.y, masafumi.oe}@nao.ac.jp, ^{††}yasuda@icrr.u-tokyo.ac.jp, ^{†††}masunaga@is.ocha.ac.jp,
^{††††}{ishi, tsutsumi}@ssd.ssg.fujitsu.com, ^{†††††}{nakamoto, kobayashi, m-sakamoto}@sec.co.jp

Abstract In the observational astronomy, astronomical data in multiple wavelength range are obtained and are compiled to database systems. Virtual Observatory (VO) projects are promoted in many countries as new tools for development of astronomical studies by connecting various astronomical databases and computing resource all over the world. Japanese Virtual Observatory (JVO) is also developing in order to apply to image and catalog data of the SUBARU telescope and so on. Our JVO project enabled to cooperate various VO projects all over the world in 2004, and, as a new function of JVO, we implemented a faculty to treat a workflow into JVO portal. The workflow function is developed as a subset of BPEL4WS and workflow is described in XML format. We emphasize that JVO is a realistic science tool including functions for database search and data analysis.

Key Words scientific DB, XML, workflow, astronomical DB

1. はじめに

近年、観測天文学はその守備範囲を飛躍的に広げてきた。地上においてはまず1993年に口径10mのKeck I望遠鏡が稼働を開始し、1996年にはKeck IIが完成した。2000年前後には、国立天文台のすばる望遠鏡を始めとして、Gemini Telescopes, VLTなどの8m級光学赤外線望遠鏡が次々に稼

働を開始した。さらに、現在でも地上大型望遠鏡の建設は世界各地で続けられている。このような大型望遠鏡の建設ラッシュは、“深い”観測 - より暗い天体まで含んだ観測 - を可能にし、天球上での観測領域を大きく広げることも可能にした。光学赤外線望遠鏡以外にも電波天文学では年間数ペタバイトを生み出すALMA望遠鏡がチリに建設中である。また、宇宙空

間の衛星軌道上には、多数の天文観測衛星が打ち上げられ、地上では大気によって吸収されて観測ができない遠赤外線・中間赤外線・紫外線・X線・γ線などの観測データを収集している。このように、波長域・観測領域・観測の深さにおいて格段の進歩を遂げた天文観測機器は、必然的に大量のデータを生み出している。数多くの波長域のデータを組合せ、天文現象について統一的な理解をすることで、現代天文学に残された数多くの謎・ダーク・マターの解明、コスミック・ストリングの存在の検証、等々が解明されると期待されている。しかし、逆説的ではあるが、この大量のデータそのものが従来の手作業を基本とした観測データ解析手法を破綻させ、迅速に結果を導き出すことを困難にしている。ここにおいて天文学は、特に統計的手法を必要とする場面では、これまでとは違ったデータ処理法を導入する必要がある。近年発展の著しい電子計算機資源を活用した情報処理技術は大いに期待されている。

電子計算機はハードウェア・ソフトウェア両面において急激に発展中であり、また計算機同士を接続する通信技術もここ数年で格段に進歩した。天文学においてもこの流れを採り入れる動きが盛んで、観測データはアーカイブ化されて公開されるのが当たり前になっている。加えて通信技術の進歩がこれらアーカイブ化されたデータの検索・取得を支えている。このような状況が整ってきたことで、世界中の大量のアーカイブデータにアクセスし、多波長のデータから必要なものを検索・取得しつつ高速処理し、総合的な研究を速やかに行なうことが可能になりつつある。

例えばすばる望遠鏡は一年当たりおよそ30テラバイトの観測データを得ることができる。この膨大なデータを速やかに解析し、世界各地に存在するすばる望遠鏡と同程度かまたはそれ以上の性能を持つ望遠鏡群との国際競争に立ち後れないためには、これまで我々が培ってきたデータ解析技術を最新の計算機技術を使って発展させ、分散環境におけるデータ解析システムとして構築し、また、サイエンスを支えるデータそのものの供給源としてデータアーカイブシステムを確立する必要がある。そして、これらを組み合わせることで一つのサイエンスアーカイブシステムを構築することが重要になる。

