

すばる望遠鏡公開データアーカイブシステムの開発 5

山田 善彦, 小澤 友彦^{*1}, 西澤 淳, 古荘 玲子, 西村 高德^{*2},
榎 基宏^{*3}, 吉野 彰^{*4}, 古澤 順子, 高田 唯史, 市川 伸一

(2009年4月25日受付; 2009年8月31日受理)

Development of Public Science Archive System of Subaru Telescope 5

Yoshihiko YAMADA, Tomohiko OZAWA ^{*1}, Atsushi NISHIZAWA ^{*2}, Reiko FURUSHO, Takanori NISHIMURA,
Motohiro ENOKI ^{*3}, Akira YOSHINO ^{*4}, Junko FURUSAWA, Tadafumi TAKATA, Shin-ichi ICHIKAWA

Abstract

We have improved the public science archive system, SMOKA (Subaru-Mitaka-Okayama-Kiso Archive system). We developed of the “overlapped sky area search” by using HEALPix indices. This search makes users be able to know where are the sky areas covered by multiple observations, those observed with several filters, and so on, from archived frames of Suprime-Cam. It must be a useful tool for data search especially aiming to study variable objects, moving objects etc. We updated the data transfer system between observatories and SMOKA system, for more effective and secure data handling and operation. We started providing new observational data; those are the data from Kyoto3DII and MOIRCS which are new instruments on Subaru Telescope, CCD cameras on MITSuME Telescope, and HOWPol on the KANATA Telescope. We also improved Web user interface for involving and browsing new environmental data for Subaru telescope, such as image data of sky monitor and seeing information from Auto Guider and DIMM.

1. はじめに

天文学研究において観測は主要な柱であり、観測データはその天域・観測時間における唯一無二の貴重な記録である。そこで、それを散逸しないよう保存管理し、後の研究・教育目的の利用に供するのがデータアーカイブシステムの最大の存在意義である。天文データアーカイブは、天文学の発展に重要な役割を担っていることは既に過去の

論文^{1), 2), 3), 4)} (以下論文1, 論文2, 論文3, 論文4) において述べてきたが、その重要性はアーカイブデータの蓄積・多様化とともに増している。

世界的に見れば、地上観測所によって得られた観測データのデータアーカイブシステムは、近年着実な広がりや成長を遂げてきている。例えば、2005年4月よりESO (European Southern Observatory)^{*1)}の各望遠鏡で取得された観測データは、専有期間後、世界中の人々に公開されるようになった。Keck望遠鏡^{*2)}の観測装置であるHIRES (High Resolution Echelle Spectrograph)のデータも、生データのみであるが世界中に公開されるようになってきている^{*3)}。また、データ・コンテンツの充実も着実に行われており、較正済みデータを公開する試みが一部のアーカイブで始まっ

^{*1}みさと天文台
Misato Observatory

^{*2}東北大学
Tohoku University

^{*3}東京経済大学
Tokyo Keizai University

^{*4}早稲田大学
Waseda University

ている。例えば、ESOはVLT (Very Large Telescope) のUVES (Ultra-violet and Visible Echelle Spectrograph) やLa-Silla 3.6m望遠鏡のHARPS (High Accuracy Radial velocity Planet Searcher) といった観測装置のデータをパイプラインで処理したものを*⁴⁾, CADC (The Canadian Astronomy Data Centre)*⁵⁾ はハワイのCFHT (Canada France Hawaii Telescope)*⁶⁾ の広視野可視光撮像装置であるMegaCamの処理済み画像を公開し始めている*^{7) 5)}。これらの処理済みデータ公開は、現在世界的な潮流となりつつある仮想天文台構想 (VO: Virtual Observatory)*⁸⁾ の中核をなすものとして積極的な取り組みが行われた結果であるが、今後のデータ検索機能の充実と相まって更に効率的なデータ再利用を促進するものとして期待されている。

我々は、2001年6月以来、国立天文台ハワイ観測所のすばる望遠鏡、岡山天体物理観測所188cm望遠鏡、ならびに東京大学木曾観測所105cmシュミット望遠鏡の観測データを保管・提供するアーカイブシステムSMOKA (Subaru-Mitaka-Okayama-Kiso Archive system)*⁹⁾ の運用を行ってきた。そして、上に述べたような世界的な潮流に合わせ、新機能の開発や新しい観測装置データの組み込みなどを行ってきた (論文1, 論文2, 論文3, 論文4)。SMOKAの公開から8年経ち、SMOKAを用いた天文学研究が浸透してきている。SMOKAは多くの利用者に活用され、それを利用した研究論文も50本を超え、その分野・利用方法も多岐に渡っている*¹⁰⁾。最近では、国立天文台のハワイ観測所・光赤外研究部と連携し、大学学部学生や大学院生ら国内外の若手の研究者を対象に開催しているデータ解析講習会の一環として、SMOKAの利用講習・案内を行っており、若手の研究者の一層の利用が期待されている。SMOKAのデータは、研究活動だけでなく、教育普及方面でも活用されている。その一例としてPAOFITS WG*¹¹⁾ による教材開発が挙げられる。PAOFITS WGは、FITS画像を教育に活用する目的で活動しているワーキンググループで、小・中・高校および大学の教員、国立天文台を始めとする研究機関職員、科学館・博物館職員、および公開天文台職員などが参加している。その成果として、SMOKAデータを用いた教材開発についての教育論文も出版されている⁶⁾。

SMOKAは利用者数も増え、SMOKAから取得

したデータを用いた成果も増えてきたが、天文学研究・教育をより推進していく上では、まだ開発・改良すべき余地が多くある。これまで、我々は、基本となる単純検索・アドバンスド検索の他にも、すばる望遠鏡のSuprime-Cam⁷⁾ のデータに特化したSuprime-Cam専用検索、日付による検索を可能にするカレンダー検索、FITSヘッダ情報を詳細に検索する全文/全項目検索など、用途に応じた様々な検索機能を開発してきた。しかし、これらの検索は全て、天体・日時・座標・観測者といった項目で「特定の」データを探し、絞り込むためのものであった。つまり、総露出時間が何秒以上である領域、何回以上露出されている領域、といったような場所・日時を特定しない検索はできなかった。このような検索は、移動天体や変光天体を捜したり、露出時間やカラーバンド数を条件にしたサンプルで統計的研究を進めるために必要な機能であるが、これまでは検索の仕組みなど技術的な問題によって、実現が困難であった。また、SMOKA利用者の増加、および利用者の要求する機能の多様化によって、システムの複雑さや運用の煩雑さが増している。このような状況下、限られた人数で運用を行うには運用の効率化が必要であり、可能な限りの処理の自動化・省力化が望まれる。

本論文では、以上の課題を解決すべく開発・補強された機能について論ずる。2章では、本論文の主要な開発項目である重複領域検索について解説する。3章では新観測装置データと観測所からのデータ転送の仕組みの改良を、4章ではデータ請求機能・検索機能の改良および環境データの表示・検索の改良などについてを述べる。2008年春に行われた計算機システム更新に伴うSMOKAの改訂、SMOKAの利用状況とSMOKAを用いた研究成果、運用に関する課題については5章、今後の課題とまとめを6章で行なう。

2. 重複領域検索

本章では、HEALPix (Hierarchical Equal Area and iso-Latitude Pixelization)*^{1) 8)} を用いた、すばる望遠鏡のSuprime-Camデータに対する重複領域検索機能の開発について述べる。2.1節では開発の目的について、2.2節では空間分割法とインデックスについて、2.3節では検索機能の開発の詳細について、2.4節では検索試験について、2.5

