

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・准教授
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・助教
	大江 将史	国立天文台・天文データセンター・助教
	高田 唯史	国立天文台・天文データセンター・准教授
	安田 直樹	東京大学・宇宙線研究所・准教授
研究協力者	田中 昌宏	国立天文台・天文データセンター・研究員

1 研究目的の概要

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。

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。一般的に天体は多波長で放射をしているため、各種天体現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。一方、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) 構想であり、その構築にあたっては世界の主要国が協力して相互の資源を活用するための標準を定めてきた [1]。

これらの標準化活動の結果、2008年12月現在、1600を越える日米欧の主要な天文台やデータセンターにあるリソースがVOインターフェースを通じて相互に接続されている。

3 ローカルグリッド環境の構築

JVOでは世界の天文台などが保持する天文データをシームレスにアクセス・取得することが可能となった。本研究開発の目的の一つである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ためには、取得したデータを解析し、様々な天体现象の理解に繋げる環境を整える必要がある。

天文学におけるデータ解析では、当然ながら、様々なソフトウェアパッケージ(例:光赤外観測データ解析のためのIRAF、電波天文観測データを解析するAIPS)をはじめ、世界各地で開発されたソフトウェアを利用することが想定される。そこで、遠隔地に存在するデータ解析ソフトウェアとの連携に向け、ローカル環境にある複数サーバをシームレスに連携しデータ解析処理を進める環境を試験構築した。

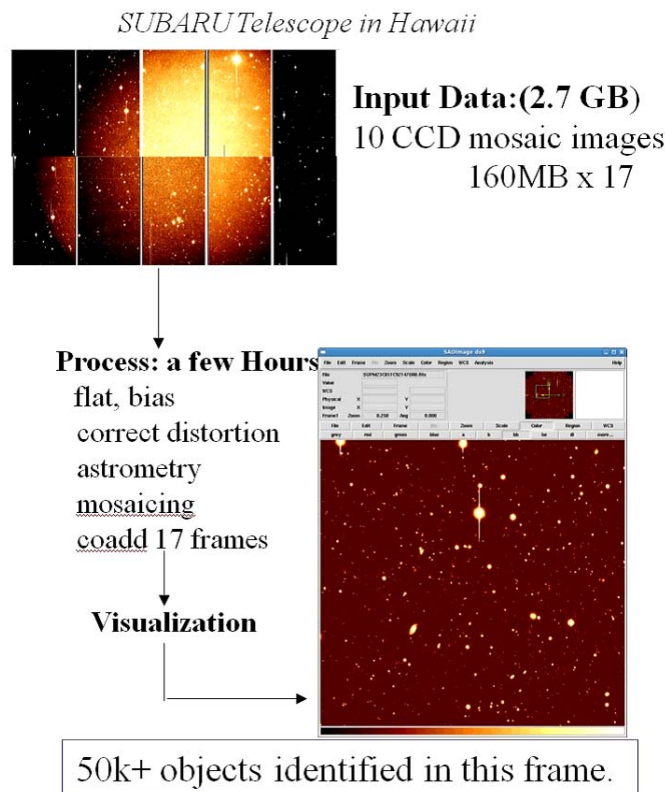


図1: すばる望遠鏡によって取得した天体画像の自動モザイク処理

そのユースケースとして、国立天文台ハワイにあるすばる望遠鏡に設置された大型カメラである Suprime-CAM により取得された CCD 画像の自動処理を選んだ。SuprimeCAM は 10 枚の CCD チップから構成される。CCD チップの間には隙間があり、観測時には隙間を埋めるようにカメラを移動させながら何枚も画像を取得し、観測後に CCD チップ内の感度ムラの補正 (flat) や暗電流の補正 (bias)、また望遠鏡の光学系特性による画像歪みの補正 (distortion) や画像内の天体位置測定 (astrometry) を行った後、位置合わせをしながら画像のつなぎ合わせ (モザイク) 処理や重ね合わせをすることになる。従来はこれらの処理は各観測者が自分自身で行っていたが、ヴァーチャル天文台を経由して SuprimeCAM データを活用したいという

