

観測データアーカイブ MOKA と SMOKA

市川 伸一

〈国立天文台天文学データ解析計算センター 〒181-8588 三鷹市大沢 2-21-1〉

e-mail: ichikwsn@cc.nao.ac.jp

観測データアーカイブは、観測データの再利用をはかり、天体の時間変化の追跡や多面的・統計的な天体の把握などの天文学研究や教育の材料を提供するためのシステムである。日本の地上観測で初めての本格的データアーカイブ MOKA と、それを発展させてすばる望遠鏡の観測データの公開を行っている SMOKA についてその歴史と現状、課題と展望について述べる。

1. 観測データアーカイブの意義と必要性

基本的に観察の学問である天文学においては観測データは最も重要な要素である。観測データには時々刻々変化しつつある宇宙のその時の姿が刻まれている。また、観測データには観測者が意図せぬ、あるいは、使いきれない膨大な情報が含まれている。このような貴重な観測データを整理・保管して有効に再利用することは、天文学の発展にとって重要なことであり、そのためのシステムを(観測)データアーカイブと呼ぶ。

観測データアーカイブによって時間的に変化する天体の過去の姿を調べることができるし、また、様々な波長のデータを統合することによる多面的な天文学、多数の天体を扱った統計的な天文学の道もひらけていく。個々の観測データは時間・空間・波長などからなる多次元空間の一点を占めるだけであるが、複数の観測データを組み合わせることによって未知の観点や発想に至ることが期待できるだろう。また、すばる望遠鏡の観測申し込みがかなりの高倍率である、という事実が示しているように、全ての研究者がそれぞれの望む通りに大型観測装置を利用できるわけではない。とりわけ大学院生などにとって、アーカイブデータを使って研究を進め、次の観測申し込みに備えること

ができるという意義は大きいであろう。

教育の現場においては、実際に天体を観測することの意義はもちろん大きいですが、第一線の観測データに触れることの意義もまた大きく、次世代の研究者育成という面と一般の方々に天文学研究への理解を求めるという面で観測データアーカイブは重要である。

望遠鏡や観測装置の製作や運用には莫大な費用がかかっているが、日本ではほとんどの場合、国民の税金がそれにあてられている。観測データから最大限の研究成果を上げるのは研究者の重大な責務であり、観測データを人類共通の資産として様々な面で活用していかねばならないことは明らかであろう。情報やデータの輸入超過の傾向が強い日本の天文学で、観測データを公開するという国際貢献を果たすことの意義も大きいであろう。研究成果も出さないまま観測データを長い期間占有し死蔵に任せることはあってはならない。

観測データアーカイブとして世界で最も優れたものは Hubble Space Telescope (HST) のものであろう。衛星観測であることによる、データの統一管理や観測条件の均一さの面での有利さがあるとはいえ、データ処理や使い勝手まで含めたシステム全体が実に良くできている。多賀正敏氏(現早稲田大学)らが 1999 年 8 月に行った調査によると HST によっ

て産み出された論文 2339 編のうち、211 編がアーカイブを用いてのものであった。全体の約 1/10 という数は 1 つの観測装置に匹敵するものであり、その意義が理解できよう。

本稿では、地上観測における観測データアーカイブの先駆けとなった MOKA と、その発展形であり、すばる望遠鏡の観測データの公開を行っている SMOKA について、その歴史と現状、課題や展望などについて述べる。

2. まず MOKA を

MOKA (Mitaka Okayama Kiso data Archive) は、岡山天体物理観測所と木曾観測所の観測データを扱うシステムである。日本における地上観測の初めての本格的データアーカイブである。

