

最新情報技術を活用した超大規模天文データ解析機構の研究開発

研究代表者	大石 雅寿	国立天文台・天文データセンター・助教授
研究分担者	水本 好彦	国立天文台・光赤外研究部・教授
	白崎 裕治	国立天文台・天文データセンター・助手
	大江 将史	国立天文台・天文データセンター・助手
	高田 唯史	国立天文台・天文データセンター・助教授
	安田 直樹	東京大学・宇宙線研究所・助教授

1 研究の概要

研究目的 国立天文台のすばる望遠鏡を始めとして、世界では Gemini, VLT など次々と 8m 級大型望遠鏡が稼働を始め、観測データを大量に産出する時代を迎えた。衛星による観測データも合算すると、天文学研究に利用できる観測データは数値データとして年間 10TB の規模で増大する。これらの様々な波長のデータを同時解析することによって、現代天文学に残された多くの謎—宇宙の暗黒物質の解明、宇宙ひもの存在の検証、宇宙における生命誕生の解明など— が解明されると期待されている。しかし、従来の手作業を主体としたデータ解析手法に基づく研究ではこの膨大なデータを迅速に処理することは極めて困難であり、現代天文学の謎の解明に至るとは到底考えられず、高性能な計算機資源を活用した新たな情報処理技術を天文学に導入することが必須である。

すばる望遠鏡が毎年算出する 30TB に及ぶ膨大な量の観測データを解析して、口径 10m の KECK 望遠鏡や Gemini, VLT といった欧米の 8m クラス望遠鏡との激しい国際競争の中で現在天文学に残された謎を世界に先駆けて解明するためには、我々がこれまで野辺山にある電波望遠鏡やすばる望遠鏡で培ってきたデータ解析技術を GRID や Java を始めとする技術を駆使することにより高度化した分散データ処理システムとして発展させ、かつ、大量観測データの供給源となるデータアーカイブシステムに合体させたサイエンス・アーカイブシステムとして再構築することがカギとなる。

これらの科学的要求を実現するため、我々は近年発展が著しい情報学の研究成果と大量観測データを生み出す最新の望遠鏡技術の融合として Japanese Virtual Observatory (JVO) の構築を進めてきた。JVO は国内にとどまらず世界中に分散している天文データアーカイブ・データベース (DB) やデータ解析機能を連携させ、いつでもどこからでも天文学を推進することが可能な研究基盤を構築することを目的としている。

本年度の研究成果の概要

2 ヴァーチャル天文台とは

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関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) 構想である。

図1にヴァーチャル天文台の概念図を示す。

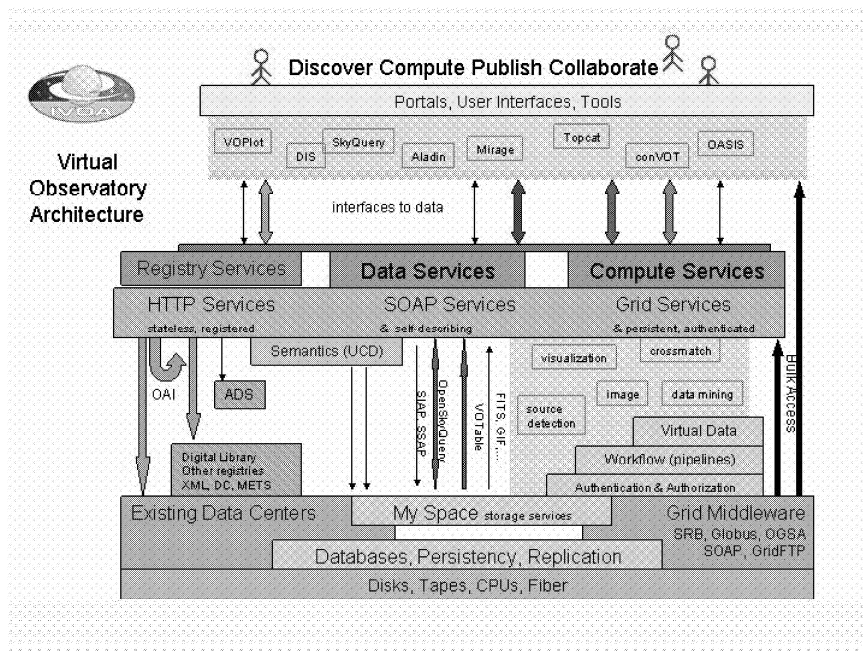


図1: ヴァーチャル天文台の概念図

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合って構成される。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としての安定稼働が保証される。即ち、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。従って、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、2006 年現在、16 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、以下の観点为满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 (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
- 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、即ち、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のち上げと保守