要望は強く、この処理を誰でも利用できるようにすることで天文学コミュニティの要請に応えられるものと考えた。

図 1 に、国立天文台三鷹キャンパス内のローカル環境に構築したシステムのフローを示す。本システムでは、図 1 の例で言えば、17 回の観測で取得した画像（各 CCD 毎に 1 枚の画像が得られるので総計 170 枚の画像、データ量は 2.7 GByte）を、それぞれの観測毎に、flat, bias, distortion, astrometry の処理を複数サーバで並列に処理し、得られた処理済み画像を集積した後に位置合わせをしながら重ね合わせやモザイク処理を行う。この例では全ての処理を終えるのに数時間を要するが、利用者は JVO ポータルに用意されたインターフェースにおいて処理したいデータを指定するだけであり、ジョブを投入した後はオフラインで処理終了を待っていれば良い。そして得られたモザイク済み画像を、JVO が提供する画像ビューワで見ることになる。

4 グリッドミドルウェアによる他研究機関との連携試験

前章に述べた成果を踏まえ、次のステップとして国立天文台以外の研究機関との連携試験を実施した。解析ソフトウェアを遠隔サーバで実施するための環境として、国立情報学研究所が中心になって開発したグリッドミドルウェアである NAREGI を試用することとした。連携試験のパートナーとしては、これまでも共同研究等を行ってきた相手である高エネルギー加速器研究機構 (KEK) にお願ひし、両研究機関に NAREGI ミドルウェアバージョン $\beta 2.0.1$ をインストールし、前章に示したデータ処理を両研究機関のサーバで実施した。図 2 に、その連携環境を示す。