1990 年頃、日本でも CCD データが当たり前ものとなり、データ量の増大と、デジタルデータという写真乾板に比べて四散しやすいその性質から、観測データアーカイブの意義と重要性が強く叫ばれるようになってきた。1991 年のすばる望遠鏡の建設開始もその議論を後押しした。MOKA の開発はそのような状況の下、国立天文台天文学データ解析計算センター、岡山天体物理観測所、東京大学木曾観測所、およびその他の機関に属する有志が集まり、天文情報処理研究会の WG の形で始められた。1994 年 2 月の木曾観測所における合宿が開発のはじまりであり、1995 年 6 月には MOKA の最初の版 (MOKA1)¹⁾ が運用開始となった。しかし、システムに対する基本的考え方がまだまだ未熟であり、また、ネットワーク環境や関連技術が猛烈な速さで発展していた時期でもあり、改良と開発はその後も継続的に進められた。Name Resolver (天体名を座標に変換する) や Coordinate Converter といったツールの開発²⁾ も進めつつ、1996 年 9 月には WWW と分散データベースを用いた MOKA2³⁾ の、1998 年 2 月には JAVA applet を用い、地上観測データの宿命的な伴侶である気象や環境のデータの参照機能を採り入れた MOKA3⁴⁾ の運用が開始

された。開発は、予算の面でも概念設計の完成度の面でも、仕様書を書いて外注するという形をとることはできず、自分達で紆余曲折・試行錯誤を重ねて進めるという形をとった。外注と自力開発という 2 つの開発方法にはそれぞれ得失があるのだが、自力開発の過程で得た様々な有形無形の財産がその後も我々の中に生き続けていることを思えば正解であったように思う。

このようにして MOKA は運用段階に入っていたのだが、残された課題も少なくなかった。その中で最大のものは、計算機資源の問題 (要はお金がなかった) とネットワークの未発達のために、データ本体は各観測所から磁気テープで利用者に送られるという offline 方式であったことである。

3. そして SMOKA へ

SMOKA (Subaru Mitaka Okayama Kiso data Archive) は MOKA の発展形であり、すばる望遠鏡の観測データをも扱うためのシステムである。

世界第一線級の性能を持つすばる望遠鏡の観測データアーカイブには期待も大きく、ファーストライト (1999 年 1 月) に先駆けて 1998 年 1 月から開発がはじまった。開発は MOKA と同様に天文情報処理研究会の WG の形で行われた。SMOKA の開発にあたっては MOKA 開発の際に判明した数々の問題点や残された課題を解決しつつ、より完成度が高く、有用なシステムを目指した。SMOKA の (すばる望遠鏡対応箇所) の開発では具体的に以下の点で進歩があった。

まず、観測データのデータ形式 (FITS) についての規約を早期に定めた。観測データがいつどこをどのように捉えたものであるかの属性情報をもれなく FITS ヘッダに記載するようにしたのである。この作業は、各観測装置や望遠鏡制御系の開発の段階にそれぞれの開発グループと協同でおこなう必要があり、1997 年 4 月から検討を開始し、12 月にはルール第 1 版をまとめた⁵⁾。

次に、すばる望遠鏡が観測後直ちに観測データ

を一括管理し、観測者へ観測データを供給する STARS (高田唯史氏の稿を参照されたい) を採用したことが SMOKA の助けとなった。これによって観測データが四散することなく、また、属性情報を正しくもれなく記録することも推進されたのである。現在、SMOKA は STARS の三鷹のミラーである MASTARS から観測データや関連情報を得ている。

また、何とか計算機資源も確保できたため、観測データの配送を online にすることができた。使いやすさは格段に向上し、SMOKA の登場によって、木曾観測所、岡山天体物理観測所の観測データの利用も格段に増えている。

様々な処理のどの部分をサーバー側とクライアント側 (利用者の WEB ブラウザ) のどちらに担わせるかは悩ましい問題である。ネットワークの発展や、WWW や JAVA といった関連技術の動向、計算機の高速度化、など技術的要素と、利用形態や利用頻度を見きわめて定めることになる。SMOKA では、サーバー側の処理に重点を置くことにし、JAVA Servlet と JAVA Server Pages を採用した。それによって、表示 (User Interface) と内部のアルゴリズムを分離することもでき、開発や維持や新たな観測装置データの組み込みが容易になった。MOKA 時代の分散データベースや分散データ配置の方式は現在の運用体制の実態にそぐわないため採用せず、公開するデータや情報の管理は三鷹に一元化した。

さらに、MOKA の特長の一つであった画像早見システムも大幅に強化した⁶⁾ (表紙参照)。

SMOKA は入念なる検討を含む約 3 年あまりの開発期間を経て、2001 年 6 月より運用を開始し⁷⁾、8 月には MOKA を統合した (MOKA は運用終了)。