節では実装へ当たっての問題と方向性について論じる。

2.1 開発の目的

これまで、SMOKAにおいては、単純検索・アドバンスド検索の他、Suprime-Cam専用検索・カレンダー検索・全文／全項目検索といった複数の検索機能を開発してきた（論文1，論文2，論文3，論文4）。その中で基本となる検索は、位置情報となる天球座標（赤経・赤緯など）を元にしたものである。すばる望遠鏡の共同利用観測装置であるSuprime-Camは、10枚のCCDチップを持つモザイクカメラ（一度の露出で10フレームのデータを生成する）である。この場合、視野の端のフレームであっても視野中心付近の赤経・赤緯がFITSヘッダに書き込まれており、SMOKAで検索する際もこの座標を用いているため、「ある領域があるフレームに含まれているか」を厳密に判断しているわけではない。したがって、何回露出されているか、総露出時間が何秒であるかを領域毎に正確に数え上げることができず、複数回露出されているような領域を検索することはできなかった（図1）。

Suprime-Camの共同利用が始まってから10年が経った今、デザインングによって意図的に重ねられた場合を含め、全天の中で重複して撮像された領域も多いと考えられる。小惑星などの移動天体や、超新星・変光星など変光天体の検出には時間的に間隔をおいた同じ領域の画像が必要であり、重複して観測された領域を検索する必要がある。合計で一定以上の露出時間をもつ領域を得たい場合にも、露出時間の情報を持った重複領域を検索する必要がある。重複して観測された領域というのは、必ずしも同じフィルターを用いて撮像されているわけではなく、様々な天文学的興味により、多くが複数種類のフィルターを用いて撮像されている。このような領域を検索するためには、各領域がどのようなフィルターで観測されている

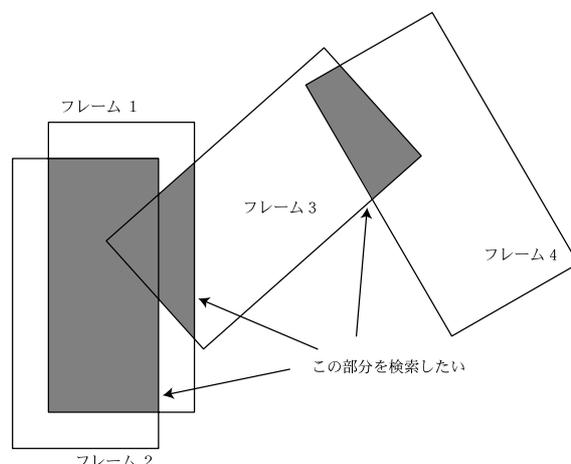


図1. 重複領域検索の概念図. 幾つかのフレームの重なっている領域（灰色部分）を検索し、それを含むフレームを結果として与える。

かという情報と何回露出されているかという情報が必要で、重複領域検索の派生と言えらる。

このような検索を実現するため、空間分割法の一つであるHEALPixを用いた検索機能を開発することとした。SMOKAがデータを扱っている撮像観測装置は16あるが、今回我々はSuprime-Camデータのみを対象にこの検索機能を開発した。その理由としては、既に正確な位置較正が行われている⁹⁾ データが存在することと、装置の視野が広く、多くの重複領域の存在が期待できることなどが挙げられる。

2.2 空間分割法HTMとHEALPix

天体の天球での位置情報は天文学の中でも重要な情報の一つである。通常、位置情報は赤経・赤緯などの座標で表されており、天体の同定・検出を行うような際には、この座標が与えられれば十分である。しかし、天球の領域同士を比較するような場合には、赤経・赤緯のような座標値では、非常に扱いづらい場面が多くなる。それは、赤経・赤緯などは天球上の「点」を表す情報であって、領域を示す「面」の情報ではないからである。これは、天球の任意の形・面積の領域を赤経・赤緯（と何らかのパラメータ）で表現しようとした時に、非常に困難であることから理解できる。

今回我々が実現しようとする重複領域検索では、ある領域がSuprime-Camフレームに含まれるか否か、ということが重要となってくる。この機能の実現のためには、天球座標といった点で位置を表すようなものでは不十分であり、領域とい

^{*1} <http://www.eso.org/public/>

^{*2} <http://www.keckobservatory.org/>

^{*3} <http://www2.keck.hawaii.edu/koa/public/koa.php>

^{*4} http://archive.eso.org/eso/eso_archive_adp.html

^{*5} <http://www2.cadc-ccda.hia-ihh.nrc-cnrc.gc.ca/cadc/>

^{*6} <http://www.cfht.hawaii.edu/>

^{*7} <http://www.cadc.hia.nrc.gc.ca/megapipe>

^{*8} <http://www.ivoa.net/>

^{*9} <http://smoka.nao.ac.jp/>

^{*10} <http://smoka.nao.ac.jp/about/publish.ja.jsp>

^{*11} <http://paofits.nao.ac.jp/>

う「面」を扱う目的に適した空間の表現方法を用いる必要がある。そこで、天文学で利用されている2つの空間分割法（領域指定法）HTM (Hierarchical Triangular Mesh)^{*2) 10)}とHEALPixの使用を比較検討した。これらの空間分割法は、インデックスというものをを用いて領域の名前付けを行っている。インデックスとは、実態のデータ（空間座標）とは別の、索引用の空間の位置データであって、機能的に領域を扱うことができるものである。

上記2つの空間分割法の共通点として、再帰的な領域分割が挙げられる。これらの分割法は、その分割に階層を持ち、階層数が深くなるに従って、空間の分割は細かく、一つの領域は小さくなる。階層が一段階深くなると、各領域は4分割され、それぞれに新しいインデックスが振られる。ただし、そのインデックスの上位桁は、その親領域と一致するものである。例として、HEALPixの場合を解説する。図2の左側はHEALPixの、ある階層での領域とそのインデックス番号で、右側は同じ領域をもう一段階階層を深くした場合のインデックス番号を表している。図の左側には、4つの領域がありそれぞれ2桁のインデックス（2進数表記）が付与されている。階層を深くする場合、左側で01のインデックスが付いていた領域は、右側では4分割されており、そのそれぞれの領域には、右下→右上→左下→左上の順に00から11までのインデックスが付与される。ただし、この場合2進数表記の上2桁は必ず01となっており、一つ上の階層の情報を保持している。このように、任意の階層のインデックスを2進数表記で表した場合、上2桁は最も浅い階層でのインデックスを表しており、順次、次に深い階層のインデックスの情報を保持している。HTMにおいても表現される領域の形状は違うものの、基本的には同様にインデックス番号が振られる。このシステムによ

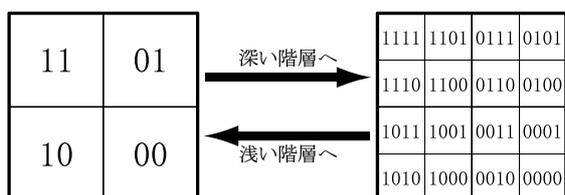


図2. HEALPixの任意の階層におけるインデックスの振り方。数字はインデックスを表している。右側は左側のインデックスから階層を1階層深くした場合。

って、どちらの空間分割法も階層の昇降（領域の拡大・縮小）を容易に扱うことができるようになっている。

HTMはSDSS (Sloan Digital Sky Survey)¹¹⁾用に開発されたもので、まず天球を8個の象限毎の球面三角形に分割し、それぞれを各辺の中点を結ぶ大円で再帰的に4つの球面三角形に分割していくという球面分割法である(図3)。北半球に属する領域のインデックスには先頭にNが、南半球のそれには先頭にSが付けられている。HTMの分割法は、数式を用いずとも、非常に直感的なものとなっておりわかりやすい。しかし、各領域は必ずしも等面積ではない。