これらの標準化活動の結果、2006年現在、日米欧の主要な天文台やデータセンターがVOインターフェースを通じて相互に接続されている（図2）。

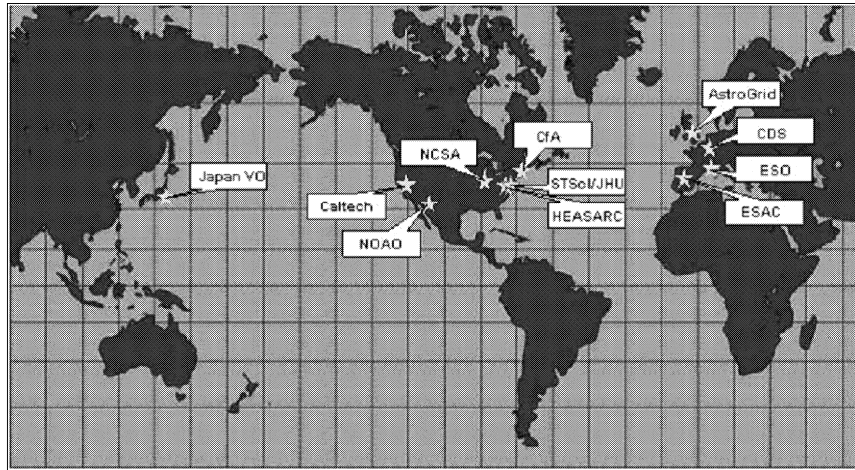


図 2: 2006年現在、相互接続され、ヴァーチャル天文台を構成している世界の天文台やデータセンター

3 データサービスの公開

JVOは、まず、世界にある天文データ資源への容易なアクセスを実現する。このデータサービスを天文コミュニティに提供するための簡易マニュアルや不安定な動作をする部分の強化などの環境整備作業を行い、2006年9月に公式にサービス開始をアナウンスした。そのURLは <http://jvo.nao.ac.jp/portal/> である。データサービスでは、本報告執筆の時点で、200を越えるデータサーバへのアクセスが可能となっている。利用者が上記サイトにブラウザからアクセスすると、アカウント登録なしでも guest user としてサービスが利用できる。ただし、guest user の場合は、取得した天文データ等を一時的に利用できるだけであり、永続的な使用はできない。これは guest user が、ヴァーチャル天文台とはどのような機能を持っているかを気軽に利用してもらうための“お試し”との位置づけであるからである。検索結果等を JVO ポータルサーバに永続的に蓄積するためには、ユーザー登録をすればよい。

4 ワークフロー機構の研究開発

ヴァーチャル天文台は天文学の新しい研究基盤を目指している。そのためには天文データの共有だけでは不十分で、大規模データを解析して宇宙の新しい知見を見いだす環境を提供する必要がある。そのためワークフロー処理機構の標準化を国際共同の元で進めた。

4.1 ワークフロー処理言語の定義

まず、ビジネス向けのワークフロー言語である BPEL4WS (Business Process Execution Language for Web Services) を参照しながら天文解析に必要なワークフロー記述言語 (WFDL) を定義した。これは BPEL4WS がビジネス向けに定義されたものであり、既存のデータ解析プログラムへの入力パラメータを渡したり、ループ処理や異なる天空域のデータを独立に並列処理できるようタグ定義を拡張する必要があったためである。定義した主なタグを以下に示す。