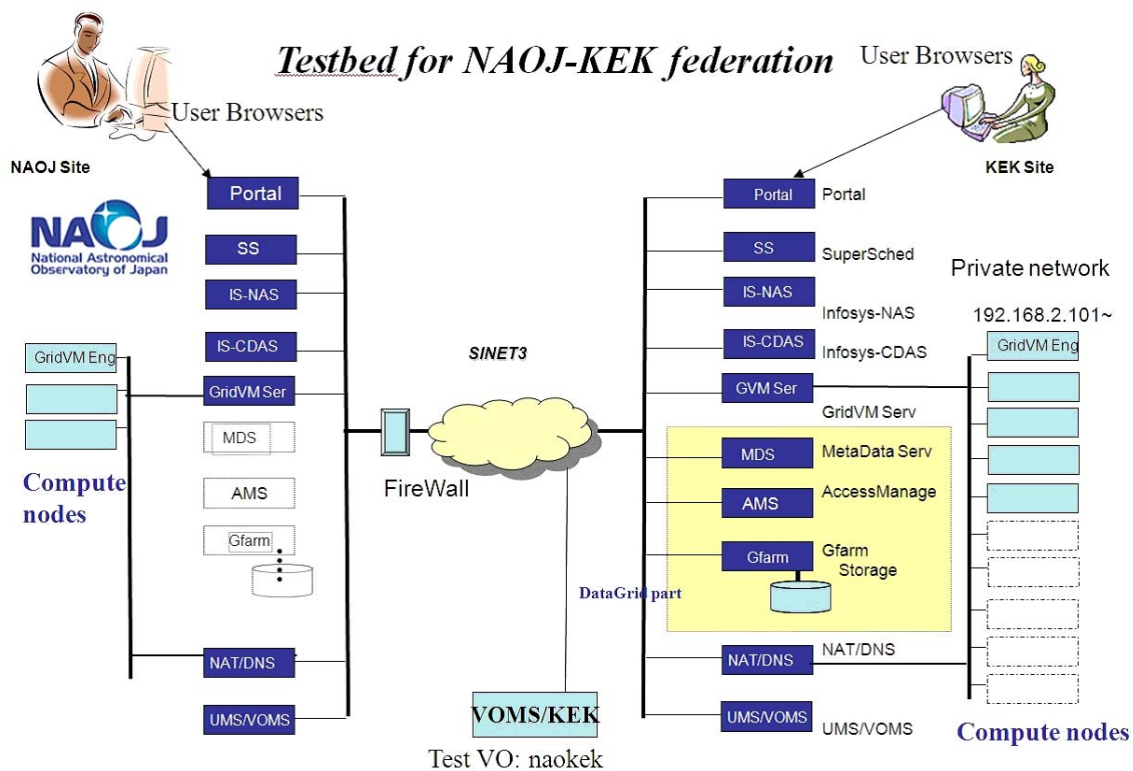


図 2: 国立天文台-KEK 間のグリッド連携環境

この連携環境において、flat, bias, distortion, astrometry などを行う天文データ解析プログラム（エンジン）を NAREGI PSE(Program Solving Environment) 上で動作させるための基本機構を構築し、その

実行制御のために、これらの天文データ解析エンジン等を NAREGI Workflow Tool で利用した。データ解析プログラムを実行するためには、相手方（今回は KEK）のサーバ上に解析プログラムのソースコードを GridFTP 転送し、KEK のサーバ上でコンパイル処理・登録処理をしなければならない。国立天文台と KEK はいずれも SINET3 のノードとなっているため、プログラム転送を含めてもコンパイル等に必要な経過時間はローカルに実行する場合と大差がないことが分かった。

一方、NAREGI の Super Scheduler (SS) を通じてジョブを実行する場合には、国立天文台の計算資源と KEK の計算資源を利用する場合で、経過時間に大きな差が生じることが分かった。例えば、国立天文台の計算資源にジョブを投入する時には 38 秒でジョブが開始するのに対し、KEK にジョブを投入する場合には 3 分 27 秒かかるという例があった。これは、SS がジョブを実行するまでに多くのパケットのやり取りを行っており、通信負荷が相当大きいことが考えられる。しかし、すばる望遠鏡による大量画像処理など CPU を多く利用するジョブの場合は、相対的に通信負荷の割合が小さくなるので、実際には大きな問題にはならないと考えられる。

この連携試験環境では、KEK から国立天文台の計算資源を利用する高エネルギー物理用の処理プログラムも実行されており、VPN を介さない本当に意味でのグリッド環境が NAREGI ミドルウェアにより試験構築できた意義は大きいと考えられる。

5 JVO ポータルの公開状況

JVO ポータルを通じて世界の天文データへのアクセス機能を 2008 年度の直前より公開を始めた。その URL は <http://jvo.nao.ac.jp/portal/> である。公開開始以降のポータルへのアクセス状況を見ながらより良い機能を研究開発できるよう、そのアクセスログを取る機能を盛り込んである。図 3 はその機能を用いて作成した本報告書執筆直前までのアクセスログを示す。