HEALPixは、宇宙背景放射の非等方性を探る研究のために、NASAが開発したもので、全天を等面積に分割する手法である。HEALPixの領域は2本の極を通る大円と赤道、それから極角の余弦と方位角の関数として記述される球面上の曲線により12分割される(図4(a);詳細は8)の図3と8)の式19-22を参照)。これがHEALPixで表すことのできる最も浅い階層の領域で、1領域の大きさは差し渡し約60度である。深い階層の領域は、これを4等分していくことで再帰的に得られる(図4(b)(c))。HEALPixによるインデックス化によって、全天は $N_{index} = 12 \times 2^{2N}$ 個の等面積な領域に分割される。 N_{index} はインデッ

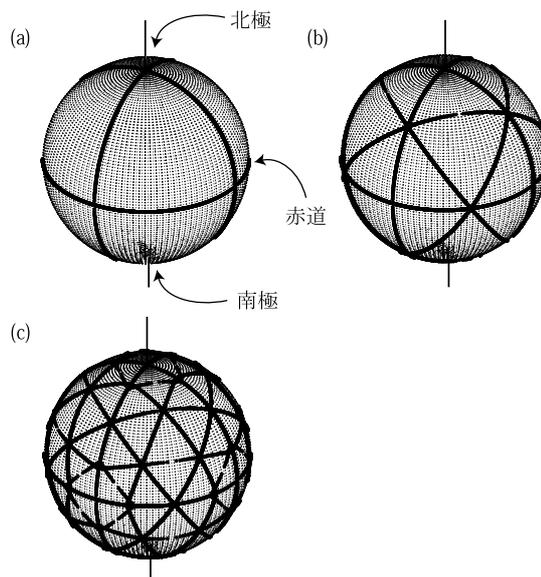


図3 (a), (b), (c). HTMの空間分割法。(a) 天球を8つの球面三角形に分割する。(b), (c) それぞれの球面三角形を、各辺の中点同士を結ぶことで、再帰的に4分割していく。

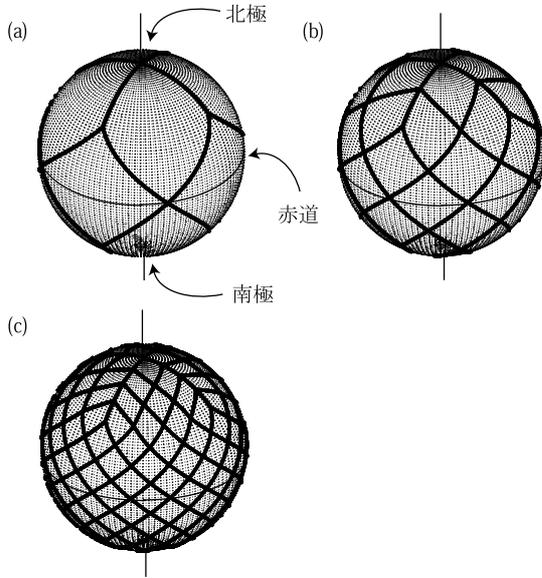


図4 (a), (b), (c). HEALPixの空間分割法. 参考のため、北極・南極と赤道(細い実線)を示している。(a) HEALPixの領域は極を通る2本の大円と赤道、それから極角の余弦と方位角で記述される単純な式が表す球面上の曲線により12分割される。(b), (c)より階層の深い領域は、これを4等分していくことで再帰的に得られる。詳細は8)の図3と8)の式19-22を参照のこと。

クス数, N_{res} は解像度変数で, N_{res} 階層と呼ぶ。表1に, 階層, インデックスの総数, 各インデックスの領域面積を示す。

HTMと違い, HEALPixは同じ階層での各領域が等面積であることが大きな利点である。階層を浅くしていった(一つの領域は大きくなる)も, 領域同士の比較が等面積で容易であるため, 統計的研究にも適当である。また, HEALPixの領域は, その定義から等緯度線上に並んでいる。このことは球面調和関数との親和性が良いことを示しており, ある量の全天での分布を簡単に統計的に扱えるなど, 天文学的研究にも大きな利点がある。実際にこれを用いて, 宇宙背景放射分布に関する研究(WMAP¹²⁾などで大きな成果を上げている。将来的にSuprime-Camデータが全天の多くの部分を埋めるようになった時に, あるいは, Hyper Suprime-Cam¹³⁾がすばる望遠鏡に搭載されサーベイ観測が行われるようになった時に, HEALPixを用いて領域が検索できることは, 前述の宇宙背景放射の研究などとも親和性が非常に高く, データおよびSMOKAの利用価値が高まることを期待され, 将来的にも有益である。また, HEALPixは, 天文学のデータの標準的形式であるFITS¹⁴⁾の天球座標表現の規格として採用されている*³⁾ ¹⁵⁾ことも重要な要素である。このよう

表1. HEALPix インデックスの階層と, インデックス桁数, 全天でのインデックスの総数, それぞれの領域面積。

階層 N_{res}	インデックス の桁数	インデックス数 N_{index}	領域面積		
			deg^2	arcmin ²	arcsec ²
0	5	12	3437.747		
1	7	48	859.437		
2	9	192	214.859		
3	11	768	53.715		
4	13	3072	13.429	48343.313	
5	15	12288	3.357	12085.828	
6	17	49152	0.839	3021.457	
7	19	1.97×10^5	0.210	755.364	
8	21	7.86×10^5	0.052	188.841	
9	23	3.15×10^6	0.013	47.210	
10	25	1.26×10^7		11.803	42489.240
11	27	5.03×10^7		2.951	10622.310
12	29	2.01×10^8		0.738	2655.577
13	31	8.05×10^8		0.184	663.894
14	33	3.22×10^9		0.046	165.974
15	35	1.29×10^{10}		0.012	41.493
16	37	5.15×10^{10}			10.373
17	39	2.06×10^{11}			2.593
18	41	8.25×10^{11}			0.648
19	43	3.22×10^{12}			0.162
20	45	1.32×10^{13}			0.041
21	47	5.28×10^{13}			0.010

NASA 1234567
パッケージは
← 12 階層まで対応

← 1 秒角以下の精度

← 21 階層まで拡張

な様々な利点から、今回我々が実現しようとする重複領域検索には、HEALPixを用いることにした。

2.3 検索機能の開発

2.3.1 HEALPixプログラムの改造

HEALPixのパッケージはNASAから提供されている。このパッケージには、天球座標とインデックスの変換プログラム、任意の円形領域に含まれるインデックスの検索プログラムなどが含まれており、我々の重複領域検索では基本的にこれを用いることにした。我々は検索精度を1秒角以下にすることを目標としているため、インデックス化を1秒角以下の領域まで行う必要があり、 $N_{\text{res}} = 18$ 階層の分割が必要になる(表1)。しかし、NASAから提供されているHEALPixパッケージでは、内部の整数変数は全て4バイト整数で宣言されているため、 $N_{\text{res}} = 13$ 階層(約1分角に相当)以上に細分化することができない。そこで我々は関連する箇所の整数を全て8バイト整数で宣言し直し、最大で21階層まで分割できるようプログラムに改良を加えることで、目標とする検索精度を実現した。