- 変数定義：<variables> ~ </variables>, <variable> を要素にとる
- 外部サービス呼出：<invoke> ~ </invoke>, <input>・<output> を要素にとる, 属性値でサービスを指定
- 内部コマンド実行：<command> ~ </command>, <input>・<output> を要素にとる
- 逐次実行：<sequence> ~ </sequence>, <command>などを要素にとる
- 単純ループ処理：<for> ~ </for>, 要素を順次実行する
- 並列ループ処理：<parfor> ~ </parfor>, 要素を並列実行する
- 条件判定：<if> ~ </if>

定義したこれらの XML タグの相互関係を図 3 に示す。

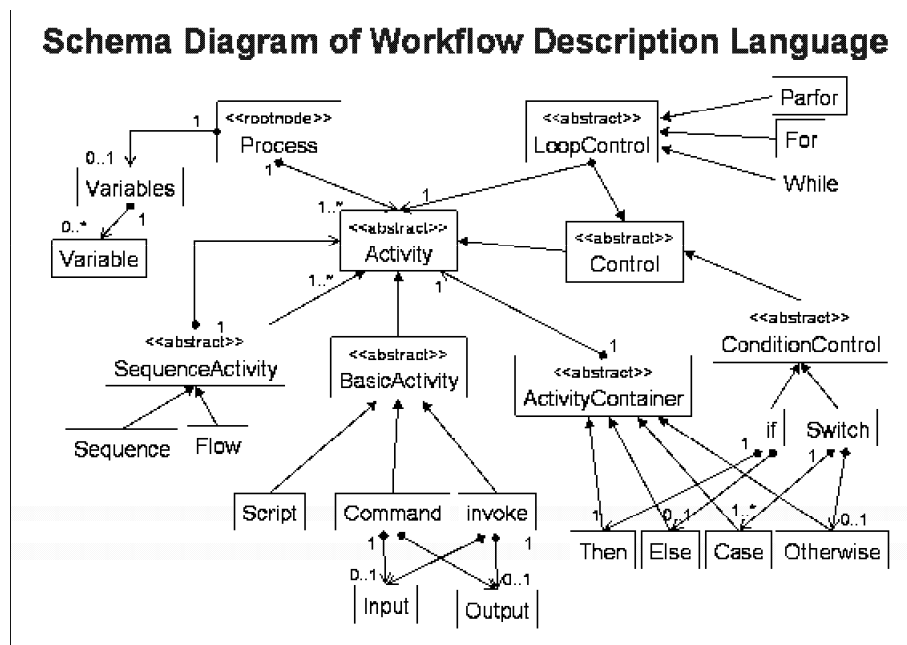


図 3: ワークフロー処理言語を記述するタグ類の関係

4.2 ワークフロー実行機構の構築

WFDL で記述された各処理ステップは、まず、Java インタプリタである Groovy に変換され、Groovy スクリプトを Java スクリプトに変換された後に実行される。これにより世界各地の解析サーバで実行できるワークフロー実行が可能となる (図 4)。Java / Groovy を用いたのは、実行されるプラットフォーム依存性を排除するためである。

ワークフローの各ステップは適切なサーバで実行される必要がある。このため、使用可能なサーバの状況を MDS (Monitor and Discovery Service) サーバが常にモニターし、ワークフローを解釈するポータルからの問い合わせに応じて適切なサーバを通知するようにした。ポータルはその情報に基づいて実際にジョブを解析サーバに投入する。処理が終わると終了情報がポータルに通知されて次のステップに進む。図 5 に、構築したワークフローの実行状況を示すモニター画面のスナップショットを示す。この画面は自動的にリフ

リフレッシュされ、ユーザーは容易にワークフローの進行状況を把握できる。その後、利用者は最終結果を受け取ることとなる（図6）。

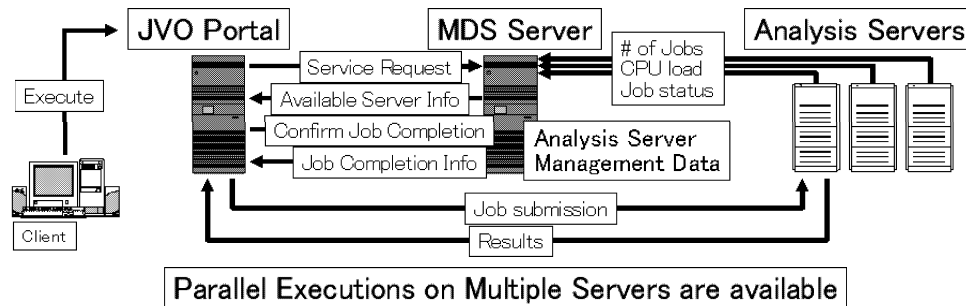


図 4: データ処理サーバの負荷モニター機構

図5に、構築したワークフローの実行状況を示すモニター画面のスナップショットを示す。この画面は自動的にリフレッシュされ、ユーザーは容易にワークフローの進行状況を把握できる。

