

すばる望遠鏡公開データアーカイブシステムの開発

馬場 肇^{*1}, 安田直樹, 市川伸一, 八木雅文, 岩本信之, 高田唯史,
洞口俊博^{*2}, 多賀正敏^{*3}, 渡邊 大^{*4}, 奥村真一郎^{*4},
小澤友彦^{*5}, 山本直孝^{*6}, 濱部 勝^{*7}

(2002 年 3 月 29 日受理)

Development of Public Science Archive System of Subaru Telescope

Hajime BABA^{*1}, Naoki YASUDA, Shin-ichi ICHIKAWA, Masafumi YAGI, Nobuyuki IWAMOTO,
Tadafumi TAKATA, Toshihiro HORAGUCHI^{*2}, Masatoshi TAGA^{*3}, Masaru WATANABE^{*4},
Shin-ichiro OKUMURA^{*4}, Tomohiko OZAWA^{*5}, Naotaka YAMAMOTO^{*6},
and Masaru HAMABE^{*7}

Abstract

We have developed a public science archive system, Subaru-Mitaka-Okayama-Kiso Archive system (SMOKA), as a successor of Mitaka-Okayama-Kiso Archive (MOKA) system. SMOKA provides an access to the public data of Subaru Telescope, the 188 cm telescope at Okayama Astrophysical Observatory, and the 105 cm Schmidt telescope at Kiso Observatory of the University of Tokyo. Since 1997, we have tried to compile the dictionary of FITS header keywords. The accomplishment of the dictionary enabled us to construct an unified public archive of the data obtained with various instruments at the telescopes. SMOKA has two kinds of user interfaces; Simple Search and Advanced Search. Novices can search data by simply selecting the name of the target with the Simple Search interface. Experts would prefer to set detailed constraints on the query, using the Advanced Search interface. In order to improve the efficiency of searching, several new features are implemented, such as archive status plots, calibration data search, an annotation system, and an improved Quick Look Image browsing system. We can efficiently develop and operate SMOKA by adopting a three-tier model for the system. Java servlets and Java Server Pages (JSP) are useful to separate the front-end presentation from the middle and back-end tiers.

1. はじめに

天文学は、基本的に観測を柱とした学問である。観測データは、決して繰り返すことのない宇宙の歴史の一瞬をとらえた記録であり、観測時刻における天体の情報を持つ唯一の存在である。これらを散逸しないように管理し保存することがアーカイブの第一の目的である。一つの観測データの中に含まれる情報量は非常に多いため、アーカイブシステムを整備しておくことにより、元々の観測者の意図した情報とは別の観点を持つ研究に役立てることができる場合がある。たとえば、複数の観測データを組み合わせれば、変

光天体あるいは移動天体などの位置と明るさといった物理量の変化の追跡や固有運動の検出など、天体の時間変動する現象を追うことができるようになる。新天体が発見されたときに、過去のデータに遡って天体の物理量を調査することもできる。あるいは、大規模サンプルを用いて種々の天体の統計的性質を調べることもできる。同じ天体を可視と X 線と電波などの複数の波長域にまたがって調査することができるようになる。

ここ数年で、観測データのアーカイブシステム、特に地上観測データの公開アーカイブシステム（たとえば ESO における VLT アーカイブ¹⁾ など）が世界中で整備されるようになってきた。前述のアーカイブの持つ元来の意義に加え、アーカイブ整備の動きに拍車をかけているのは、近年の望遠鏡の大型化および観測装置の高性能化に伴ったデータ取得にかかるコストの肥大化傾向である。このため、望遠鏡の運用効率を向上させ、観測者以外の研究者のデータの再利用を促し、データをより有効に活用するべきであるという考え方が広まりつつあるのである。さらに、ネット

^{*1}宇宙科学研究所 (The Institute of Space and Astronautical Science)

^{*2}国立科学博物館 (National Science Museum)

^{*3}早稲田大学 (Waseda University)

^{*4}宇宙開発事業団 (National Space Development Agency of Japan)

^{*5}みさと天文台 (Misato Observatory)

^{*6}東京理科大学 (Science University of Tokyo)

^{*7}日本女子大学 (Japan Women's University)

表 1. 利用可能なアーカイブデータ.
(2002 年 5 月 29 日現在)

望遠鏡	観測装置名	観測 モード	アーカイブされた フレーム数	専有期間
すばる	Suprime-Cam	撮像	19,466	18ヶ月
	FOCAS	撮像	1,126	
		分光	980	
	HDS	分光	1,240	
	OHS/CISCO	撮像	52,791	
		分光	4,157	
	CIAO	撮像	8,698	
	COMICS	撮像	1,261	
		分光	1,257	
	IRCS	撮像	4,630	
		分光	1,580	
	CAC	撮像	5,895	
	MIRTOS	撮像	7,423	
岡山 188 cm	SNG	分光	32,908	2 年
	OASIS	撮像	9,824	
木曾 シュミット	1kCCD	撮像	54,372	1 年
	2kCCD	撮像	12,669	

ワークも含めた計算機能力の急速な向上と、これら資源を購入・維持するためのコストの相対的な低下が、大規模アーカイブ開発と整備を実際に可能にした大きな要因である。

我々は 1994 年以降、国立天文台岡山天体物理観測所 188 cm 望遠鏡および東京大学木曾観測所 105 cm シュミット望遠鏡で取得された観測データのアーカイブシステム MOKA (Mitaka-Okayama-Kiso Archive system) を 3 世代にわたって構築・運用してきた²⁻⁵⁾。MOKA は、統一したユーザインターフェースから異なる望遠鏡および異なる観測装置のデータを検索できるようにした先進的なシステムであった。しかし、MOKA シリーズではデータ本体がオンラインでは参照できず、データ配布部分で人手を介する必要があるという大きな問題があった。また、MOKA3 開発当時の最先端技術であった Java アプレットを用いたクライアントサイド技術は、時代を経るにつれ、web ブラウザによる非互換性が次第に明らかになるなど、現在では技術的に時代の流れに合わなくなった部分も見られるようになった。さらに天文学の進歩に見合う高度な検索機能も要求されるようになってきている。そこで我々は、抜本的なシステムの再構築を行うべく、すばる望遠鏡の公開観測データを組み合わせた新しい公開データアーカイブシステムとして、Subaru-Mitaka-Okayama-Kiso Archive System (SMOKA) を開発することとした。

すばる望遠鏡⁶⁾ は、国立天文台がハワイ島マウナケア山頂で運用している口径 8.2 m の光学赤外線望遠鏡である。1999 年初頭のファーストライト以降、その高度な機能を十

分に発揮した装置群により、最先端の観測が数多く行われている（たとえば、“Subaru First Light Papers”⁷⁾ を参照）。すばる望遠鏡は北半球で運用されている口径 6～10 m クラスの望遠鏡の中で、もっとも広視野かつ高解像度の主焦点を備えており、その特性を生かした主焦点カメラにより、宇宙遠方の広領域を探索したり、太陽系外縁部に広がる小天体をサーベイするなどの研究が行われている。また近赤外線領域で雑音の原因となる、地球大気の発する明るい OH 夜光を取り除くことができる観測装置である OHS により、遠方の暗い銀河などの観測をより高い S/N で行うことができる。その他の観測装置もそれぞれ独自の特長を備えており、これらの観測装置から得られたデータを適切にアーカイブして一定の専有期間の後に広く一般に公開し研究者の再利用に供することは、前述したアーカイブの意義に照らしても極めて重要である。

SMOKA は、すばる望遠鏡、岡山天体物理観測所 188 cm 望遠鏡、および木曾観測所 105 cm シュミット望遠鏡で取得された天文観測データを、世界中の研究者に対して提供することを目的として 1998 年 1 月から開発が開始された（アーカイブの基礎となる FITS のキーワードの検討は 1997 年 4 月から）。MOKA シリーズの延長線上に立ち、さらに高度な天文学的検索機能と、より使い勝手の良い、また多機能な検索機能を追加した SMOKA は、2001 年 6 月からインターネットを通じて、非営利の研究あるいは教育普及の目的の利用者のためにデータを公開している。URL は <http://smoka.nao.ac.jp/> である。

本論文は、SMOKA の機能およびその特長について述べる。第 2 章では SMOKA の機能および実際の利用方法について概説する。第 3 章では SMOKA の構成と技術的特長について述べる。第 4 章では課題及び今後の開発の方向性について触れ、第 5 章でまとめを行う。