2.3.2 Suprime-Cam フレームの視野のHEALPix インデックス化

すばる望遠鏡のSuprime-Camの場合、1フレームの視野は13.5分角×6.8分角であり、1フレーム全てを18階層のみの領域を用いてインデックス化すると、約51万個という膨大な量のインデックスが必要となる。これでは検索の速度も遅くなるばかりか、ディスク領域も圧迫されてしまう。そこで、細分化しなくてもフレームの位置情報が損なわれないインデックスについてはより浅い階層(=大きな領域)のインデックスを割り当てるといった手法を取った(図5)。具体的には、フレームの境界付近では1秒角精度で境界を定義しなければならないが、フレームの境界から遠い中心部では、より浅い階層(大きな領域)でインデックス化すれば良い。実際の手順は、11階層から領域が大きな順に、フレームに完全に含まれる領域を埋めていき、18階層ではフレーム全体を覆うようにするものである。この手法により、

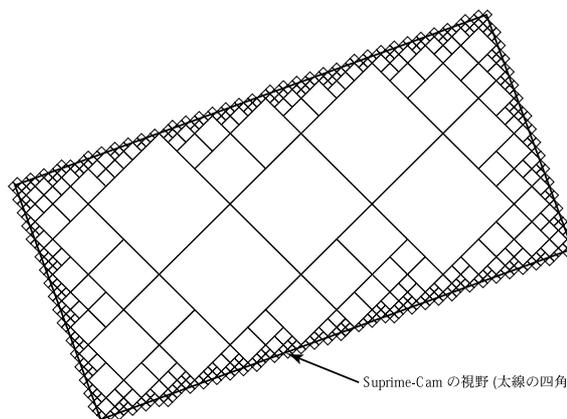


図5. Suprime-CamのフレームへのHEALPix領域の埋め込み方(概念図)。Suprime-Camのフレームを太線で示す。インデックスの数を減らす為、まず最初にフレームに入る最も大きな面積を持つ領域を置き、それから順次、下の階層(面積1/4)の領域で埋めていく。17階層まで繰り返し、最後18階層で周辺部を覆う。

Suprime-Camの1フレームを、18階層のみのインデックスで表記した場合の約1/100、平均5300個のインデックスで記述できることになる。なお、天球上に投影されたフレームは、光学的歪みによって歪んだ四角形となっている。したがって、フレームが持つWCS(World Coordinate System)情報からその歪みを考慮して、天球に投影されたフレームに対してインデックス化を行っている。

2.3.3 HEALPixインデックスのデータベース化

重複領域検索を実現するためには、データベース(リレーショナルデータベース)上にHEALPixインデックスを適当なテーブルとして入力しておく必要がある。ここではそのテーブルと入力情報について述べる(図6)。リレーショナルデータベースを用いて検索機能を実装する理由は、入力・検索プログラム開発の容易さ、検索速度を高速化可能なこと、他の検索機能との親和性が良いためである。

(1) テーブル1: 各フレームに含まれるインデックスを示したテーブル

最大18階層で表された、各フレームを「面」で表現するためのインデックスを入力したテーブル。カラムはフレームIDとインデックス、その階層である。

(2) テーブル2: 各フレームに含まれる12階層で示されたインデックスと、露出時間・フィルター情報を持つテーブル

*1 <http://healpix.jpl.nasa.gov/>

*2 <http://skyserver.org/HTM/index.html>

*3 http://fits.gsfc.nasa.gov/iaufwg/IAUtrans_2006.pdf

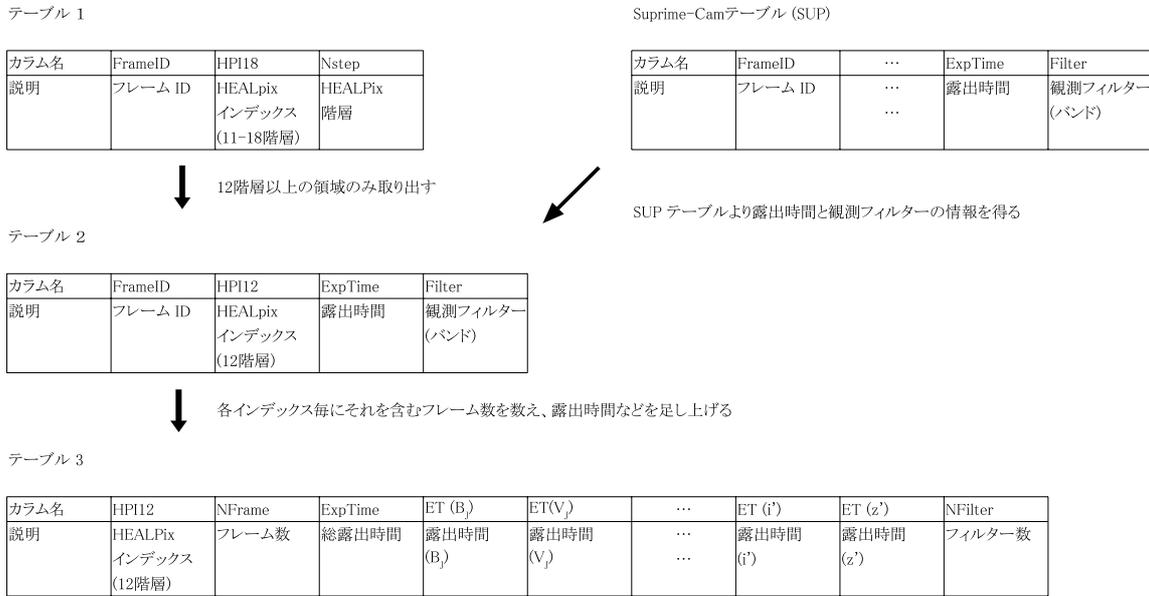


図 6. データベースのテーブル構造と、テーブル生成の流れ. テーブル 1 : 各フレームに含まれる 11-18階層の HEALPix インデックスを示す. テーブル 2 : テーブル 1 から 12階層以上の領域のインデックスを抽出し, Suprime-Cam テーブル (SUP) より各フレームの露出時間とフィルターの情報を付加したもの. テーブル 3 : 各 HEALPix インデックスを含むフレームが何フレームであるかと, その総露出時間等を示したテーブル. ピンポイント検索ではテーブル 1 のみを, 重複領域検索ではテーブル 2 と 3 を用いる.

ピンポイント検索 (後述) では, 1 秒角の精度で, ある赤経・赤緯が含まれるフレームを検索しているが, 重複領域検索においては, 僅か 1 秒角の領域が重なっていても利用価値が少ないと考えられるため, 1 分角以上が重なっている領域を検索することにした. したがって, このテーブルには, 12階層の HEALPix インデックスの情報をテーブル 1 より抽出することにし, Suprime-Cam の諸データが記録された運用テーブル (SUP) より抽出した各フレームについての露出時間とフィルター情報を併せて入力した.

- (3) テーブル 3 : 各 HEALPix インデックス (12 階層) を含むフレーム数, 総露出時間, フィルター毎の総露出時間, 撮像されたフィルター数を情報として持つテーブル

テーブル 2 より算出されるテーブルで, 各インデックスが何枚のフレームに含まれるか, また, その総露出時間がどれだけであるか, 何枚のフィルターで観測されているかを示したものである.