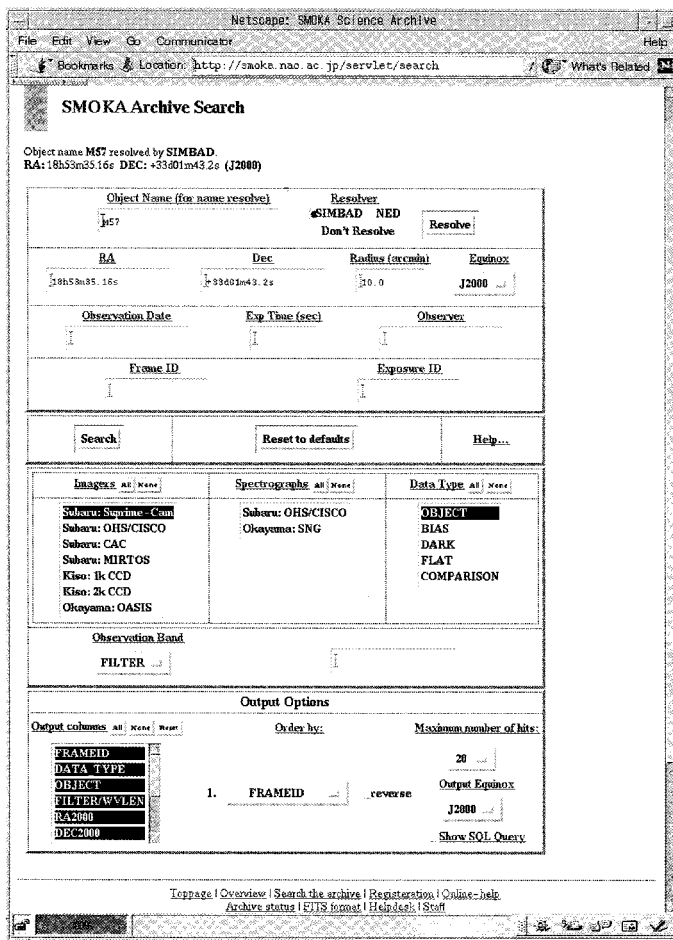


図1 SMOKA の検索画面。利用者の便や将来の統合も考え、HST data archive に酷似させている。

SMOKA は、(運用組織を超えた) 複数望遠鏡の複数観測装置の観測データが一度に検索できるという MOKA の特長を引き継いでおり、現在 3 望遠鏡 13 観測装置の観測データ、総計 20 万フレームを扱っている。すばる望遠鏡の観測データについて、現在はファーストライト期のデータが公開されているだけだが、2002 年 3 月からは観測後 1 年半を過ぎた (ごく一部を除いた) すべての試験観測のデータが、6 月からは同じく共同利用観測のデータが順次公開されていく予定であり、公開フレーム数は格段に増えることになる。

SMOKA ではデータの検索とデータの早見は誰でも行えるようになっている。データ本体の請求(配送)に関しては、システムの過大な負荷を避けるとともに、データの利用状況を把握するために登録制としている。なお、登録のための資格条件は、非営利の研究教育の目的である、ということだけである。

4. 観測データアーカイブの急所

MOKA/SMOKA の開発構築にあたって最も労力と時間を費やしたのは、観測データを観測後一定期間経過後に研究教育の目的で公開するという、データ公開のルールを確立することであった。

MOKA 開発の初期の頃は、観測データの再利用という概念が希薄であり、観測データは観測者が独占するものである、という考え方が根強かったのである。研究には競争の面が強いわけであり、そうした考え方も理解はできるが、日本、あるいは世界全体として最大限の研究成果を目指すためにはそうした意識は変えていかねばならない。木曾観測所は利用者も含めてこの面で先駆的であり、シュミット乾板という非デジタル時代の大容量記録媒体に関して観測後3年という観測者占有期間のルールを既に定めていた。CCD データに関しても特に問題なく「観測後1年で公開」が定められた。しかし、岡山天体物理観測所のデータ公開は、利用者の頑強な抵抗があり難航した。年配の方々からの反対はある程度想定のうちであったが、若い方々からも反対があったのは意外なことであった。最初は不完全な形での公開を余儀なくされ、「観測後2年で公開」が定められたのは運動開始から5年後のことであった。すばる望遠鏡の場合もすんなりとはいかなかった。それでも、SMOKA 開発グループと関係各位の根気強い活動により、当初案よりも少々後退はしたものの「観測後1年半で公開」が定められた。