我々は、このような科学的要求を実現するために、発展著しい情報学・計算機科学の研究成果と、大望遠鏡と最新の観測技術から生み出される大量のアーカイブデータとを融合させ

た一つのシステムとして Japanese Virtual Observatory(JVO)の開発を進めてきた[1][2]。JVOは国内研究機関によるデータサービスへのアクセス機能を持つだけではなく、世界各地でそれぞれ公開されている天文学データアーカイブ・データベースや天文学データ解析サービスを連携させ、時刻・場所を選ばずに天文学を推進することが可能となる研究基盤を整備することを目的としている。

これまでの研究開発において構築し、高機能化を進めてきたJVOプロトタイプは現在第3版まで達している。現在、JVOは、世界各地で行なわれている同様のプロジェクトが提供するアーカイブデータベースの検索・アクセスが容易に行なえる機能を持ち、また、限定的ではあるがデータ解析サービスの提供も可能になっている。そして、このようなデータベース検索機能・データ解析機能をより汎用性をもって有機的に結合し、来るべき超大量データによる天文学の時代に備えるための機能としてワークフロー処理機能を追加した。本論文では、このワークフロー処理機能について詳細を述べる。

2. JVOの開発 - プロトタイプ第3版 -

JVOプロトタイプ第3版は、海外のVOプロジェクトとの連携を主な目標として開発された。International Virtual Observatory Alliance(IVOA)[3]の進めるVOの標準化に基づいて実装を行ない、海外VOに対して相互にデータベースへのアクセスが可能なシステムが構築されている。

JVOプロトタイプ第3版は以下のような4つの構成要素からなっている[2]。

- 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インターフェース部
- 検索および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、その処理からの結果を返すコントローラ部
- ポータルサーバからの検索サービスや解析サービスへの要求に対して実際の処理を実行して結果を返すデータサーバ部
- メタデータの管理・検索を行なうレジストリサーバ部

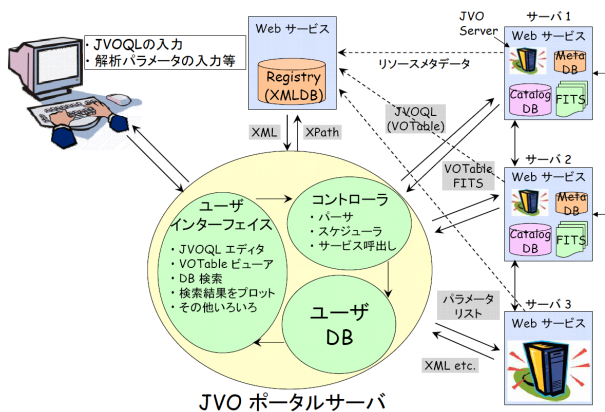


図1 JVO ポータルサーバの構成

Fig.1 Architecture of JVO portal server.

システム開発にはJavaを用いている。このことで、ハードウェアやオペレーティングシステムに大きく依存しないで開発を行なうことができる。また、分散配置されたサーバ群はWebサービスで連携されている。

JVOのユーザはポータルサーバにアクセスし認証を行った後に、データ検索条件を記述する言語であるJVOQL[4]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。ユーザーインターフェース部はWebベースで作成しており、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。

海外のデータベースサーバと接続する際には、どこに・どのように接続して、何が返ってくるかを知っていなければならない。VOの世界では、データを公開するデータベースサーバは各VOプロジェクトに対して、アクセスURLやテーブル情報・カラム情報などのメタデータをレジストリに登録し、他のVOに対して公開する必要がある。データサーバがメタデータを登録するPublishing Registryと、各地のPublishing Registryからメタデータを取得して持っておき、登録済みのメタデータを検索することのできるSearchable Registryが存在する。この仕組みによって海外のデータベースサーバにアクセスすることが可能になるが、データベースサーバは必ずしも不変のものではなく、そのメタデータも変更を受ける可能性がある。このため、各地のVOプロジェクトでは常にメタデータを交換し合い、最新の情報を保持しておく必要がある。ここでもIVOAによって、OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)[5]