1 秒角の精度で必要な領域を含むフレームを効率よく検索するピンポイント検索は, テーブル 1 のみを用いる. 具体的には, 利用者から入力された天球座標の赤経・赤緯をインデックスに変換

し, その領域を含むフレーム ID をテーブル 1 から検索し, 結果として与えるという流れである (図 7 上). 一方, 重複領域検索は, テーブル 2 と 3 を用いる. まず, 与えられた条件 (何回以上露出されたか, など) に合致するインデックスをテーブル 3 から検索する. 次に, そのインデックスをテーブル 2 から検索することで, 重複領域を含むフレーム ID を結果として与えるという流れである (図 7 下).

元々は重複領域検索を実現するために空間分割法の一つである HEALPix を用いたが, このことによって, ピンポイント検索も単純なアルゴリズムで実現できるようになった. ピンポイント検索機能は, 従来の Suprime-Cam 専用検索でも実現されていた (論文 4) が, そのアルゴリズムが SMOKA 独自のもので, 新しいフレームが追加される度に, インデックスを一から作り直さなければならないという問題があった. このため, 重複領域検索の開発とともに, ピンポイント検索のアルゴリズムも HEALPix を用いたものに変更した. HEALPix を用いた検索では, 新しいフレームが追加された場合は, そのフレームに含まれる HEALPix インデックスの行のみ変更を加えればよいので, 大幅に運用が効率化される.

精度 1 秒角のピンポイント検索を行うために

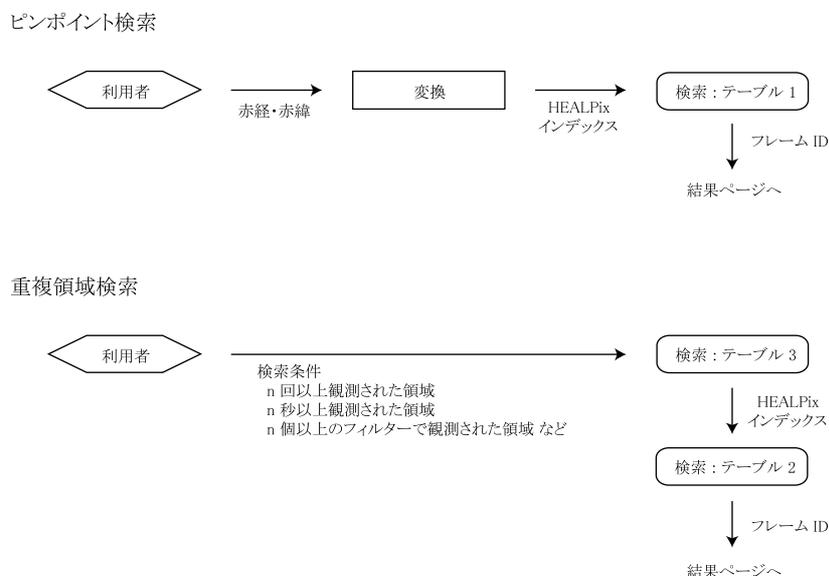


図7. ピンポイント検索と重複領域検索における検索の流れ。ピンポイント検索は、利用者から入力された天球座標の赤経・赤緯をHEALPixインデックスに変換し、そのインデックスを含むフレームIDをテーブル1から検索し、結果として与える。重複領域検索は、与えられた条件（何回以上露出されたかなど）に合致するインデックスをテーブル3から検索し、そのインデックスをテーブル2から検索することで、重複領域を含むフレームIDを与える。

は、フレームが持つ元々の位置精度が1秒角以内より十分小さい必要がある。しかし、Suprime-Camの生データの状態では、その位置を示すWCSの精度は最大20-30秒角程度の誤差があり⁹⁾、そのままそのフレームに含まれるインデックスを計算してテーブルに入力する訳にはいかない。したがって、0.2-0.3秒角程度の位置精度で位置較正されているフレームに対して、データベーステーブルを作成した。本研究で用いたデータは、2002年9月から2003年6月までの18886フレームである。実際に情報を入力した結果、テーブル1は約1億行、テーブル2は約179万行、テーブル3は約19万行のテーブルとなった。入力に要した時間は、FUJITSU PRIMEPOWER 450 (CPU: SPARC64V (2GHz) × 4, Memory: 8 GB) を用いて、テーブル1は約60時間、テーブル2は約6時間、テーブル3は約60時間であった。なお、データベース管理ソフトウェア (DBMS) としてSybase 12.5を用いた。

2.4 検索試験の結果

2.3.3節においてテーブルを作成したのと同じ環境で、検索の試験を行った。

(1) ピンポイント検索：

赤経・赤緯を入力すると、インデックス (18階層) に変換したものをテーブル1から検索し、

結果としてそのインデックスが含まれるフレームIDを得る。実現される検索機能は、論文4に示されたピンポイント検索と同等である。すべてのインデックスを一つのテーブルに入力した場合、検索に要する時間はおおよそ3分であった。この時間は天域によらない。しかし、今後フレーム数が増加し、検索により時間がかかるであろうことも考えると、これは現実的な検索時間ではない。これを解決すべく、赤緯1度毎に分割したテーブル (テーブル1aとする) を作成した。このテーブルの入力には約6日を要した。このテーブル1aに対しての検索時間は検索する赤緯により異なるが1秒以下～数秒程度であり、これは、今後フレーム数が増えても現実的な検索速度であるといえる。さらに、検索インデックスを分割されたテーブル1aに対して張ることで、全ての領域で0.1秒以下にまで検索速度を速くすることが出来るが、検索インデックスを張るとテーブルの入力速度が遅くなる。そのため、テーブル入力速度との兼ね合いを計りつつ、検索インデックスの張り方の最適化が必要となる。

(2) 重複領域検索：

例えば150回以上観測された領域を条件とする (図8 (a)) と、これをテーブル3から検索し、その条件に該当する領域のリストを得ることが出来る (図8 (b))。この検索に掛かる時間は1秒

以下であった。図 8 (b) は 150 回以上観測された領域が 8 箇所であることを示す。この領域のインデックスをテーブル 2 で検索する (利用者側にはインデックスは表示されない) ことで、目的のフレーム ID を得る (図 8 (c))。この結果はそれを含むフレーム数は 1248 フレームであったことを示している。フレーム数が多いため、検索には時間がかかっている (約 1 分)。

2.5 実装への課題

この機能を SMOKA に実装して運用に至るためには、幾つかの課題が残されている。

まず第一に、検索とテーブル入力に要する時間である。ピンポイント検索については、Suprime-Cam 専用検索に搭載されていた従来のピンポイント検索 (詳細は論文 4) の代わりとして実装される予定である。将来的に位置較正が為されるであろう 40 万フレーム程度を対象とした場合に、現実的な時間内でテーブル入力と検索ができることが必要となる。前節でテーブルを赤緯 1 度毎に分割する、検索インデックスを張るという試行をおこなったが、さらなる最適化が必要であろう。

第二に、検索結果の表示をいかに利用者

りやすく行うかである。図 8 に示されたようなテーブル形式ではなく、よりグラフィカルな表示を含めて、利用者が結果を直感的に理解できるような仕組みを検討せねばならないだろう。