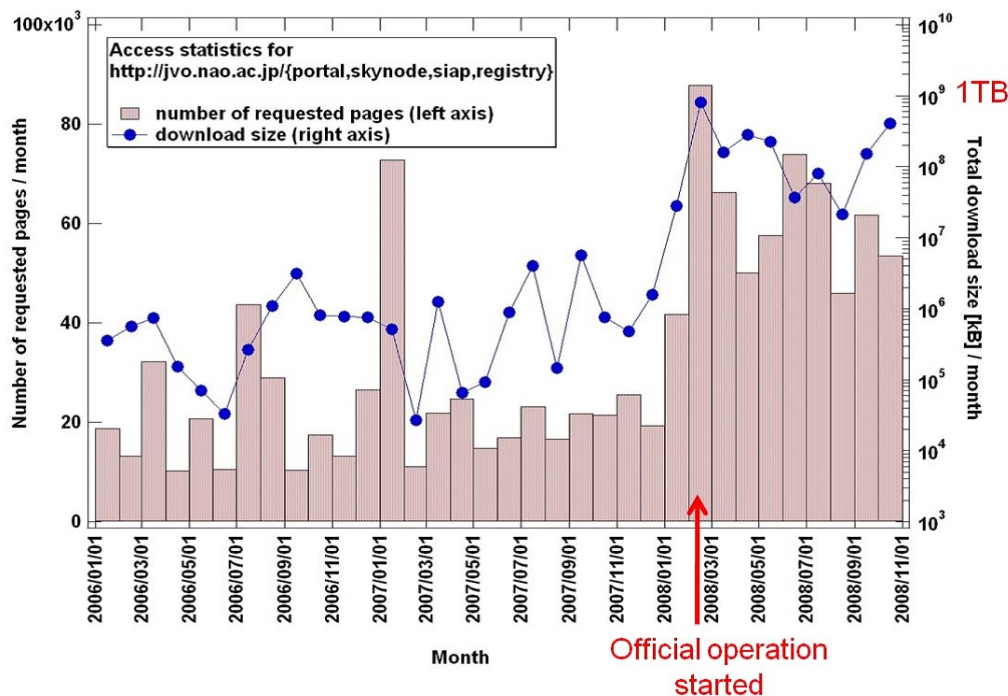


図 3: 2008 年 11 月までの JVO へのアクセス状況

JVO のデータサービスシステムを正式に公開したのは 2008 年 3 月末であるが、実際には 2 年前より試

験公開という位置づけでシステムは関係者には公開されていた。このためアクセスログは2006年当初から取っている。図3の縦軸は、対数スケールになっていることに注意していただきたい。

JVOのログ機能では、JVOを経由して他のデータサーバにデータ請求を出した場合のダウンロード量などは分からない。これは、国際ヴァーチャル天文台連合(IVOA)のデータアクセスプロトコルは、そのデータサーバに対してデータ請求をするからであり、各天文台のヴァーチャル天文台システムは、利用者の要求を満たすためにはどこにアクセスするべきかを水面下で調査し、利用者が知らないうちにその利用要求を必要サーバに転送してしまうからである。従って、このログは、JVOが直接管理するデータサービスに対する要求だけを記録している。

2008年3月の正式公開時に、JVOシステムに対する要求ページ数(左軸)やダウンロードされたデータ量(右軸)が1-2桁増加したことが明確に分かる。正式公開以降は、夏休み中にアクセスが低下したものの夏の終わりと共にアクセス量が増えている。ダウンロードされたデータ量は1月当たり最大1Tbytesにも及び、ネットワークを通して天文観測データを様々に活用したいという要請に大きく応えることができたことが理解される。

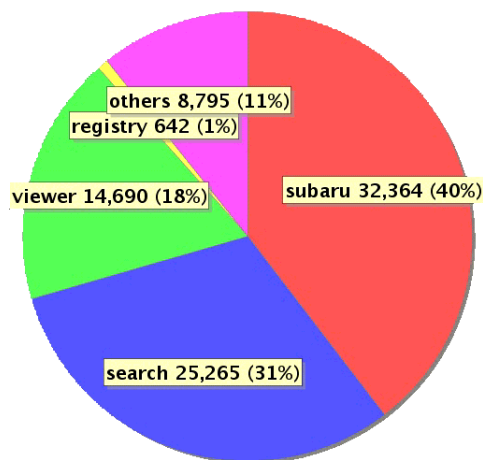


図4: JVOへのアクセスの分類

さらに、どのデータに対するアクセスが多いのかをログから調査してみると、図4に示すように、すばる望遠鏡によるデータへのアクセスが約40%、他のJVO管理のデータに対する検索要求が約30%となっている。現在JVOには、利用頻度が高いと思われる大規模カタログ(SDSSや2MASS等)を登録しており、情報技術を活用した大規模データへの容易なアクセスを実現したことにより、国内のみならず海外の天文研究者などの利便性を大きく向上できたことが分かる。

6 今後の展望

今後は、下記の観点を持って、JVOシステムの機能向上のための研究開発を実施し、情報技術によって様々な科学分野においてその研究基盤が飛躍的に向上できることを示してゆきたい。

- より大規模なデータをより高速に検索するための、並列検索システムの研究開発
- そのようにして得られた天文データをより高速にデータ処理するための遠隔並列処理システム
- 言うまでもないことであるが、検索可能なデータを増やすための、JAXA/ISASをはじめとする国内外の研究機関との連携強化

参考文献

- [1] 大石雅寿 他: “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”, 平成 18 年度特定領域研究「情報爆発」成果報告書, 2006.