2. SMOKA アーカイブシステムの概要

2.1 SMOKA が公開しているデータ

SMOKA では、波長域としては可視から赤外、また観測手法として撮像および分光と、さまざまな種類のデータを公開している。2002 年 5 月 29 日現在で、SMOKA を通じて公開されている観測データの一覧および、観測者による専有期間は表 1 のとおりである。

現在、すばる望遠鏡の観測データはファーストライト期および試験観測期のデータが公開されている。第一期観測装置として運用されている 7 つの観測装置 (Suprime-Cam, FOCAS, HDS, OHS/CISCO, COMICS, CIAO, IRCS) と、2 つのファーストライト期試験観測装置 (CAC, MIRTOS) によって取得されたデータが対象である。すばる望遠鏡で 2000 年 12 月から始まった共同利用観測や、機能試験あるいは性能試験で取得されたデータは、原則的に表 1 にある専有期間が切れたデータから SMOKA で順次公開される。また、観測者などが研究教育への早期利用に鑑みて、観測者専有期間終了前に例外的に早期公開する場合もある。なお、アーカイブ構築に必須である FITS ヘッダ項目に不備があるデータは公開されていない場合がある。

岡山天体物理観測所 188 cm 望遠鏡の観測データは、SNG と OASIS の観測データが公開対象となっている。木曾観測所 105 cm シュミット望遠鏡の観測データは、1kCCD と 2kCCD の二つの観測装置の観測データが公開対象となっている。

図 1 は、各望遠鏡 (すばる、岡山 188 cm、木曾 105 cm シュミット) のそれぞれの観測装置ごとの公開データを天球面上にプロットしたものである。図 2 は、各望遠鏡 (すばる、岡山 188 cm、木曾 105 cm シュミット) のそれぞれの観測装置ごとの公開データを時系列にプロットしたものである。図 1、図 2 を見ると、どのようなデータがアーカイブ中に含まれているのかという状況を一目で確認することができる。これらは逐次更新されており、最新の図は web 上で参照することができる。

2.2 SMOKA の機能とユーザーインターフェース

図 3 は、SMOKA アーカイブシステム (<http://smoka.nao.ac.jp/>) にアクセスしたときのトップページである。アーカイブの概要、検索ページ、ユーザ登録ページ、オンラインヘルプ、各観測装置のデータの特徴などの情報、アーカイブ全体の状況、問い合わせ窓口などへのリンクと、最新情報などが示されている。

データ検索のモードには二種類を用意してある。単純検索 (Simple Search) と高度検索 (Advanced Search) である。

図 4 は、単純検索画面である。各観測装置ごとに観測された天体名 (FITS ヘッダの OBJECT の値) の一覧が表示されるので、目的の天体名を選択すれば検索できるようになっている。アーカイブは一般に研究者の興味のある天体全ての画像を含んでいるわけではない。そこで、利用者に対してアーカイブで利用可能なデータの一覧を示しておくことで、複雑な検索条件式を用いることなく簡単に、しかも必ずヒットする条件で検索できるようにしておくことには作業効率をあげる上で意味があると考えられる。このように天体名一覧から簡単に検索できる機能は、他のアーカイブシステムには見られない SMOKA 独自の特長である。

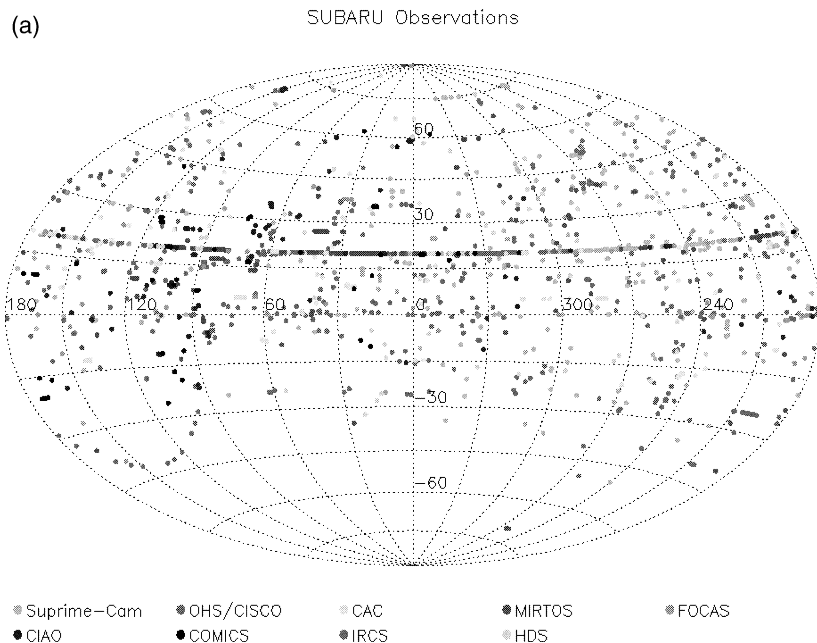
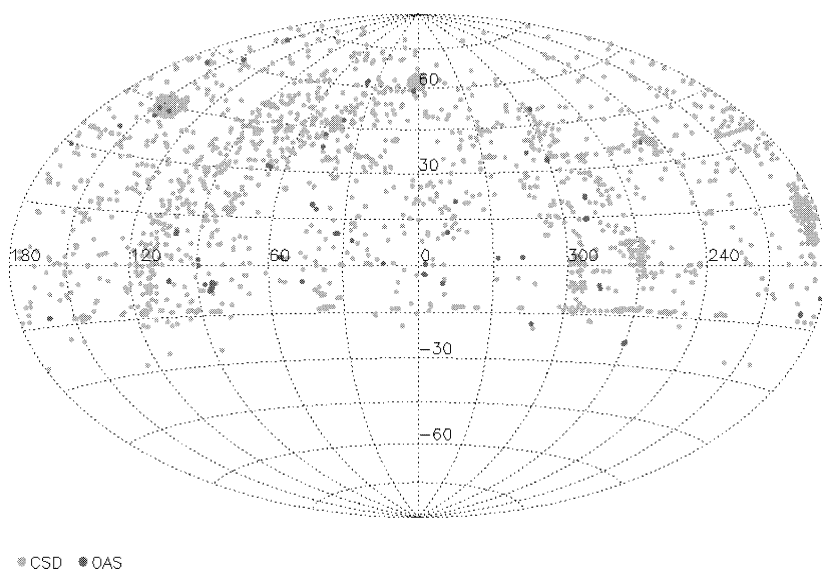


図 1

(b)

Okayama Observations



(c)

KISO Observations

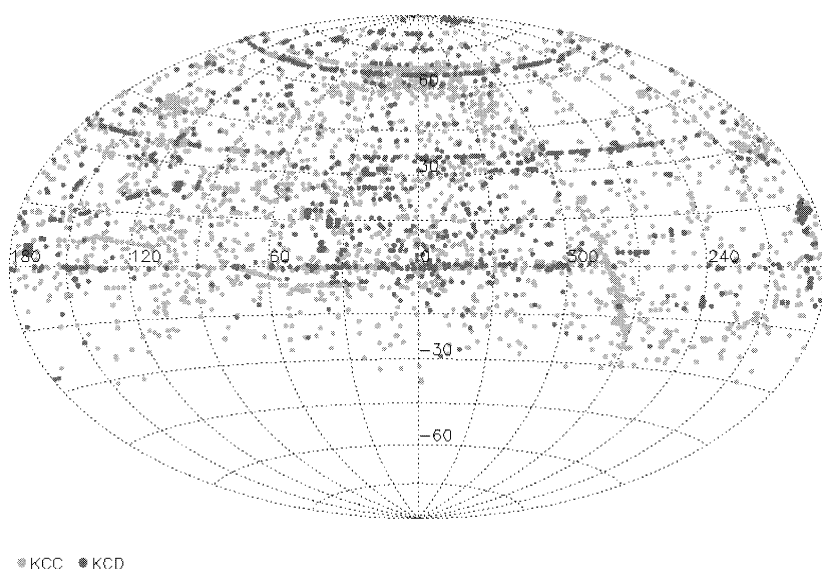


図 1. 各観測所・観測装置ごとの観測天域の天球上の分布図（較正データを除く）。図 1-a がすばる望遠鏡，図 1-b が岡山 188 cm 望遠鏡，図 1-c が木曽 105 cm シュミット望遠鏡に対応する。撮像・分光の区別はない。座標系は赤経・赤緯である。なお，図中の CSD は岡山 SNG を，OAS は岡山 OASIS を，KCC は木曽 1kCCD を，KCD は木曽 2kCCD をそれぞれ表している。