第三に、フレームのインデックス化と、テー

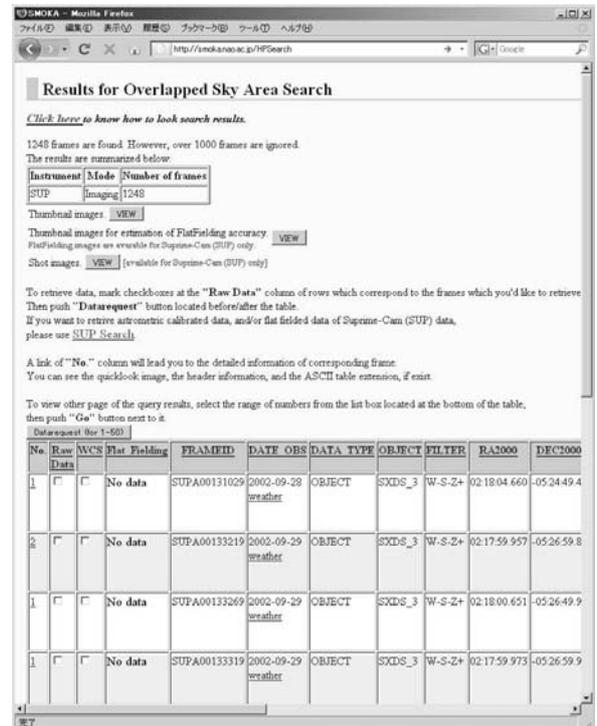
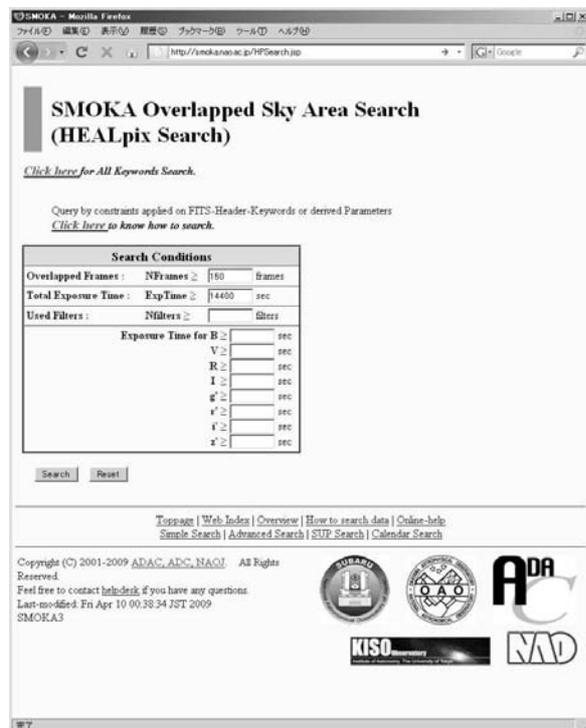
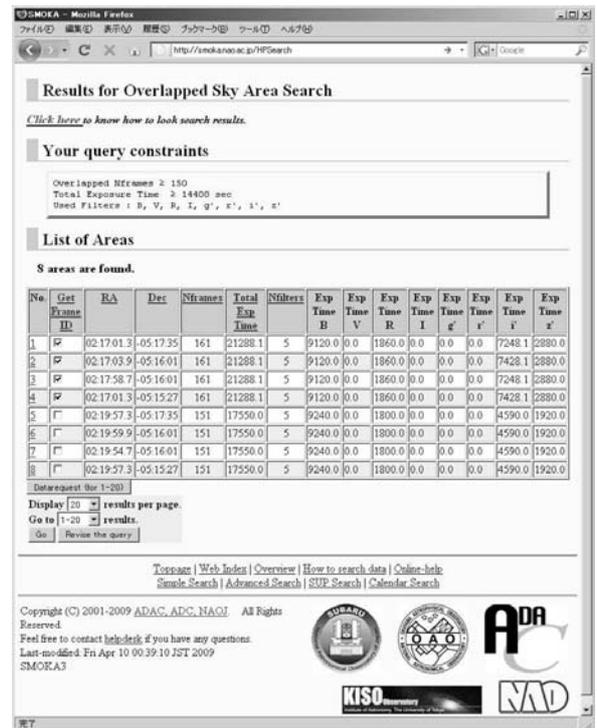


図 8 (a), (b), (c). 重複領域検索の Web インターフェース。(a) 検索ページで条件を入力し、(b) 結果ページ 1 で合致した領域の数を得て、その中でフレーム ID のリストを欲しい領域についてチェックを付け検索すると、(c) 結果ページ 2 でそのフレーム ID が得られる。

ルへの入力の効率化・高速化である。今回の試験では、サーバー1台のみを用い、19000フレームのインデックス化とテーブル入力に、全体で3週間程度掛かっている。今後は、インデックス化アルゴリズムの更なる検討と、サーバーを複数台用い並列化させるなどの工夫が必要となる。

3. 新観測装置データの組み込みと観測所からのデータ転送システムの開発

SMOKAは光赤外線望遠鏡のデータアーカイブシステムとして、国内では主要なものとして認知されてきている。従来よりSMOKAがデータをアーカイブしてきた国立天文台ハワイ観測所・岡山天体物理観測所・東京大学木曾観測所だけではなく、国内各地の天文台から、観測データをアーカイブして欲しいとの要望も上がってきており、各地の天文台との連携もますます重要となってきた。しかも、大学や地方自治体単位での天文台運営は、予算・人材の問題から、自前では観測データのアーカイブや公開まで達成するのがなかなか困難なのが現状である。SMOKAでデータを公開することは、各地の天文台にとって、その手間や予算の節約とデータの有効利用、SMOKAグループにとっては一大データ集積地として利用価値を高めることができるため、双方にとって利益が大きい。

本章では、論文4以降、新規にSMOKAに組み込まれたデータと、各観測所とSMOKA間のデータ転送システムについて述べる。上述のように、SMOKAではより多くの観測装置の観測データを組み込むことを目標としている。3.1節では、すばる望遠鏡の新規観測装置 Kyoto3DII (Kyoto tridimensional spectrograph II)^{16), 17)} と MOIRCS (Multi-Object InfraRed Camera and Spectrograph)^{18), 19)} の観測データの組み込みについて、3.2節では MITSuME (Multicolor Imaging Telescopes for Survey and Monstrous Explosions) 望遠鏡群²⁰⁾ の観測データの組み込みについて述べる。3.3節では各観測所からのデータ転送方式の統一化について述べる。

3.1 すばる望遠鏡 Kyoto3DII と MOIRCS データの組み込み

Kyoto3DIIは、京都大学を中心としたグループによって作製された(1) ファブリペロ分光撮像、

(2) マイクロレンズアレイを用いたIntegral Field 分光、(3) 長スリット分光、(4) 狭/広帯域フィルタ撮像、の4つの観測モードを持つ可視光の多モード分光装置で、すばる望遠鏡のカセグレン焦点に設置されている。視野は3.6秒角×2.8秒角であり、4000–7000 Åの波長領域をカバーしている。

MOIRCSは、(1) 広視野撮像、(2) 多天体分光、の2つの観測モードを持つ、近赤外観測装置で、すばる望遠鏡のカセグレン焦点に設置されている。視野は4分角×7分角であり、0.85–2.5 μmの波長領域をカバーしている。