研究成果リスト

著書、論文

1. Mizumoto Y., 他: “Observation of Very High Energy Gamma Rays from HESS J1804-216 with CANGAROO-III Telescopes”, *Astrophysical Journal*, Vol. 683, pp.957–966, 2008.
2. Mizumoto Y., 他: “CANGAROO-III Search for Gamma Rays from Kepler’s Supernova Remnant”, *Astrophysical Journal*, Vol. 683, pp.383–388, 2008.
3. Mizumoto Y., 他: “Observation of an Extended Very High Energy Gamma-Ray Emission from MSH 15-52 with CANGAROO-III”, *Astrophysical Journal*, Vol. 677, pp.297–305, 2008.
4. Mizumoto Y., 他: “The JEM-EUSO Project: Observing Extremely High Energy Cosmic Rays and Neutrinos from the International Space Station”, *Nuclear Physics B Proceedings Supplements*, Vol. 175, pp.237–240, 2008.
5. Yasuda N., 他: “The Sloan Digital Sky Survey-II Photometry and Supernova IA Light Curves from the 2005 Data”, *Astronomical Journal*, Vol. 136, pp.2306–2320, 2008.
6. Takahashi N., Doi M., and Yasuda N.: “Light-curve studies of nearby Type Ia supernovae with a Multiband Stretch method”, *MNRAS*, Vol. 389, pp.1577–1592, 2008.
7. Yasuda N., 他: “A Measurement of the Rate of Type Ia Supernovae at Redshift $z < 0.1$ from the First Season of the SDSS-II Supernova Survey”, *Astrophysical Journal*, Vol. 682, pp.262–282, 2008.
8. Takata T., Yasuda N., Mizumoto Y., 他: “The Subaru/XMM-Newton Deep Survey (SXDS). II. Optical Imaging and Photometric Catalog”, *Astrophysical Journal Supplement Series*, Vol. 176, pp.1–18, 2008.
9. Yasuda N., 他: “First-Year Spectroscopy for the Sloan Digital Sky Survey-II Supernova Survey”, *Astronomical Journal*, Vol. 135, pp.1766–1784, 2008.
10. Oda, T., Totani T., Yasuda N., Sumi T., Morokuma T., Doi M., and Kosugi G.: “Implications for Galaxy Evolution from Cosmic Evolution of the Supernova Rate Density”, *PASJ*, Vol. 60, No. 2, pp.169–182, 2008.
11. Shirasaki Y., Tanaka M., Kawanomoto S., Honda S., Ohishi M., Mizumoto Y., Yasuda Y., Masunaga Y., Ishihara Y., Tsutsumi J., 他: “Data processing for ‘SUBARU’ telescope using GRID”, *Fusion Engineering and Design*, Vol. 83, pp.438, 2008.
12. Mizumoto Y., Shirasaki Y., 他: “The cosmic ray primary composition at the knee region from lateral distributions of atmospheric Cerenkov photons in extensive air showers”, *Astroparticle Physics*, Vol. 29, pp.453, 2008.
13. Shirasaki Y., 他: “Intrinsic properties of a complete sample of HETE-2 gamma-ray bursts. A measure of the GRB rate in the Local Universe”, *Astronomy and Astrophysics*, Vol. 491, pp.157–171, 2008.