Workflow Status

[Status](#) | [Registry](#) | [Search](#) | [Workflow](#) | [Result](#) | [Database](#) | [QSO](#) | [DataViewer](#) | [Link](#) | [MemoryMonitor](#) | [Logout](#)
 ⇒ [All](#) | [Detail](#)

Workflow Name : work_20060123200255167

Activity Name	Host	Elapsed Time (s)	Flag	Status
1_executeQuery	cda.harvard.edu	1.93		success
		0.0		success
2_executeQuery	jvo.nao.ac.jp	2.143		success
		0.0		success
3_executeQuery	pma.iso.vilspa.esa.es	4.632		success
		0.0		success
4_executeQuery	www.cadc-ccda.hia-ih.nrc-cnrc.gc.ca	0.0		executing
		0.0		waiting
5_executeQuery		0.0		waiting

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Mon Jan 23 19:56:18 JST 2006

図 5: ワークフローの実行例

5 今後の予定

最新情報技術を活用することにより、いつでもどこでも天文学研究を推進する基盤ができつつある。しかし現状での計算資源共有は公開DBに限定されている。VOが普及すると、未公開観測データもVOで処理したいという要請が出てくると予想される。従って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。そのためには世界各地で開発が進められているGrid middlewareが提供する共通認証機構を活用することとなるであろうが、そのために

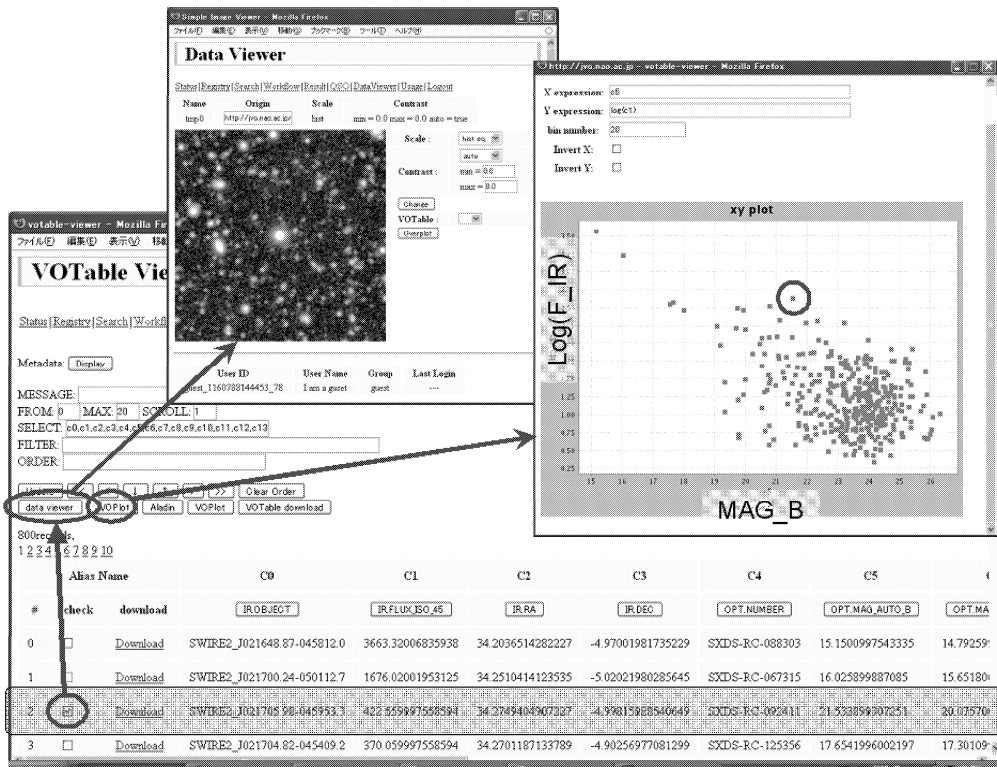


図 6: データ検索・解析の実例

は複数存在する Grid middleware 間の相互運用性が確保されていなければならない。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を VO の上で利用できるようにするための共通インタフェースの開発が求められている。今後はこれらの課題を踏まえながら大規模な実用 VO システムの構築を進めてゆきたい。