プロトコルを使ってメタデータの交換を行なうことが決められている。

また、データベースのデータそのものの取得のためにも標準プロトコルが必要であり、これに関してもIVOAではSimple Image Access ProtocolやSkyNodeなどの標準が定められている。各VOプロジェクトはこれらの外部インターフェースを採用しさえすれば自由なシステム設計、自由なDBMS使用が可能となる。

JVOプロトタイプ第3版では、このようなデータベース連携については、世界各地のサーバに対して100件を超えるようなデータサービス・解析サービスに関する連携が実現している。各国において得意とする研究分野は自ずと異なることから、このようなデータの連携は必須のものである。

3. ワークフロー機能の詳細

昨年度までのJVOプロトタイプ第3版では、データ処理手続きの流れは、JVOQLの表現の許す範囲内での限定的な構造化しかなされていなかった。今回、更に高度な処理を一括して記述し、自動的に実行するためのワークフロー機能をJVOに追加するべく、Webサービス連携用の標準的なWF記述言語であるBPEL4WSを基にして、WF記述言語をXMLで定義した。現在開発中のJVOのワークフロー機能は、BPEL4WSで定義される多数のアクティビティを完全に再現するものではなく、そのサブセットに相当するアクティビティ群で構成される。いくつかの天文学的サイエンス・ユースケースを想定し、これらを実行するための最小限の命令セットをまず実装して、ワークフロー機能の実効性を検証するためである。BPEL4WSのアクティビティセットを完全に取らない理由は、数例のユースケースを満足する命令セットが実装されていれば、大半の天文学WFを記述するには必要十分と判断したためである。また、このような基本的な命令セットの実装ができていれば、今後開発を続けていく際に必要に応じて機能を拡張することが容易になると考えられ、開発開始初年度の目標としては限定された機能を確実に動作させることを重視した結果でもある。以下、現在実装されているJVOのWF機能についてその構成と個々の機能について解説する。

XML形式で記述されたWFをJVOポータルに実行させる際には、Groovyを使用している。GroovyはJava仮想マシン上で動作するスクリプト言語であり、Javaと直接関係できる。

先述の通り JVO ポータルは Java で開発されているため、WF 機能の実現のために Groovy を使用することは、実装の効率の点から見て非常に都合が良い。XML 形式の WF は、JVO ポータル上で Groovy スクリプトの形に変換して実行される。構造化された WF の実行は、Groovy スクリプトの機能によって実現されているため、Groovy の言語仕様で可能な機能は原理的にはすべて WF を記述する際に使えるが、現在はスクリプト変換エンジンの実装上の問題により、Groovy の機能を完全には使えない。しかし、現状でも WF 処理のプロトタイプとしての機能は完全に備えており、サンプル WF を実行することが可能になっている。

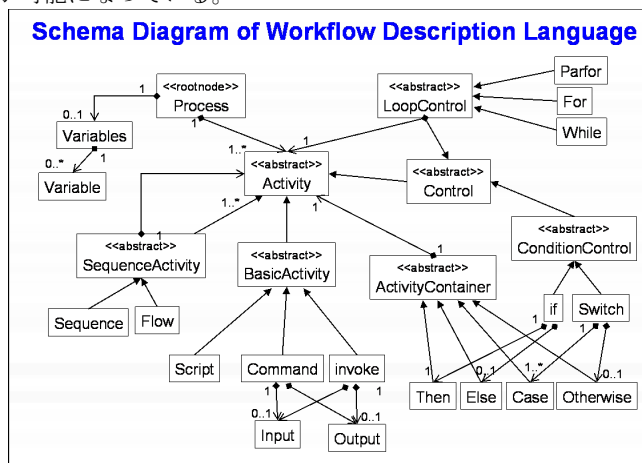


図2 JVO ワークフロー言語スキーマ図

Fig.2 Schema diagram of JVO workflow language.