単純検索では観測時に入力された天体名を検索条件として用いていることから，入力時の誤りや天体名称の使い方（同一天体に付けられる複数名称のうちどれが記載されるかは観測者の任意）などにより，必ずしも目的の天体が漏れなく探索するとは限らない。そのため，FITS ヘッダ中の OBJECT に含まれる文字列で検索するだけでなく，ネームレゾルバを用いて，天体名から天球座標に一度変換してから座標範囲で検索する機能も提供している。ネームレゾルバとしては，現在 SIMBAD (<http://cdsweb.u-strasbg.fr/Simbad.html>) および NED (<http://nedwww.ipac.caltech.edu/>) が利用できる。

図 5 は高度検索画面である。高度検索は，天体名だけでなく様々な条件を同時に指定して検索する利用者のためのインターフェースである。この検索画面には複数の検索条件入力フィールドがあり，天体名の他に，天球座標（赤経・赤緯/銀経・銀緯/黄経・黄緯）での指定が可能である。座標指定では，たとえば中心座標の回りに 10 分角以内などと範囲を指定できる。また観測日時，観測者名，フレーム番号，観測装置，データ種別などの条件を設定して検索できる。結果出力する項目も選択可能になっている。高度検索では，複数の望遠鏡・観測装置によるデータを同時に検索でき，多波長の様々な観測モードのデータの横断的検索が可能になっていることが特長である。なお，この画面は

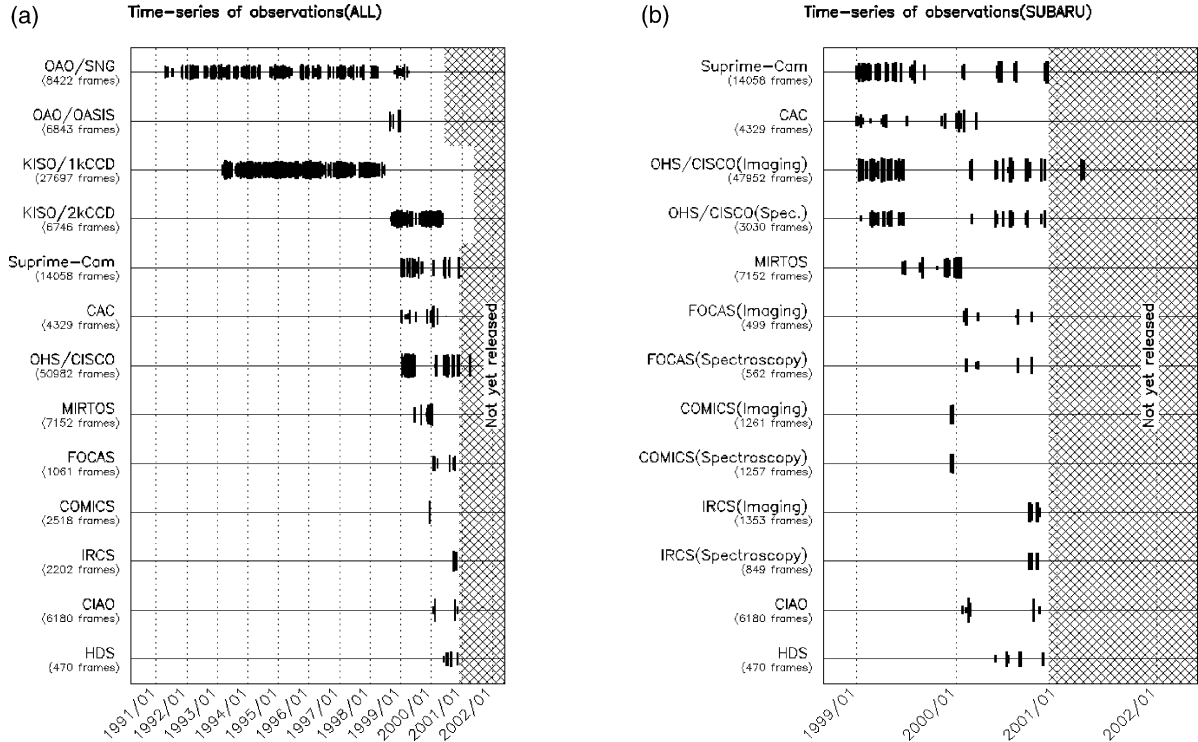


図 2. SMOKA 全観測装置のデータ取得時刻の時系列プロット (1991 年以降) (a), および, すばる望遠鏡観測装置のみの拡大プロット (1999 年以降) (b). 時系列プロットは, オブジェクトフレームのみをプロットしており, それぞれの観測装置の線の縦幅がフレーム数を表す. 網掛け部分は観測者専有期間中であるため非公開とされている期間に対応する. 網掛け部分の中にある一部のデータは, 早期公開データを示している.

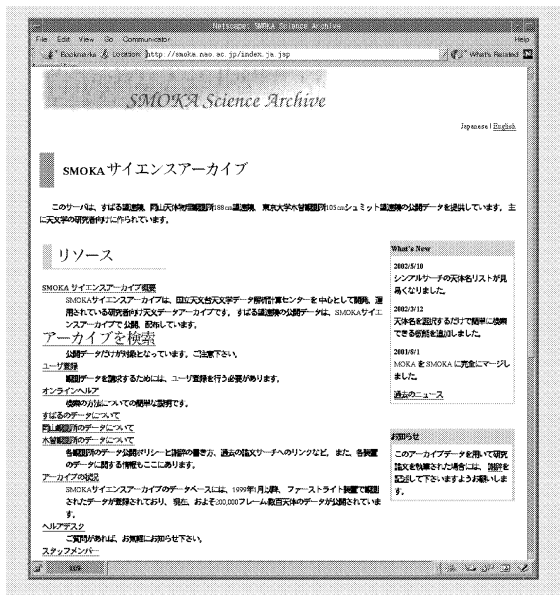


図 3. SMOKA アーカイブシステムのトップページ.

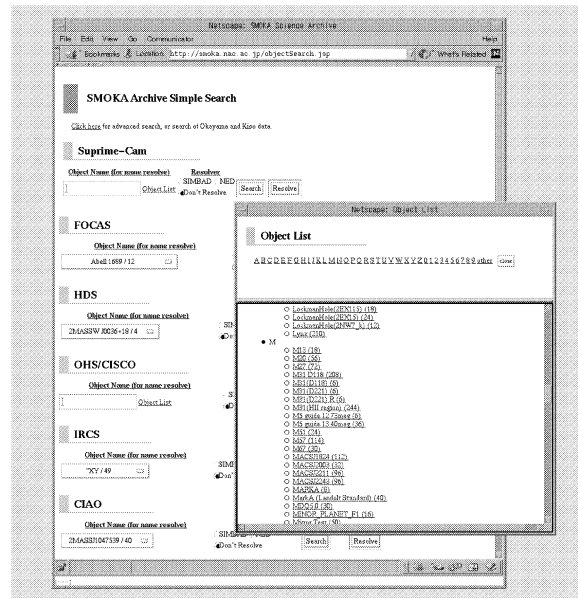


図 4. 単純検索画面. 観測装置ごとに天体名とフレーム数がスラッシュで区切られて表示されている. 利用者は観測装置を選び天体名を選択することで, 観測データを簡単に探すことができる.