最後に、JVO の基本的機構は、天文学にとどまらず DB を活用する他の研究分野にも応用可能であることを指摘しておきたい。研究分野ごとに DB に格納している属性値やデータ保護の考え方は異なるが、巨視的に見れば共通の枠組みが利用できると考えられる。

研究成果リスト

著書、論文

1. Nakagawa, Yujin E., Yoshida, Atsumasa, Shirasaki, Yuji, 他: “An Optically Dark GRB Observed by HETE-2: GRB 051022”, Publications of the Astronomical Society of Japan, Vol. 58, No. 4, pp.L35–L39, 2006.
2. Kawabata, K. S., Ohyama, Y., Ebizuka, N., Takata, T., Yoshida, M., Isogai, M., Norimoto, Y., Okazaki, A., Saitou, M. S.: “Low- and Medium-Dispersion Spectropolarimetry of Nova V475 Scuti (Nova Scuti 2003): Discovery of an Asymmetric High-Velocity Wind in a Moderately Fast Nova”, The Astronomical Journal, Vol. 132, pp.433–442, 2006.
3. Simpson, C., Martinez-Sansigre, A., Rawlings, S., Ivison, R. J., Akiyama, M., Sekiguchi, K., Takata, T., Ueda, Y., Watson, M.: “Radio imaging of the Subaru/XMM-Newton Deep Field - I. The 100-microJy catalogue, optical identifications, and the nature of the faint radio source population”,

MNRAS, Vol. 372, pp.741–757, 2006.

4. Takata, T.,Sekiguchi, K.,Chapman, S. C.,Geach, J. E.,Swinbank, A. M.,Blain, A.,Ivison, R. J.: “Rest-Frame Optical Spectroscopic Classifications for Submillimeter Galaxies”, *The Astrophysical Journal*, Vol. 651, pp.713–727, 2006.
5. Kawai, N.,Kosugi, G.,Aoki, K.,Yamada, T.,Mizumoto, Y.,Takata, T.,Shirasaki, Y., 他: “An optical spectrum of the afterglow of a γ -ray burst at a redshift of $z = 6.295$ ”, *Nature*, Vol. 440, pp.184–186, 2006.
6. Enomoto, R.,Tsuchiya, K.,Adachi, Y.,Kabuki, S.,Edwards, P. G.,Mizumoto, Y., 他: “A Search for Sub-TeV Gamma Rays from the Vela Pulsar Region with CANGAROO-III”, *The Astrophysical Journal*, Vol. 638, pp.397–408, 2006.
7. Kashikawa, Nobunari,Yoshida, Makiko,Mizumoto, Yoshihiko,Yasuda, Naoki, 他: “Clustering of Lyman Break Galaxies at $z = 4$ and 5 in the Subaru Deep Field: Luminosity Dependence of the Correlation Function Slope”, *The Astrophysical Journal*, Vol. 637, pp.631–647, 2006.
8. Watanabe, K.,Gros, M.,Stoker, P. H.,Kudela, K.,Lopate, C.,Valdes-Galicia, J. F.,Hurtado, A.,Musalem, O.,Ogasawara, R.,Mizumoto, Y., 他: “Solar Neutron Events of 2003 October–November”, *The Astrophysical Journal*, Vol. 636, pp.1135–1144, 2006.
9. Akiko Aizawa and Jun Adachi: “Information Retrieval from Bursty Information”, *Proceedings of Intl. Conf. on Informatin Retrieval*, pp.312–324, 2006.
10. 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, 他: “JVO の研究開発 (全体進捗)”, 日本天文学会 2006 年春季年会., 2006.
11. 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, 他: “JVO の研究開発 (ワークフロー機能の実装)”, 日本天文学会 2006 年春季年会., 2006.
12. 田中 昌宏, 白崎 裕治, 川野元 聡, 本田 敏志, 大石 雅寿, 水本 好彦, 大江 将史, 安田 直樹, 増永 良文, 石原 康秀, 他: “JVO の研究開発 (JVO ポータルのデモンストレーション)”, 日本天文学会 2006 年春季年会., 2006.
13. 白崎 裕治: “バーチャル天文台を利用した QSO 周辺環境の統計的調査”, 日本天文学会 2006 年春季年会., 2006.
14. 安田 直樹, 土居 守, 諸隈 智貴, 高梨 直紘, 時田 幸一, 井原 隆, 小西 功記, Scot Kleinman, 新田 敦子: “SDSS Supernova Survey 2005”, 日本天文学会 2006 年春季年会., 2006.
15. Zheng, C.,Becker, A.,Yasuda, N.: “SDSS-II Supernova Survey - Fall 2005 Spectroscopy Results”, *American Astronomical Society*, 208, 02.02-(2006), 2006.
16. 前田 啓一, 川端 弘治, 田中 雅臣, 野本 憲一, 高田 唯史: “後期スペクトルで探る Ib/c 型超新星爆発の性質”, 日本天文学会 2006 年秋季年会, 2006.
17. 下川辺 隆史, 河合 誠之, 小谷 太郎, 谷津 陽一, 高田 唯史, 吉田 道利, 柳澤 顕史, 黒田 大介: “MITSuME 望遠鏡データベースの開発”, 日本天文学会 2006 年秋季年会, 2006.