WF 記述言語は以下のような要素を持つ。

- 変数定義 Variable definition
- 外部呼出 Invoke external services
- 内部呼出 Invoke built-in Java classes
- 逐次実行 Sequential execution
- 並列実行 Parallel execution
- 反復構造 Loop
- 選択構造 Condition

これらの要素により、構造化されたワークフローが記述でき、データの検索・収集からその内容による条件分岐・反復処理などが可能となる。以下にそれぞれの要素についての詳細を記す。

変数定義機能とは

ワークフロー内で使われる変数を定義する機能

である。variables 要素の中に variable 要素を複数持つことができる。variable 要素は name(変数名) 属性、type(変数の型)属性と value(値・複数指定可能)要素を持つ。

外部呼出とは

外部のサービスを呼び出すための機能である。識別子・URL・ポートタイプ名・オペレーション名で一意に定まる外部サービスに対して、サービスの呼び出しおよび入出力の受け渡しを行なう。識別子・URL・ポートタイプ名・オペレーション名はそれぞれ identifier・url・portType・operation 属性として持ち、入力と出力は input・output 要素として持つ。input・output 要素はそれぞれ 0 個以上の子要素を持つことができ、それはあらかじめ定義済みの変数または任意の定数から選択される。

内部呼出とは

内部のコマンドや処理を実行するための機能である。XML 表記上、xsi:type 属性を使って型指定を行なう。具体的に取り得る値は、ビルトイン型とクラスメソッド型である。ビルトイン型は利用頻度の高い操作をあらかじめ用意しておき、これら呼び出すために使用する。この時、name 属性にコマンド名をとる。入出力に関しては外部呼出と同等の機能を備える。クラスメソッド型は任意の Java クラスに定義されたメソッドを呼び出すために使用する。この時、class 属性にクラス名を、method 属性にメソッド名を記述する。入出力に関してはビルトイン型と同様である。

逐次実行とは

子要素を順次実行するための機能である。子要素として複数の機能をとることが可能である。この機能が実装されることで、実際に記述できるワークフローの幅は大きく広がる。単一の機能を実行できることと、複数の機能を接続して実行できることとは質的な差がある。

並列実行とは

子要素を並列実行するための機能である。子要素として複数の機能をとることが可能である。実際

にワークフローを実行する際、端末に対するレスポンスを向上させるために重要な機能である。具体的には、外部サービス呼び出しの際、回線状態などの影響で必ずしも反応が返ってくるとは限らない。その場合、タイムアウトを待つことになるが、逐次実行と並列実行では制御が戻るまでの時間に大きな差が生じる。

反復構造とは

条件に沿って繰り返し実行を行なう制御のための機能である。基本的には、構造化プログラミング言語に良く現れる `for` 文の機能を実現するためのものである。属性として、アイテムの並びを示す変数を指定できる。アイテム並びのうち、ループで実行中のアイテムは `var` 属性に指定した変数名で参照できる。ループの内部に記述できるのは単一の機能または逐次実行機能でまとめられた機能群である。

選択構造とは

状態に応じて実行する機能を分岐制御する機能である。構造化プログラミング言語においては `if` 文として現れる機能に相当する。条件は `condition` 属性に記述し、一致するなら `then` 要素内が実行され、一致しない場合は `else` 要素内が実行される。また、`else` 要素については省略可能である。`then` 要素、`else` 要素ともにその内部には単一の機能または逐次実行機能でまとめられた機能群が記述できる。