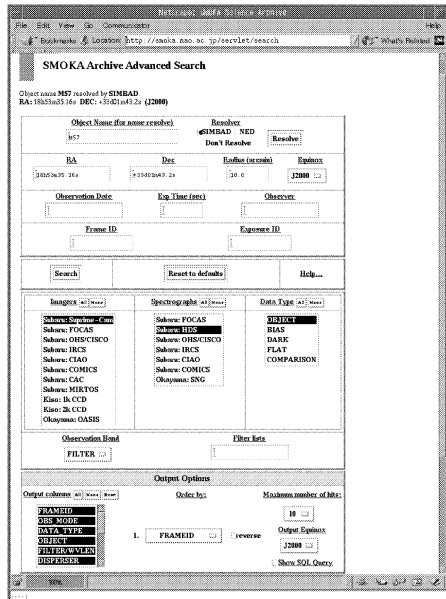


図 5. 高度検索画面. 天体名, 天球座標 (赤経・赤緯/銀経・銀緯/黄経・黄緯), 観測日時, 観測者名, フレーム番号, 観測装置, データ種別, 出力項目などの条件を設定して検索できる。

MOKA3 あるいは HST や VLT アーカイブなどの既存システムと機能的に類似させてある。ユーザはこのインターフェースになじめば、他のアーカイブシステムも同様の操作方法ですぐに利用できるようになるだろう。

図 4 あるいは図 5 の検索画面で条件を入力し Search ボタンを押すと、検索条件に適合するフレームの一覧が表示される。図 6 は検索結果画面の一例である。検索結果画面では、まず検索結果の概要 (観測装置ごと, 撮像/分光モードごとのそれぞれの枚数) が表示され、簡単な説明文の後

に検索結果一覧が表示される。一覧で表示される項目は検索画面で取捨選択できる。

図 6 の検索結果フレーム一覧からは、必要に応じて FITS ヘッダや早見画像 (QLI: Quick Look Image), そのフレームに関する情報が記載されたメモなどを閲覧することができる (図 7, 8)。この QLI 表示システムは、観測データを縮小した画像の他, 統計情報やスペクトルトレースなどを表示することができる。複数チップを用いた観測装置の場合, モザイク化した表示も可能である。ユーザは, これらの情報から観測データの質や天文学的情報を評価し, 検索結果一覧の中から必要なものを選別することができる。

図 6 の検索結果フレーム一覧から必要なデータを選択して Datarequest ボタンを押すと、データの請求ができる。図 9 にデータ請求画面を示す。確認のために, 請求したデータの概要とファイルサイズ合計が表示されている。この画面で OK ボタンを押すと, 利用者登録時 (後述) に登録したアドレスに電子メールで請求の受理が通知され, 数分~1 時間の後にデータが取得可能になった旨が電子メールで通知される。データ転送の手段は FTP を標準としているが, テープ媒体による郵送を依頼することも可能であり, 請求時に選択できる。

なお, SMOKA にアーカイブされた観測データを取得するためには, 簡単な利用者登録を行う必要がある。図 10 は利用者登録画面である。名前やメールアドレス, 所属機関, 利用目的などの情報を入力するようになっている。これは, データ請求者を把握し, 機械的なアクセス等による不必要なデータ請求を抑止するためのものであって, 利用者に制限を課すためのものではなく, 非営利の研究あるいは教育目的であれば, 誰でも登録可能である。利用者 ID の有効期限は当該年度内としており, 年度末には登録ユーザに対して更新に関する連絡を通知している。

No.	FRAMED	OBS MODE	DATA TYPE	OBJECT	FILTER/WHEEL	DISPENSER	RAHMS	DECIMS	DATE/ORS	UT START	EXP TIME	OBSERVER	EXP. ID	Note
1	HD540000341	SPRC	OBJECT	M57	637.0 - 739.0	schell	18:53:35.61	+33:01:38.00	2000-07-06	18:59:56.71	1:00:00	Aoki et al.	18N240WN	SIMBAD HST MASON
2	HD540000341	SPRC	OBJECT	M57	637.0 - 739.0	schell	18:53:35.61	+33:01:38.00	2000-07-06	18:59:56.71	1:00:00	Aoki et al.	18N240WN	SIMBAD HST MASON
3	HD540000341	SPRC	OBJECT	M57	637.0 - 739.0	schell	18:53:34.909	+33:02:03.32	2000-07-06	11:10:45.536	901.00	Aoki et al.	18N240WN	SIMBAD HST MASON
4	HD540000341	SPRC	OBJECT	M57	637.0 - 739.0	schell	18:53:34.909	+33:02:03.32	2000-07-06	11:10:45.536	901.00	Aoki et al.	18N240WN	SIMBAD HST MASON
5	HD540000653	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON
6	HD540000658	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON
7	HD540000657	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON
8	HD540000658	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON
9	HD540000659	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON
10	HD540000650	Imaging	OBJECT	Play/Melior (M57)	N-A-1651		18:53:00.172	+33:01:02.24	1999-04-21	15:04:03.117	1:00:00	Komatsu	SUPER000653	SIMBAD HST MASON

図 6. 検索結果画面. M57 を検索した結果. 検索結果の概要 (観測装置ごと, 撮像/分光モードごとのそれぞれの枚数) が表示され, 簡単な説明文の後に検索結果フレーム一覧が表示される。

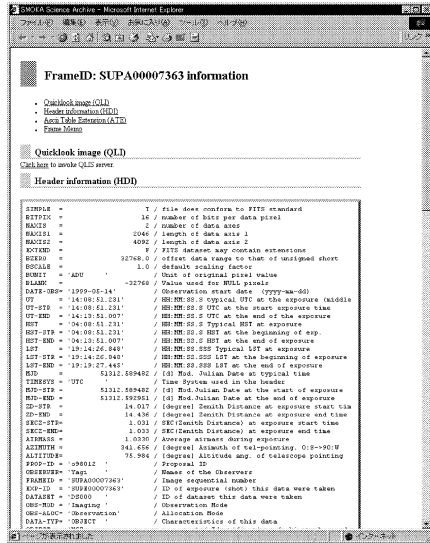


図 7. FITS ヘッダ (HDI) 表示画面.

3. SMOKA アーカイブシステムの開発

3.1 STARS/MASTARS と SMOKA の関係

すばる望遠鏡には, STARS (Subaru Telescope ARchive System^{8,9)}) という, 望遠鏡システムの一部として開発・運用されているアーカイブシステムがある. これはハワイ観測所内に設けられており, すばる望遠鏡で観測を行った観測者グループが各自の観測データを取得するために利用されている. また国内利用者のため, 国立天文台三鷹キャンパスに設置された光赤外データ解析システムに MASTARS (Mitaka Advanced STARS) と呼ばれる STARS のミラーセットが設置されている.

STARS および MASTARS は観測者グループ用アーカイブであり, 専有期間内にあるデータのセキュリティを守る機能を重視していることから, 観測者ではない一般研究者が制限なしに利用することはできない. また, STARS は観測システムの一部として稼働しており, 一般利用者向けに一部分だけを改造することは容易ではない. 加えて, STARS/MASTARS は岡山・木曽のデータを組み込んで公開するような用途は想定しておらず, 現在運用中のシステムでもあることから高度な検索機能などの先進的な実験的機構を組み入れるのが難しい面がある. そこで, STARS および MASTARS に蓄積されているすばる望遠鏡の観測データのうち, 観測者の専有期間を過ぎて公開可能になったデータだけを SMOKA 側にコピーして一般の研究者に公開するという方針をとっている.

図 11 は, STARS/MASTARS と SMOKA との関係を表した図である. マウナケア山頂のすばる望遠鏡で得られた観測データは STARS に自動的に蓄積され, データベースに登録される. STARS から MASTARS へも専用のネットワーク回線を通じてデータ転送が行われ, データベースへの登録も行われる. まず MASTARS のデータベーステーブルのうち, 2.1 節で述べた公開条件を満たすフレームを選んでコピーしている (この時点で MASTARS と SMOKA で利用しているデータベース管理システムの違いを吸収している). そのコピーされたデータベーステーブルの中から検索に必要な項目だけを抜き出し, SMOKA の高度な検索機能に必要な項目を付加したテーブルと, データ本体や QLI などへのパス名やファイルサイズなどの管理情報を含むテーブルを作成する. そして, 対応するデータ本体を MASTARS から SMOKA の磁気テープライブラリにコピーする. ヘッダ情報 (Header Information: HDI) や QLI は SMOKA 上でデータ本体から再作成している.

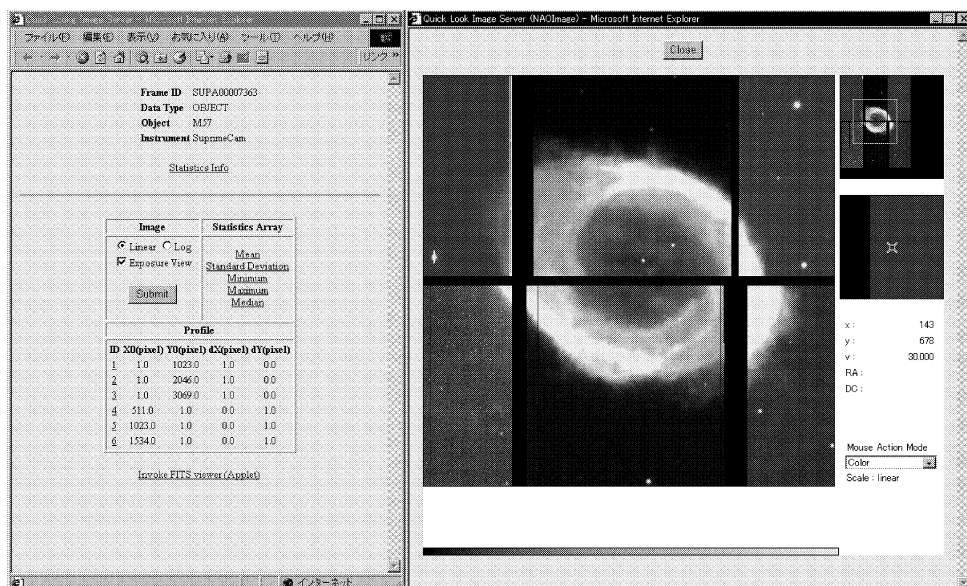


図 8. 早見画像 (QLI) 表示画面. QLIS が起動している. 複数の検出器を持つ観測装置の場合, モザイクして表示することも可能である.