14. Kitagawa M. and Ohishi M.: “Measurements of the Radiated Electric Field and the Common Mode Current from the In-house Broadband Power Line Communications in Residential Environment”, EMC Europe 2008, pp.433–438, 2008.
15. Tanaka M.,Shirasaki Y.,Ohishi M.,Mizumoto Y.,Kawanomoto S.,Yasuda N. and Honda S.: “Astronomical Researches using Virtual Observatories”, Fifty Years after IGY, 2008.
16. Tanaka M.: “Water ice absorption map of Pipe Nebula”, Workshop for Interstellar Matter 2008, 2008.
17. Shirasaki Y.,Tanaka M.,Ohishi M.,Mizumoto Y.,Kawanomoto S.,Honda S.,Yasuda Y.,Masunaga Y.,Ishihara Y.,Machida Y., 他: “The Japanese Virtual Observatory in action”, The Astronomical Data Analysis Software and Systems Conference, 2008.
18. Shirasaki Y.: “Design and Implementation of the Japanese Virtual Observatory (JVO) system”, The 10th Asian-Pacific Regional IAU Meeting, 2008.
19. Ohishi M.: “Special Item: ITU/Leap Second”, IVOA interoperability meeting 2008 Autumn, 2008.
20. Ohishi M.: “JVO Spotlight”, IVOA interoperability meeting 2008 Autumn, 2008.
21. Shirasaki Y. 他: “Substructure spectral analysis of HETE-2 burst GRB060115”, GAMMA-RAY BURSTS 2007: Proceedings of the Santa Fe Conference. AIP Conference Proceedings, Volume 1000, 2008.
22. Shirasaki Y. 他: “Spectrum Feature of the Underlying Soft Component of GRB041006”, GAMMA-RAY BURSTS 2007: Proceedings of the Santa Fe Conference. AIP Conference Proceedings, Volume 1000, 2008.
23. Shirasaki Y. 他: “Development of A New Background Reduction Method for WXM/HETE-2 and Its Application for Bright GRB Spectra”, GAMMA-RAY BURSTS 2007: Proceedings of the Santa Fe Conference. AIP Conference Proceedings, Volume 1000, 2008.
24. Tanaka M.,Shirasaki Y.,Ohishi M.,Mizumoto Y.,Ishihara Y.,Tsutsumi J.,Machida Y.,Nakamoto H.,Kobayashi Y., and Sakamoto M.: “Construction of Multiple-Catalog Database for JVO”, Astronomical Data Analysis Software and Systems ASP Conference Series, Vol. 394, pp.261–264, 2008.
25. Ohishi M.: “JVO and NaReGi (Japanese Grid middleware initiative)”, Mem. Soc. Astr. It., pp.in press, 2008.
26. Ohishi M.: “Astrobiology-related research activity in the National Astronomical Observatory of Japan”, Proceedings of the SPIE, Volume 7097, 2008.
27. 小林 かおり, 常川 省三, 松島 房和, 山本 智, 大石 雅寿: “ALMA に向けた分子線データベース (Toyama Microwave Atlas) の公開”, 日本天文学会 2008 年秋季年会, 2008.
28. 田中 昌宏: “Pipe Nebula における H₂O 氷の吸収の空間分布”, 日本天文学会 2008 年秋季年会, 2008.
29. 大石 雅寿: “NAREGI による分散天文データ 解析機構の試験構築”, 国立情報学研究所 CSI 交流報告会 2008, 2008.

30. 井上 一, 小山 勝二, 高橋 忠幸, 水本 好彦: “宇宙の観測 (3) 高エネルギー天文学”, シリーズ現代の天文学, 日本評論社, 2008.

招待講演

1. Ohishi M.: “Astronomical Virtual Observatories through International Collaborations (招待講演)”, Fifty Years after IGY, 2008.
2. Ohishi M.: “IUCAF activities in the International Telecommunication Union (ITU) (招待講演)”, 2008 URSI General Assembly, 2008.
3. Ohishi M.: “Astrobiology-related research activity in the National Astronomical Observatory of Japan (招待講演)”, SPIE, Astrobiology conference, 2008.
4. Ohishi M.: “Protection of Space-based observations and the International Telecommunication Union (ITU) (招待講演)”, The 37th COSPAR Scientific Assembly, session E110, 2008.
5. Ohishi M.: “Construction of the Virtual Observatory in the World (招待講演)”, The 10th Asian-Pacific Regional IAU Meeting, 2008.
6. Ohishi M.: “JVO and NaReGi (Japanese Grid middleware initiative) (招待講演)”, EURO-VO DCA Workshop “Grid and the Virtual Observatory”, 2008.