SMOKAでは、すばる望遠鏡の新規観測装置のデータ組み込みについては、データベースの作成や検索部の改造などの際に、一定の作業手順が確立されており、これらの装置のデータについても、その手順に従って組み込みを行った。具体的な手順は以下の通りである。

- (1) 検索に用いる各装置ごとのデータベーステーブルを新たに作成する。
- (2) GUIの観測装置選択メニュー、カレンダー検索などに新規装置を追加する。
- (3) 検索に用いるSQL文を生成するプログラムに、検索条件として新規観測装置を選択できるように追加作業を行う。
- (4) 各装置についての概要と利用者ドキュメントを書き加え、アーカイブ状況を表示できるようにする。

今回もこの手順で問題なく組み込みが行えた。また、データの登録作業については、すばる望遠鏡の他の観測装置と同様に行っている。

Kyoto3DIIについては2006年6月30日から、MOIRCSについては2006年7月6日からSMOKAでの公開を開始した。2009年3月現在、近赤外の撮像装置としては広い視野を有しているMOIRCSのデータのSMOKAへの請求数は一ヶ月平均1200フレームに上っており、Suprime-Camに次いでデータ請求の多い装置となっている。

3.2 MITSuMEデータの組み込み

MITSuMEは、東京工業大学理工学研究科宇宙物理学研究室(以下、東工大)の河合誠之教授らの研究チームによって作製された、ガンマ線バースト残光の観測を行うための、東京大学宇宙線研究所明野観測所(以下、明野)と国立天文台岡山天体物理観測所(以下、岡山)にそれぞれ設置さ

れた可視50cm望遠鏡と、既存の岡山天体物理観測所の赤外91cm望遠鏡の3台の総称である。明野と岡山の可視50cm望遠鏡には視野が30分角×30分角でg', Rc, Icの三色同時撮像 CCD カメラが設置され、稼働している。さらに将来、岡山の赤外91cm望遠鏡には、56分角×56分角の視野を有する OAO Wide Field Camera (OAOWFC)²¹⁾ が設置される予定となっている。MITSuMEは、ガンマ線バースト残光観測のための望遠鏡・装置であるが、多数の天域の画像を3色同時に広視野撮像しているため、その観測データはアーカイブ化し公開する価値が高いと考えられる。

岡山50cm望遠鏡と明野50cm望遠鏡で取得されたMITSuMEデータは観測後速やかに東工大河合研究室に転送され、データベースへの登録や位置較正、早見画像の作成などの処理が施される。河合研究室の研究者や共同研究者などは東工大からこのデータを取得する。SMOKAでは観測翌日の午後（東工大での処理が済んだ後）に三鷹への観測データ転送を行い、SMOKAの非公開部に蓄える。この蓄積はMITSuMEデータのバックアップを兼ねているものである。転送にあたっては他の観測所のデータと同様にMD5チェックサムによる確認を行い、ネットワークの不調などによって転送の失敗があった場合（週一度程度起きている）には再転送を行っている。そして、あらかじめ東工大で設定された公開日（通常は観測後1年）に達するとデータをSMOKAの公開部に移すと同時に、SMOKAで取り扱いやすい標準フレーム名への変換を行い、検索用データベースへの登録を行った後、利用者に公開している。以上のデータの流れを図9に示した。

明野のデータについては2007年4月25日からSMOKAでの公開を開始した。しかし、東工大での位置較正によって、FITSヘッダに記入された赤緯が90度を超えているなど、不正であるフレームが多数含まれていることが判明したため、一旦公開を停止した。その後、不正な天球座標値を望遠鏡のポインティング値に置き換えて提供することにした。そのための作業が完了した後の2008年6月16日より、岡山のデータと共に提供を開始した。

3.3 データ転送システムの統一化

これまでも述べたとおり、我々SMOKA開発チームではより多くの観測所・望遠鏡システムで

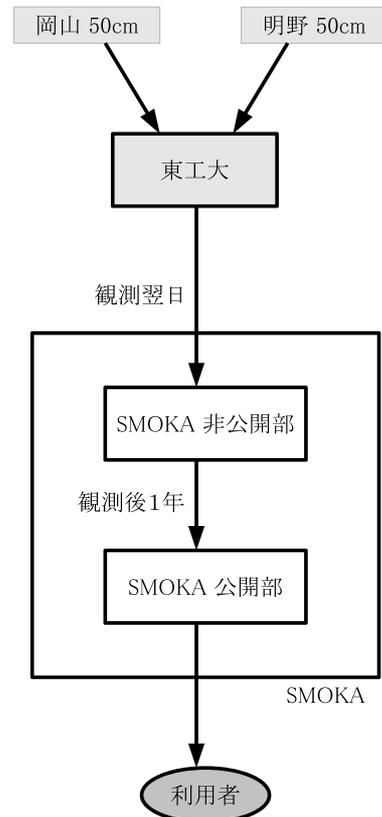


図9. MITSuMEデータ（岡山50cm望遠鏡，明野50cm望遠鏡）のSMOKAでの公開までの流れを示した図。両望遠鏡で取得された観測データは観測後速やかに東工大（河合研究室）に転送され、データベースへの登録や較正等の処理が施される。SMOKAでは観測翌日の午後（東工大での処理が済んだ後）にデータを転送して非公開部に蓄える。あらかじめ設定されている公開日（通常は観測後1年）に達したデータは順次公開部に移され、フレーム名の書き換え等の処理を施され、検索用データベースへの登録が行なわれた後に利用者に公開される。

取られた観測データの組み込みを目指しており、今後も新規にデータを受け入れSMOKAに組み込む観測所を増やしていく計画である。現在、各観測所・望遠鏡システムのデータは、それぞれのネットワーク環境や事情に合わせてさまざまな方法で転送され、SMOKAへ組み込まれている。岡山天体物理観測の観測データはネットワーク経由で自動転送されている一方で、東京大学木曾観測所では人が出向いてデータ搬送を行っていた。そのため、観測データを観測所からSMOKAシステムに運ぶためのシステム（転送済データの管理を含む）やSMOKAへの組み込み手順もまちまちで煩雑な状況にあった。これらを簡略化することで、新たな観測所から観測データを転送し組み込む際

の開発要素を最小限にでき、かつ運用での負担も軽減され、新たな観測所からのデータ受け入れを容易にする。

今回、新たに広島大学東広島天文台かなた望遠鏡²²⁾のHOWPol (Hiroshima One-shot Wide-field Polarimeter)²³⁾データのSMOKA組み込みを目指したデータ転送が開始された。また、東京大学木曾観測所のネットワーク環境が增強されてネットワークによる観測データ転送が可能になる予定である。これを機に、観測所からSMOKAへのデータ転送システムの統一化を図った。本論文執筆時点では、ハワイ観測所・岡山天体物理観測所・東工大・東広島天文台・東京大学木曾観測所の5ヶ所からのデータ転送を行っている。そのデータの流れの概略を図10に示す。今回の統一化は、東広島天文台、岡山天体物理観測所、木曾観測所の3観測所からのデータ転送システムを対象とした。すばる望遠鏡とMITSuME望遠鏡データが含まれないのは、他の3観測所と違い、すばる望遠鏡はSTARS^{*1)}とMASTARS^{*2)}というハワイ観測所の管理システムが、MITSuMEでは東工大が、観測直後データおよび転送を管理しており、SMOKA側でそれを管理する必要が無いからである。