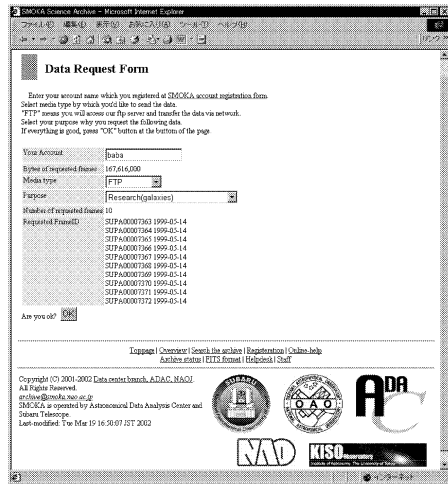


図 9. データ請求画面.

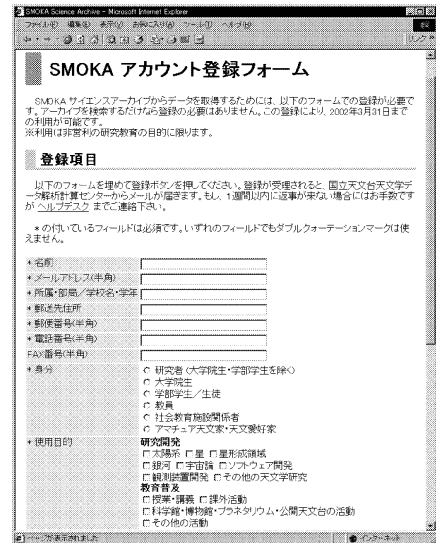


図 10. 利用者登録画面.

3.2 システムの構造

SMOKA に用いたハードウェアおよびソフトウェアの詳細は表 2 にまとめてある。運用環境の Solaris 7 に加えて、複数の開発者の環境に合わせて Linux や Windows 2000 なども併用した。Java には異なる OS 環境であっても同じソースコードのまま動作するという特徴 (Write once, Run anywhere) があり、これによって開発効率を上げることができた。

図 12 は、SMOKA の内部の処理とデータの流れを示している。SMOKA は、標準的な三層モデルを採用している。三層とはすなわち、ユーザインターフェース (UI) 層、中間 (Middle) 層、データベース (DB) 層である。

(1) ユーザインターフェース層

ユーザインターフェース層は、利用者が観測データにアクセスするためのインターフェースを提供する。図では一番左に位置するアーカイブシステム利用者は、利用者のコンピュータ上の Web ブラウザを通じてユーザインターフェース層にアクセスする。検索条件入力画面 (図 4, 5)、検索結果表示画面 (図 6)、データ評価画面 (図 7, 8)、データ請求画面 (図 9)、利用者登録画面 (図 10)、がこの層に該当する。ユーザインターフェース層は、主として web サーバ機能と JSP (Java Server Pages) 機能によって実現されている。JSP は HTML ファイルの中に Java 言語を埋め込むもので、動的なコンテンツを作成できる。加えて、web クライ

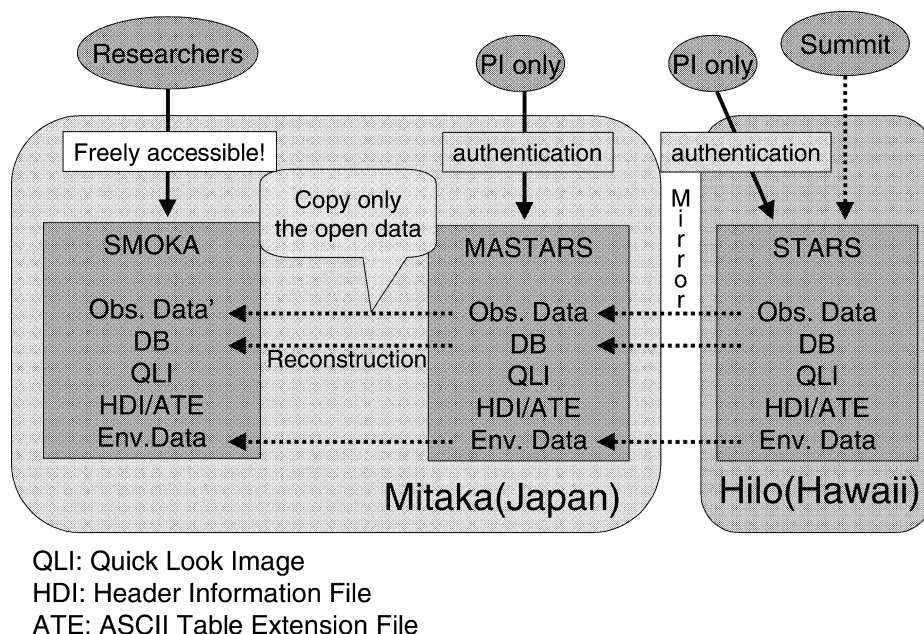


図 11. すばる望遠鏡で観測したデータの流れ。SMOKA と STARS, MASTARS との関係を示す。それぞれのシステムの中には、観測データ (FITS ファイル) の他に、データベース (DB)、早見画像 (QLI)、FITS ヘッダ情報 (HDI)、環境データなどが存在していることを示している。

表 2. SMOKA 運用サーバ計算機に用いたハードウェアおよびソフトウェア
ハードウェア

構成	利用した機器	
データベースサーバ および Web サーバ	富士通	GP-7000S-M450 (CPU: Ultra SPARC II 400 MHz×4, メモリ容量: 4 GB, ハードディスク容量: 2 TB)
磁気テープライブラリ	SONY	PetaSite (データ容量: 60 TB)

ソフトウェア

構成	利用したソフトウェア
OS	Solaris 7
Java コンパイラ	JDK-1.4.0
Web サーバソフトウェア	Apache-1.3.22
Java サブレットコンテナおよび JSP インプリメンテーション	Jakarta-Tomcat-3.2.4
データベースマネジメントシステム	Sybase Adaptive Server Enterprise 12.0 と Oracle 8i

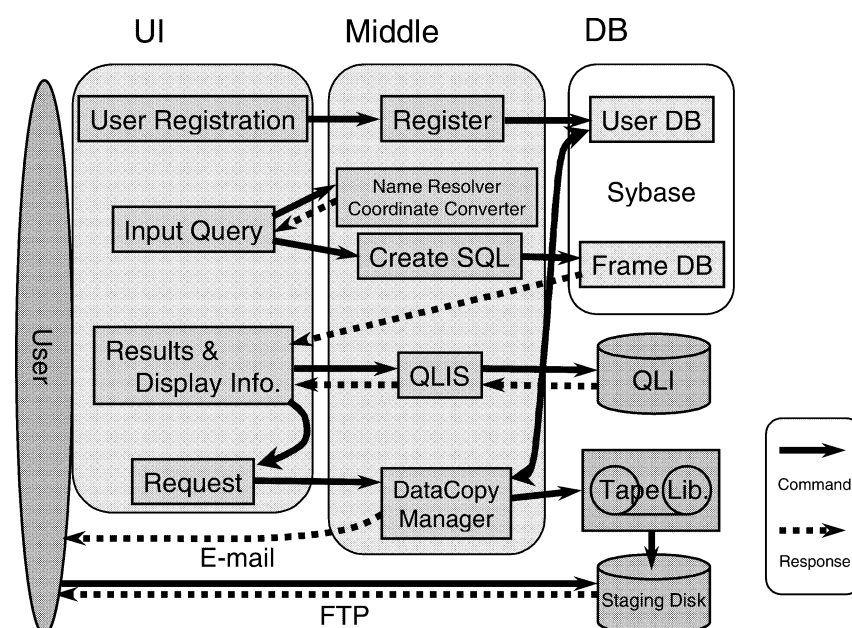


図 12. SMOKA の内部構造図。標準的な三層モデル、ユーザインタフェース (UI) 層、中間 (Middle) 層、データベース (DB) 層の三つと、コマンドと応答の流れを示している。コマンド・応答の流れは、2 章で説明したデータ取得の流れにしたがっている。