観測後一定期間経過後にデータが公開される、ということはそれまでに論文をまとめなければなら

ないという無言の圧力を観測者にかけることでもあり、観測データアーカイブの潜在的効能ではないかと考えている。

もう一つの問題(いまだ問題であり続けている)は、いつどこをどのように撮ったデータなのかはつきりしないものが少なくない、ということである。MOKA の開発では、それらの観測データを識別する情報(属性情報)を補ってデータベースに入力することに膨大な労力が費やされた。それでも、木曾観測所の1K-CCDの観測データには観測日が不明なもの(フレーム番号の前後関係から2~3日の範囲には特定できる)などが数多く残されている。移動天体や変光天体の研究の際には深刻な問題であろうが、観測データ取得時に属性情報を正しくもれなく記録しておかなければ、後ではいかんともしがたいのである。

SMOKA の(すばる望遠鏡対応箇所の)開発ではMOKAの反省に基づき、前述のように観測装置や制御系の開発の段階から観測所の担当者や観測装置開発グループと入念な検討を繰り返し、FITSヘッダの規約として属性情報を漏らさず記録することに努めた。その実現にあたっては、STARSによる観測データ一括管理が絶大な効果をもたらした。現在、機器のトラブルによりごく少数のデータで属性情報が欠落している場合があるが、ほとんどのものは問題なく属性情報を備えている。

SMOKA に組み入れて再利用を行う価値があると思われる観測データは他にもいくつかあるが、これらは属性情報が不詳であったり不十分であったり誤っていたりするものが多く、それを補う労力や時間と価値とを天秤にかけた上で組み込みを断念している。

属性情報を正しく漏らさず記録するためだけでなく、観測データの特性の理解、処理解析、品質管理など様々な面で、観測データアーカイブの開発、運用には、観測所職員や観測装置開発グループとの連携協力が必要不可欠である。開発グループの参加協力が得られない場合、その観測装置の

観測データを組み込むことは実際上不可能である。

以上の他に、技術的側面での要点も数多くあるが、それらは論文を参照していただくことにしてここでは省略する。利用者が使いやすいシステム、管理者が維持運用しやすいシステム、開発者が改良・拡張しやすいシステムを目指す、ということは言うまでもないだろう。

5. まだまだ遠い道程

このように運用を開始した SMOKA であるが、観測データアーカイブとしてより天文学研究に役立つ存在としていくために今後解決してゆくべき課題は山積みである。

SMOKA の運用体制の極度な貧弱さについては日本の天文学全体に共通する課題でもあり、ここでは論じない。

まず、地上観測データの宿命である、気象や環境の状況を把握するための機能である。気象環境データの参照機能は MOKA の時代に開発されたのだが、実際にはそれらのデータを継続的省力的に組み入れる仕組みが未完成であった。各観測所で稼働している全天モニター画像の組み込みは急務であろう。

提供する観測データの質についての目安を示すことは重要である。研究に用いるデータの質について、最終的にはそれぞれの研究者の観点と責任で判断すべきものではあるが、気象環境データだけでなく、可能な限りの判断材料（観測時のメモ等も含まれよう）を提供すべきであろう。

次に、現在 SMOKA で提供しているのは生の観測データだけである。利用者にとっては一次処理が済んで、赤経赤緯、フラックス、波長、といった物理量が得られるデータがもちろん望ましい。天体を検出してカタログ化したものが公開できれば有用性がさらに増そう。しかしながら、地上観測という気象環境の変化が避けられない宿命を負った上に、すばる望遠鏡も木曾観測所も岡山天体物理観測所もサーベイ観測という均質な観測方法では