18. 吉野 彰, 榎 基宏, 仲田 史明, 山田 善彦, 八木 雅文, 高田 唯史, 市川 伸一, 小澤 友彦, 洞口 俊博: “光赤外望遠鏡公開データアーカイブシステム SMOKA の新機能版の公開”, 日本天文学会 2006 年秋季年会, 2006.
19. Watanabe, K., Muraki, Y., Matsubara, Y., Murakami, K., Mizumoto, Y.: “Solar neutron events in association with large solar flares in November 2003”, *Advances in Space Research*, 38, 425-430, 2006.

招待講演

1. 大石 雅寿: “ヴァーチャル天文台の研究開発 – 楽しいことも苦勞もある – (招待講演)”, 情報・システム研究機構 e-science と field science が築くこれからの学問, 2006.
2. 大石 雅寿: “Recent Results on Small Carbon Molecules (招待講演)”, *Carbon in Space*, Italy, 2006.
3. 大石 雅寿: “天文学における超大規模分散データベースの活用 ~ ヴァーチャル天文台 (招待講演)”, 筑波大学計算科学センターセミナー, 2006.
4. 大石 雅寿: “International Virtual Observatory: A New, Planet-wide Research Infrastructure for Astronomy (招待講演)”, Korea Science and Technology Infrastructure Conference, Seoul, Korea, 2006.
5. 大石 雅寿: “21 世紀の新しい天文学研究基盤 ~ ヴァーチャル天文台 (招待講演)”, 理化学研究所講演会 (2006 年 10 月 24 日), 2006.
6. 大石 雅寿: “Virtual Observatory (招待講演)”, IAU General Assembly 2006, Special Session 1, Prague, Czech, 2006.
7. 大石 雅寿: “Recent Progress in Building the Virtual Observatories in the World (招待講演)”, 中国国家天文台講演会 (2006 年 7 月 19 日), 北京, 2006.
8. 大石 雅寿: “Building the International Virtual Observatory Alliance (IVOA): Programmatic Challenges (招待講演)”, COSPAR General Assembly 2006, E1.5, Beijing, China, 2006.
9. 白崎 裕治: “国立天文台におけるバーチャル天文台開発状況 (招待講演)”, 「宇宙地球系情報科学研究会」および「地球科学データの収集と公開に関する研究集会」, 2006.
10. 白崎 裕治: “Constructing the Subaru Advanced Data and Analysis Service on VO (招待講演)”, *Astronomical Data Analysis Software and Systems XVI*, Tucson, Arizona, USA, 2006.