アントサイドで動作する JavaScript も部分的に用いている。サーバ機能には Apache (<http://httpd.apache.org/>) と Jakarta-Tomcat (<http://jakarta.apache.org/tomcat/>) を連携させて用いている。

(2) 中間層

中間層は、まず、ユーザインタフェース層から受けた検索要求を、データベース検索言語である SQL に変換してデータベース層（後述）のデータベースに投げ、検索結果

を HTML 形式に整形してユーザインタフェースに返すという機能を持つ。データベースとの通信は JDBC (Java DataBase Connectivity) を用い、ユーザインタフェース層との通信は Java サブレットと JSP を連携させることによって実現されている。

データ請求に関しては、磁気テープライブラリ (PetaSite) に保管されたデータをステージングディスクにコピーし、準備ができ次第、電子メールで利用者に通知するという機能を持つ。この部分は Fortran とシェルスクリプトで書かれ

たプログラム (Data Copy Manager) によって実現されている。天体名を天球座標に変換するネームレゾルバの呼び出しと座標系の変換 (赤経・赤緯から銀経・銀緯など、あるいは分点変換) も中間層で行っている。

(3) データベース層

データベース層には、登録ユーザデータベースと観測データのデータベースが存在する。後者はデータ登録用、データ検索性、データ管理用など複数のテーブルを持ち、目的に応じてそれぞれを使い分けしている。これらはリレーショナルデータベース管理システムである Sybase Adaptive Server Enterprise (<http://www.sybase.co.jp/product/ase/>) を用いて管理されている。

登録ユーザデータベースは利用者登録画面で入力された情報を管理しており、データ請求時の受付、データコピーの完了などを通知したり、テープメディアでのデータ請求時の送付先ラベルを自動生成するために用いられる。データ登録用テーブルは MASTARS のデータベーステーブルを複写したものであり、各フレームの公開開始日の管理に使われるとともに、内容を取捨選択、追加し、検索性テーブルに再構築している。データ検索性テーブルでは観測データの検索に必要な観測座標、観測日時、露出時間、観測者、フィルターなどの情報が観測装置ごとに管理されている。データ管理用テーブルではデータが正しく登録されているかを示すフラグやデータ本体、QLI の保管されているパス名などが管理され、請求されたデータをステージングディスクにコピーするのに矛盾がないようにしている。

QLI ファイルとデータファイル本体はそのものが直接データベース化されているわけではなく、データ管理用データベースを通して OS のファイルシステム上に保管されている。QLI ファイルは高速に参照できるようにハードディスク上に置かれ、観測データファイル (FITS ファイル) の本体は大規模磁気テープライブラリ装置 (PetaSite) 上に保管されている。

3.3 システムの機能的特長

(1) FITS ヘッダの標準化

観測データである FITS ファイルのヘッダ部分には、観測データがいつどこでどのような条件のもとで得られたものであるかが記載されている。SMOKA は FITS ヘッダ項目をデータベースに展開し、それをキーにして検索を行っているが、観測装置ごとに異なった FITS ヘッダ記述法が採用されていると、別々の観測データを統一的に扱うことが非常に困難となる。同じ観測装置でありながら時期によってヘッダの記述方法が変化している場合や記載情報が不十分だったり誤っている場合も同様である。一般に、ひとつの天体を様々な角度から眺めることがその天体の本質のより深い理解につながる。そのためには、別々の観測データを統一的に眺めることができるような枠組みは必要不可欠であり、したがってヘッダ項目を正しくかつ統一的に記述することは、有用なアーカイブ構築の必須条件となる。

SMOKA ではすばる望遠鏡観測データのデータフォーマット、特に FITS ヘッダの統一に力を注ぎ、1997 年 4 月からすばるソフトウェア開発チーム、観測装置開発グループおよび望遠鏡制御グループとの検討を始め、同年 12 月に「FITS ヘッダのすばる規約」をまとめた¹²⁾。各観測装置の設計開始とほぼ同時期にアーカイブシステムの構築を視野に入れてデータの記述形式についての検討を重ねた甲斐あって、各観測装置 (岡山天体物理観測所、木曽観測所の観測装置も含む) のヘッダ情報の共通化を行うことができたので、波長横断的・観測装置横断的なデータベース検索を行うシステムを比較的容易に構築することができた。

(2) アーカイブ状況表示

利用者にとってアーカイブにどのようなデータが含まれているのかは、もっとも重要な情報の一つである。図 1、図 2 は SMOKA のアーカイブ状況を表す図であり、これらは web 経由で利用者が随時参照できるようになっている。特に時系列プロット (図 2) は他のアーカイブには類を見ない SMOKA 独自のものである。

(3) 校正データ検索機能

あるオブジェクトフレームを選択するとそのフレームに対応する適切な校正データを観測装置ごとの条件で決める校正データ検索機能を組み込んだ。たとえば CAC の場合、あるオブジェクトフレームに対応する校正データとして

- ・観測時刻前後 1 時間以内のバイアス
- ・観測日と同じ日に撮られたダーク
- ・観測日と同じ日に撮られ、同じフィルターのフラット
- ・オブジェクトの高度とプラスマイナス 20 度以内の高度で観測時刻前後 3 時間以内に同じフィルタで撮られた標準星データ

が選択されるようにしている。選ばれた校正データは、データ請求時にダウンロードすることもできるようになっている。

(4) メモ機能

公開データの中には、観測時のトラブルやデータの不備等の問題を持っているものが存在する。メモ機能は、SMOKA へのデータ登録時に明らかになったそのような問題をアーカイブユーザーに示すためのものである。メモは、シェルプログラムによって生成され、QLI や HDI ファイルなどと同様にバックエンドのハードディスク上に蓄積されており、検索結果画面から参照することができる。

3.4 システムの技術的特長

(1) サーバサイド処理の採用 (Java サブレットと JSP)

SMOKA の開発にあたっては、MOKA シリーズの開発で培われた基本構造を継承しつつ、ユーザインターフェース部分に大幅な改良を加えている。表 3 に、過去の MOKA シリーズで用いたフロントエンド技術をまとめた。MOKA3 と SMOKA は、Java 言語を用いてシステムを構築した

表 3. MOKA シリーズで用いた技術

稼動時期	フロントエンドに用いた技術
MOKA1 1995.6–1996.9	X-base with Motif client
MOKA2 1996.9–2001.2	Web-base with FORM/CGI
MOKA3 1998.2–2001.7	Web-base with Java Applet
SMOKA 2001.6–	Web-base with Java Servlet

という点では同じである。しかし MOKA3 ではクライアント側の処理を重視して Java アプレットを用いたのに対し、SMOKA ではサーバ側の処理を重視し、Java サーブレットと JSP を用いている。これは、Java アプレットが MOKA3 開発時に期待した性能をあげておらず、利用者側の web クライアントにかなりの負荷をかけたためである。また、一般に広く用いられている web クライアントである Microsoft Internet Explorer と Netscape Communicator との間に非互換があり、Java アプレット開発にかかる労力も大きくなっていった。サーバ側で処理することにした結果良好な反応速度を得ることができており、web クライアントの非互換の問題も原理的に発生しない。

また、Java サーブレットと JSP を採用したことにより、結果表示部と内部アルゴリズム実装部（特にデータベース問い合わせ部）を分離できたという点も特筆できる。JSP とサーブレットという枠組みによって自然に表示部分とロジック部分を分離するので三層モデルをうまく実装することができ、アプリケーションの品質と生産性を向上させることができた。また、実装の分離は少人数による効率的な開発を行う上で重要な役割を果たした。

(2) セッション管理の採用

SMOKA では、ユーザの与えた検索条件に多数のフレームがヒットした場合、一定数（デフォルトでは 20 件）ずつを順番に表示するようにしているが、この機能の実現のために Java サーブレットが標準で備えるセッション管理（Appendix を参照）を活用した。これまでのステートレス（Appendix を参照）なシステムの場合には、仮に 1000 件ヒットした場合、1 件目から 20 件目だけを表示し、そのほかの部分は表示せず捨てていた。したがって、次の 21 件目から 40 件目を表示したい場合、データベースにまた同じ検索要求を出して結果を得る必要があった。セッション管理によっていったん検索した情報を再利用できるようになり、ステートレスなシステムと比べて、検索結果表示時の反応速度を大いに向上することができた。