なく、時間割り当て方式で観測者それぞれの目的に沿った観測方法をとるが主である。全ての観測データに（半）自動的な一次処理を施すのはたいへん困難である。可能なものに関して可能な限りの処理を施して提供するのが現実的であろう。

さらに、システムの使い勝手の向上（各種ツールや一覧図・一覧表の表示も含む）だけでなく、高度な検索、例えば、天体名を特定して検索するのではなく、「こういう性質を持った天体」の含まれる観測データを探したり、移動天体に対応できるようにすることが望まれる。

利用の面でも多くの課題（裏返せば期待）がある。宇宙研 PLAIN センターとの共同開発である多波長画像表示システム jMAISON⁸⁾との連携を実現し、例えば、HST の画像とすばる望遠鏡の画像を手軽に重ねて表示できるようにしたい。

また、教育への活用も期待されている。例えば、球状星団の多色撮像画像から HR 図を得るなど、第一線の観測データで教材を作ろうという活動が洞口俊博氏（国立科学博物館）や五島正光氏（巢鴨高校）を中心に PAOFITS という WG で進められている。このための SMOKA の機能強化も期待されている。

最後に、SMOKA 開発運用グループも率先して研究成果をあげ、自らその有用性を示すべきであろう。様々な天文データを総合して天文学をすすめるとともに、各種天文データの扱いや統計に詳しい研究グループを形成していくことを目指したい。

観測データアーカイブをはじめとした天文データを活用して研究成果をあげるためには、ある程度以上のデータの蓄積がなければならない。すばる望遠鏡が本格的に稼働しはじめた今がまさにその時期にきているといえるだろう。

謝辞

残念ながら紙面の都合で名前をここであげることはできないが、MOKA/SMOKA の実現は、開発運用に携わってきた多くの方々の方々の尽力の賜物である。

過去の開発運用メンバーについては、参考文献にあげた各論文の著者名を見ていただきたい。また、現在の開発・運用メンバーについては SMOKA の WEB を参照していただきたい。

MOKA/SMOKA の開発にあたっては、文部省科学研究費補助金 (06554001, 07304024, 08228223, 11640341), 国立天文台共同開発研究経費 (平成 6, 7, 9 年度), 国立天文台天文学データ解析計算センター開発経費 (平成 8 ~ 13 年度) などの援助を得た。

また、国立天文台天文学データ解析計算センター、光学赤外線天文学・観測システム研究系、岡山天体物理観測所、ハワイ観測所、東京大学木曾観測所など多くの方々の御支援と御協力をいただいた。この場を借りて深く感謝したい。今後も多くの方々の期待に答えられるシステムの実現に努めていきたい。

参考文献

- 1) Horaguchi T., et al., 1994, Publ.NAOJ, 4, 1
- 2) Takata T., et al., 1995, Publ.NAOJ, 4, 9
- 3) 西原英治 ほか, 1997, 国立天文台報 3, 23
- 4) Horaguchi T., et al., 1999, PASJ, 51, 693
- 5) 金光理/天文情報処理研究会, 「FITS の手引き第 4.1 版」, 2001
- 6) Taga M., et al., 2001, Publ.NAOJ, 6, 49
- 7) 馬場 肇ほか, 2002, 国立天文台報 (投稿済)
- 8) 渡邊 大ほか, 2002, 国立天文台報 (投稿済)

参考 URL

SMOKA
<http://smoka.nao.ac.jp>
 Publ.NAOJ
<http://ads.nao.ac.jp/jp/pub/publication.html>
 国立天文台報
<http://ads.nao.ac.jp/jp/rep/report.html>
 FITS の手引き
<http://jaipa.nao.ac.jp/jfits/>
 jMAISON
<http://maison.nao.ac.jp>, <http://maison.isas.ac.jp>

Data Archive Systems: MOKA and SMOKA Shin-ichi ICHIKAWA

Astronomical Data Analysis Center, National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588

Abstract: Data archive is the system which keeps and supplies observational data for astronomical researches or educational purposes. I introduce the data archive systems, MOKA and SMOKA. MOKA is a pioneer in data archive systems for ground-based telescopes in Japan. SMOKA is the successor of MOKA, and deals with the observational data of the Subaru telescope. I mention the history and present status of them, then discuss the future task and prospect.