JVO ポータルが以上のような機能群を持つことで、過去には人間の手を介在せざるを得なかったような処理が一度に自動的に動かせるようになった。例えば、昨年度までの JVO ポータルの機能では、検索をかけ、出てきた結果を次の処理に回したい時に、ウェブブラウザ上での画面遷移を伴う状態遷移を使って次段階の処理を起動し、パラメータを与える必要があった。ここで人間の介在が必要となり、真の自動化・パイプライン処理はできていなかった。また、一ユーザーの立場で新たな処理の流れをポータルに追加することも自由にはできなかった。今回追加されたワークフロー機能で、ユーザーは自由に処理の流れを組み立てることが可能になり、自らのサイエンスに直結する処理を JVO ポータルを使って行なうことが可能

になった。

4. 可能となるサイエンス

JVO にワークフロー実行機能が実装されたことで、どのようなことが可能になるかを実証するため、過去に実際に VO を使ってなされ、論文として出版されている天文学的研究の追試を行なってみた。一例として挙げると、Padovani et al. [6]では、X線・可視光線のサーベイデータカタログからある特徴を持つものを選びだし、可視光線の画像データも使用して統計的な議論の材料にするというを行なっている。この時に重要なのは複数の波長での観測位置データをクロスマッチし、同じ天体からの X線・光であることを見つけ出すことだが、データベース間でのデータのクロスマッチという VO の基本機能を使うことで、容易に候補天体を絞り出すことが可能になっている。このように、ある天球上の領域における多波長での深いサーベイ観測 - これは複数の観測所の共同事業として一般的なものである - において、JVO は大きな力を持つことがわかった。

5. まとめ

世界各地の天文学データベースを有機的に結合してできる VO は、データの生成率が格段に向上した現在の天文学には必要不可欠なものである。我々のグループでも、海外各地の VO プロジェクトと連携して稼働する JVO の開発研究を行ってきている。昨年度までに海外 VO との結合試験にも成功し、実用的な時間内のデータ検索・取得が可能であることがわかっている。

今年度は実際の活用に向けて JVO の機能拡張を行なった。今回、特に質的向上が図られたのは、これまでは単一のサービスとして呼び出されていたデータ検索・取得・データ解析などを一連の流れとして XML を用いたワークフローに記述し、より複雑かつ現実のサイエンスに則った処理を柔軟に行なうことを可能にした点である。

最後に、データベース相互の連携、およびデータベースとデータ解析サービスとの連携という点では、JVO の成果は天文学研究・教育の分野で有効であるだけでなく、データベースを活用する多くの分野の研究活動においても応用可能なものであると我々は考えている、ということ述べておく。

文 献

[1] 大石 雅寿, 白崎 裕治, 田中 昌宏, 川野元 聡, 本田

敏志, 水本 好彦, 大江 将史, 安田 直樹, 増永 良文, 石原 康秀, 山崎 昭一, 瓦井 健二, 中本 啓之, 小林 佑介, 吉田 徳夫, "バーチャル天文台を実現する分散データベース・計算資源の国際連携機構", DEWS2005 2005年2月.

- [2] 本田 敏志, 白崎 裕治, 田中 昌宏, 川野元 聡, 大石 雅寿, 水本 好彦, 大江 将史, 安田 直樹, 増永 良文, 石原 康秀, 山崎 昭一, 瓦井 健二, 中本 啓之, 小林 佑介, 吉田 徳夫, "天文学連携データベース (ヴァーチャル天文台) の開発-プロトタイプ第3版の開発-", DEWS2005 2005年2月.
- [3] IVOA <http://www.ivoa.net/>
- [4] 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 2004年3月.
- [5] <http://www.openarchives.org/>
- [6] P. Padovani, M. G. Allen, P. Rosati, N. A. Walton, "Discovery of optically faint obscured quasars with Virtual Observatory tools", *Astronomy and Astrophysics* 424, 545, 2004.