(3) 全データのオンライン化の実現

SMOKA は、前システムである MOKA3 と異なり、すべてのデータをオンラインで参照可能にしたため、データの請求があるたびに人手で磁気テープからデータを磁気ディスクにコピーする必要がなくなった。そのため、データが入手可能となるまでの時間も大幅に短縮されている。この

目的のために、磁気テープライブラリ全体を仮想磁気ディスクとして扱える PetaSite を採用した。

(4) データベーステーブルの活用

データセキュリティとエンジニアリング（観測装置の状況把握など）を重視した MASTARS のテーブル構造は、そのままでは SMOKA には適さない。そこで、MASTARS から SMOKA にデータベーステーブルを移行する際には、MASTARS で用いている Oracle から SMOKA で用いている Sybase に変換すると同時に、高度検索機能のための項目を追加し、データベーステーブル項目の再構築を行い、検索の高速化を図っている。また、MOKA3 までは観測装置ごとの検索用テーブルだけを使っていたが、SMOKA では全データのオンライン化と、データのチェック機能を追加するために、3 種類のテーブル（データ登録用、データ検索用、データ管理用）を使用するようにデータベースを再構成している。

(5) 早見画像表示システムの改良

SMOKA の QLI 表示機能 (QLIS^{10,11}) は他のデータアーカイブにはない高度なものである。SMOKA の QLI は、縮小した画像の他、ピクセルの統計量やスペクトルトレースなど、観測データの評価に必要な情報を付加した FITS データであり、あらかじめ観測データから作成してハードディスク上に保存してある。QLI 表示要求があると、QLI 表示システムは QLI から必要な情報を取り出して PNG 画像に変換し、web クライアント上に表示する。また、画像表示アプレットの NAOimage を起動して QLI を直接表示することもできる。いったん NAOimage と QLI が web クライアントにロードされると、ユーザはサーバとの通信をそれ以上行うことなく、動的に画像を扱うことが可能になる。ユーザは表示されるさまざまな情報によってそれぞれの研究目的に合ったデータを検索結果一覧の中から選別することができる。SMOKA 用の Q LIS は、STARS 用に開発した Q LIS を改良したもので、SMOKA 本体と同様 Servlet と JSP で実装されている。PNG 画像の作成には Java Advanced Imaging (<http://java.sun.com/products/java-media/jai/>) および Pure Java AWT (<http://www.eteks.com/pja/en/>) を用いている。

4. 今後の課題

2001 年 6 月から運用を開始した SMOKA であるが、天文学の研究や教育目的の利用にとってさらに有用なものとするためには、いくつかの課題が残っている。以下ではそれぞれの課題と展望について論じる。

4.1 整約済データの提供

現在の SMOKA は、いっさいの解析処理を施していないデータ（いわゆる生データ）を利用者に提供しており、アーカイブシステムの利用者は自身でデータの整約を行う必要がある。観測装置によっては、装置特有の複雑な整約手順を取らなければならない場合もあるが、アーカイブシステムの利用者は一般に、そのような観測装置固有の最適なデータ整約方法に通暁しているわけではない。また、目的の天体観測データに対する最適な較正データを選択するための情報も持ち合わせているとは限らない。そのため、観測データの整約に慣れない研究者にとっては、処理の煩雑さのために天文学に結び付けるまでに相当の労力と時間を要することになってしまう。このような整約作業をアーカイブシステム側で行って提供するようにすれば、アーカイブ利用者はすぐに天文学的な解析を始めることができる。これは、整約者の技量などの個人差に左右されない一定の品質が保証された整約済データを提供できるということも意味する。天文観測衛星のデータアーカイブは整約済データを提供していることが多く（HST の場合は <http://archive.stsci.edu/> など）、理論の研究者などもこれらのデータアーカイブを利用して研究成果をあげている。整約済データを提供することで、データアーカイブからより多くの研究成果が得られるのである。

しかし、SMOKA のような地上観測装置のデータアーカイブで整約済データを提供するためには、いくつかの困難がある。大気の状態や天候など、観測時の環境によって最適な整約方法が変わる場合がある。また、観測装置に改良が加わることがあるため、それによっても整約方法が変わりうる。整約処理を自動的に行うパイプラインを構築するためには、このような変化にも対応可能な整約手順の確立と最適化の検討・実装を進めることが必要である。同時に、整約済データの品質を評価できるようにするために、地上観測データの質を判断するのに大きな指標となる PSF (Point Spread Function) をフレームごとに求めてデータベース化し、検索の対象とできるようにすることも検討している。

4.2 観測データの較正・カタログ化

現在の SMOKA においては、データを較正・解析して研究成果を出すのは完全に利用者によっている。較正にはフラットフィールドニングのように、観測装置固有の特性を取り除く整約過程は当然含まれるが、さらに位置較正やフラックス較正なども含まれる。これらを行うためには位置および明るさの標準天体の情報を観測データに結び付ける必要がある。我々は、観測フレームに写っている天体と

USNO-A2.0 などの大規模なカタログとの照合を行って位置較正、フラックス較正を行うことを検討している。さらに、アーカイブに含まれるデータから天体とその物理量を抽出してカタログ化して利用者に公開するなど、その一部をデータアーカイブシステム側で行うことができれば、より多くの研究成果が得られるようになると期待される。

4.3 環境データ等の提供

地上観測においては観測時の環境は刻々と変化する。すばる望遠鏡では、観測時環境の記録として天候、温度、湿度などのステータスログと全天の雲モニター画像を STARS および MASTARS に保存している。このような環境データは、より高度な解析を行うためには必須である。SMOKA においてこれらを一般研究者に供するためには、観測データと環境データの関連付けを行い、必要な情報を抽出して提供する機構が必要となる。また、研究に即した情報を効果的に抽出できるようなインターフェースの検討・実装も必要である。

4.4 検索機能の強化

現在、SMOKA では単純検索と高度検索の二つを提供しているが、アーカイブシステム利用者の様々な天文学的要求に対応すべく、さらなる検索の高度化を目指している。

- ・特定の観測領域を指定せず複数回の観測が行われている観測領域のデータ検索。移動天体や変光天体の検出に有効である。
- ・観測波長による検索。たとえば装置の正式なフィルター名を知らなくても観測波長を指定することで検索が可能になる。
- ・天球面上のプロットから、グラフィカルに領域を選択して検索できる機能。グラフィカルなインターフェースを用いて、ユーザにとってより利用しやすい検索方法を提供したい。
- ・既存の天体カタログと連携した検索。たとえば楕円銀河の写っている撮像データを探し出す、など。

これらの検索アルゴリズムの検討と実装が必要である。

4.5 早見画像の充実

2002 年 5 月現在、早見画像 (QLI) は Suprime-Cam, HDS, CAC, SNG, 1kCCD, 2kCCD, OASIS の画像について作成されており、すべての観測装置に対して作成されているわけではない。可視の撮像・分光データについては順次対応が進められているが、赤外の観測データで QLI が作成されているのは岡山観測所の OASIS の撮像データのみである。これは可視データからは 1 枚の生データからでも十分な品質の QLI を作成できるが、赤外データで十分な S/N の画像を作成するためには、多数の観測データおよびそれに対応する較正データが必要だからである。しかし、1 枚の QLI を作成するために必要なデータ群を選び出す方法は一意的に決まるものではないため、QLI を作成す

るプロセスを自動化するのが難しい。現在この問題への対応を含め、まず OHS/CISCO などすばる望遠鏡の赤外撮像データの QLI 作成から始め、ついで赤外分光データの QLI 作成に進むことを検討している。

4.6 新規観測装置の組み込み

SMOKA に対応する観測装置を新規に追加することも重要である。現時点で具体的に追加を想定している観測装置としては、岡山の HIDES (HIgh Dispersion Echelle Spectrograph) がある。すばる第二期観測装置も将来の組み込み対象として視野に入れている。他にも有用な観測データがあれば、可能な限り SMOKA に組み込みたいと考えている。

前述したように、新規の観測装置の組み込みを行う際に問題になるのは、データのフォーマット、特に FITS ヘッダの規約である。よく用いられる FITS キーワードを観測装置間で可能な限り共通化し、標準化することが重要である。これは効率的なデータ解析のためのみならず、いくつかの観測装置のデータを一括して検索できるようにするために必須の条件である。このためには、観測装置グループや観測所との連携・協力が欠かせない。

4.7 データの追加の自動化

2002 年 5 月現在、SMOKA へのデータ登録には手動で行わざるを得ない部分が残っている。これは予期できないデータの不備をチェックし、しかるべき対応を行うためである。これを完全になくすことは実際上不可能である。また、観測者からの申し出によって専有期間満了以前に早期に公開されるデータもある。これらへの対応にも手動による部分が残されている。しかしながら、データアーカイブを効率的に運用していくために省力化は不可欠であり、不完全なデータへの対応をできるだけ自動的に行えるように変更を加えるなどして、できる限りデータ登録の省力化を図りたいと考えている。

4.8 jMAISON との連携

jMAISON (<http://maison.isas.ac.jp/>, <http://maison.nao.ac.jp/>¹³⁾) とは、web 上の複数のリモート天文データ (画像、カタログ) サーバとユーザとの間を仲介する、多波長天文画像データ検索・閲覧システムである。jMAISON のリモートサーバの一つとして SMOKA を追加すれば、DSS (<http://dss.nao.ac.jp/>) などの可視サーバイデータや、X 線や電波などでのサーバイデータとの重ね合わせなどを簡単に行えるようになり、新たな研究を切り拓くことができる。jMAISON との連携を図るためには、jMAISON 用のインターフェースに適合するように SMOKA を改良する必要がある。

5. まとめ

我々は、MOKA システムの後継として、すばる望遠鏡公開データアーカイブシステム SMOKA を開発し、運用を

行っている。SMOKA は単純検索と高度検索の二つの検索インターフェースを備え、アーカイブ利用の初心者でも利用しやすいシステムと、高度な検索機能を必要とする利用者の要求にも応えうるシステムを同時に実現した。アーカイブ状況が web を通じ利用者に一目でわかるようにする仕組みも提供している。また、較正データ検索やメモ機能などの新規機能の追加、QLI 表示機能の改良を行なって、検索がより有効に機能するように工夫した。SMOKA はデータをオンライン参照可能にすることで、MOKA に比べてユーザがデータを入手するまでの時間を大幅に短縮した。また、FITS ヘッダの標準化を通じて、多くの異なる望遠鏡・観測装置のデータを統一的に扱うことを可能にした。すばる望遠鏡システムアーカイブである STARS/MASTARS との連携を行ないつつ、データベース構造の大幅な見直しを行い、より高度かつ効率的な検索を可能にしている。システム内部構造としては、標準的な三層モデルを採用した。特に ユーザインターフェース層ではサーバサイド処理を重視した JSP を用いて、開発の効率化と応答の高速化を行なった。節約済データ・較正済カタログの提供、検索機能の強化、他アーカイブとの連携、運用の効率化などが今後の課題である。

謝 辞

本研究は、天文情報処理研究会のワーキンググループ活動の一環として行われた。実際の開発作業にあたっては、天文学データ解析計算センター DB/DA プロジェクトの計算機資源を活用した。国立天文台の西村史朗氏、天文学データ解析計算センターの大越克也氏、野田祥代氏、古荘玲子氏、総合研究大学院大学の木下大輔氏、すばる望遠鏡観測データの FITS 規約検討作業を共に進めて下さったすばる望遠鏡ソフトウェア開発チームの方々、国立天文台ハワイ観測所の小杉城治氏、および福岡教育大学の金光理氏などの方々にも感謝する。また、各観測装置の詳細な情報を提供下さった観測装置開発グループの担当者の方々 (青木和光、岩室史英、岡本美子、友野大悟、本原顕太郎、小宮山裕、村川幸史、大山陽一、Mark Weber、小林尚人、寺田宏、の各氏) に感謝する。また匿名の査読者には貴重なコメントをいただいた。本アーカイブ運用には、CDS によって運用されている SIMBAD データベースおよび JPL によって運用されている NED データベースを利用した。有用な各種フリーソフトウェアを提供している開発者にも感謝したい。本研究は、国立天文台天文学データ解析計算センタープロジェクト経費、および、文部省科学研究費補助金 (11640341) の援助を得て行われた。

参考文献

- 1) M. A. Albrecht, E. Angeloni, A. Brighton, J. Girvan, F. Sogni, A. J. Wicenec, and H. Ziaeeepour: The VLT Science Archive System in Astronomical Data Analysis Software and Systems VII, *A.S.P. Conference Series*, Vol. 145, R. Albrecht,

- R. N. Hook and H.A. Bushouse, eds., pp. 363–366 (1998).
- 2) T. Horaguchi, S. Ichikawa, M. Yoshida, S. Yoshida, and M. Hamabe: An Archival System for Observational Data Obtained at the Okayama and Kiso Observatories, PNAOJ, **4**, pp. 1–8 (1994).
 - 3) T. Takata, S. Ichikawa, T. Horaguchi, S. Yoshida, M. Yoshida, T. Ito, E. Nishihara, and M. Hamabe: An Archival System for Observational Data Obtained at the Okayama and Kiso Observatories. II, PNAOJ, **4**, pp. 9–21 (1995).
 - 4) 西原英治, 洞口俊博, 伊藤孝士, 高田唯史, 青木賢太郎, 吉田道利, 吉田重臣, 市川伸一, 濱部 勝: データアーカイブシステム MOKA2 の開発, 国立天文台報, **3**, pp. 23–33 (1997).
 - 5) T. Horaguchi, E. Nishihara, M. Yoshida, K. Aoki, T. Ito, M. Watanabe, S. Ichikawa, T. Takata, S. Yoshida, and M. Hamabe: An Astronomical Data Archive System with a Java-Based User Interface, PASJ, **51**, pp. 693–701 (1999).
 - 6) N. Kaifu: Subaru Telescope, in Advanced Technology Optical/IR Telescopes VI, ed. L. M. Stepp, SPIE, **3352**, pp. 14–22 (1998).
 - 7) Subaru First Light Papers, PASJ, **52**, 1 (2000).
 - 8) T. Takata, R. Ogasawara, K. Kawarai, and T. Yamamoto: Data archive and Database System of the SUBARU Telescope, Proc. SPIE, **3349**, pp. 247–254 (1998).
 - 9) T. Takata, R. Ogasawara, G. Kosugi, Y. Mizumoto, S. Ichikawa, N. Yasuda, M. Taga, M. Yagi, T. Horaguchi, H. Baba, M. Watanabe, T. Ozawa, M. Hamabe, T. Yamamoto, and K. Kawarai: STARS (Subaru Telescope archive system) for the Effective Return from Subaru Telescope, Proc. SPIE, **4010**, pp. 181–189 (2000).
 - 10) M. Hamabe, M. Taga, W. Aoki, N. Yasuda, T. Takata, S. Ichikawa, H. Baba, T. Ozawa, M. Watanabe, and T. Horaguchi: New Image Quick-Look System for Subaru Telescope Data Archive, in Astronomical Data Analysis Software and Systems IX, *ASP Conf. Ser.*, Vol. 216, pp. 482–485 (2000).
 - 11) M. Taga, M. Hamabe, N. Yasuda, T. Horaguchi, T. Takata, W. Aoki, T. Ozawa, M. Watanabe, H. Baba, and S. Ichikawa: An Image Evaluation System for Subaru Telescope Data Archive, PNAOJ, **6**, pp. 49–58 (2001).
 - 12) 天文情報処理研究会監修, 日本 FITS 委員会協力: FITS の手引き 第 4.1 版, 国立天文台天文学データ解析計算センター 発行 (2001).
 - 13) 渡邊 大, 青木賢太郎, 三浦 昭, 安田直樹: 多波長天文画像データ検索・閲覧サービス「jMAISON」の開発, 国立天文台報, **6**, pp. 1–9 (2002).

Appendix

セッション管理について

HTTP リクエストは, 基本的にステートレス (state-less) である. すなわち, 1 回のリクエスト/レスポンスでセッションが終了し, それ以降は関係が維持されない. そのため, 複数の画面で構成されるようなアプリケーションを開発するとき, それぞれの画面での持続性を維持する機能を HTTP リクエストより上層で提供する必要がある. これを実現しているのがセッション管理である. セッション管理を使えば, あるユーザーが接続して web サイトを巡回している間, それが同一のユーザーからの接続であることをサーバプログラムが知ることができる.

なお, セッション管理を利用するためには web クライアント側でクッキー (Cookie) をサポートしている必要があるが, 現在標準的に用いられている web ブラウザはすべて Cookie をサポートしており特に問題ないと考えている.