既にネットワーク転送を行っていた岡山天体物理観測所のシステムの問題点をもとに策定された、新転送システムが満たすべき要件を以下に挙げる。

- (1) 観測所において転送するデータのファイルパスが変わっても、柔軟に対応できること。
- (2) 転送済データの把握が容易であること。
- (3) 三鷹へのデータ転送後からSMOKA公開までの期間に、観測所側がデータの公開・非公開を制御できること。
- (4) 観測所でトラブルが発生しても、あるデータが転送済みであるかどうかの照合が容易であること。

まず、(1)の要件を満たすため、転送プログラムはカレントディレクトリのデータについて動作するように構築しなおした。次に、(2)から(4)の要件を満たすため、転送済観測データを三鷹に設置したサーバー上の管理データベース(PostgreSQL^{*3)})を用いて管理することとした。これにより、まず、データの転送・未転送の把握が容易になり、さらに転送プログラム(スクリプト)の簡略化にも繋がった。観測所側では、三鷹側の管理データベーステーブルにアクセスすることで、データを三鷹へ転送した後も、容易にデータの公開・非公開を制御可能になった。さらに、三鷹側はデータの転送情報をデータベースで一元管理するので、ネットワークの不調やデータの誤削除など、観測所でトラブルが発生しても容易に転送情報を復旧できるようになった。ただし、三鷹では定期的にバックアップを取得していることが前提である。管理データベースとしてPostgreSQLを選択したのは、フリーウェアなの

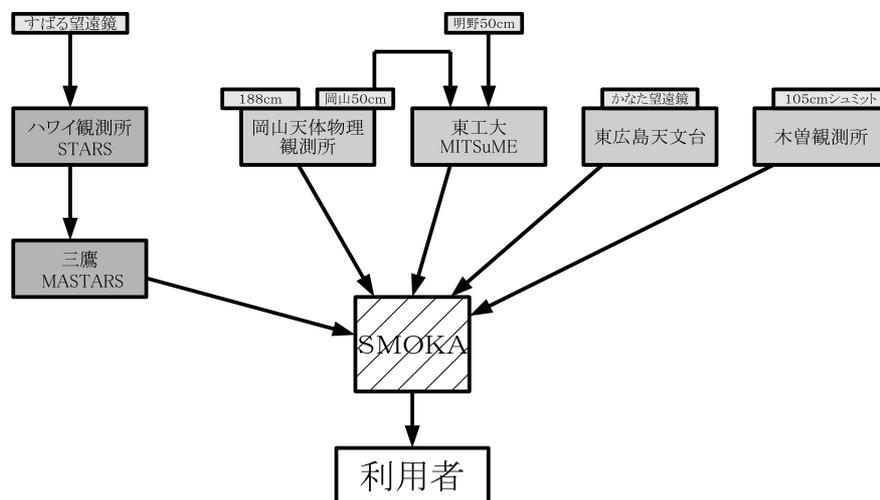


図10. 各観測所とSMOKAの間のデータの流れの概要。すばる望遠鏡の観測データは公開になった後に三鷹MASTARSからSMOKAに転送されるが、その他の観測所の観測データはバックアップの役割を兼ねて、観測後可能な限り速やかにSMOKAに転送されている。

でクライアントソフトウェアの導入コストがかからないことと、MITSuMEのデータ管理システムで運用実績があることによる。

図11に示すように、データ転送の流れは以下のようなになる。

- (1) (三鷹の) 管理用データベーステーブルから転送済ファイルのリストを取得する (各観測所)。
- (2) 管理データベーステーブルと照合して未転送のデータをリストし、MD5チェックサム値を計算しファイルに落として三鷹に転送する (各観測所)。
- (3) 転送すべきデータリストを元にデータを転送する (各観測所)。
- (4) 転送されたデータの情報とMD5チェックサム値ファイルの内容を管理データベーステーブルに登録し、転送されたデータファイルについてMD5チェックサム値を計算し管理データベーステーブルに登録する (三鷹)。

既に述べたとおり、観測所から転送される観測データは、観測所毎に分離された三鷹にある管理データベースを用いて管理されている。管理データベースのテーブルカラムは、FRAME-ID、FITSヘッダキーワード値 (DATE-OBSなどの基本的なものや各観測所の要望によるキーワード)、公開日、公開可否フラグ、各観測所における転送前データファイルのMD5チェックサム値および三

鷹転送後データファイルのMD5チェックサム値からなる。転送されてきたデータのデータベーステーブルへの入力作業は三鷹で行っている。観測所で計算した観測データファイルのMD5チェックサム値についても、一旦リストファイルに出力して三鷹に転送してから三鷹でデータベーステーブルへ入力している。これは、管理データベーステーブルへの情報入力において、ネットワークが不調な場合のトラブルを最小限に抑えるためである。観測データは観測から公開までの期間内にSMOKA非公開部に転送されるが、試験観測データなどもあるため、それらのデータ全てが公開されるとは限らない。各観測所では、PostgreSQLクライアントから三鷹の管理データベースにアクセスし、データベースの検索・更新や公開可否フラグの変更などを行うことができる。

今回の転送システムの統一化により、各観測所からのデータ転送をほぼ同一のシステムで行うことが可能になった。これにより、今後新しくSMOKAにデータを組み込む観測所が増えても、ネットワーク環境さえ整っていれば容易に対応することが可能となった。新転送システムでは、転送元にPostgreSQLクライアントをインストールした計算機の準備が必須となったため、その点では新規観測所側の手間が増えた。しかし、これに関しては、必要に応じてSMOKA運用メンバーが新規観測所に数日出向き、場合によってはデータ

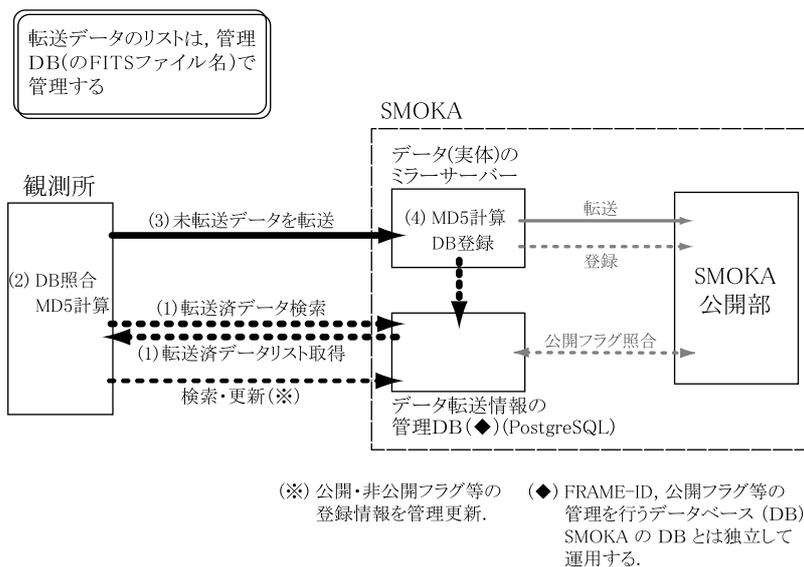


図11. 各観測所からのデータ転送の流れの詳細。実線矢印はデータファイル (実体) の移動を、破線矢印はデータベース (DB) テーブルへのアクセスを示す。(1) から (4) までが観測所からSMOKA非公開部へのデータ転送である。SMOKA非公開部から公開部へのデータ転送は灰色の矢印で示した。

転送用計算機の準備から設置・設定を行うことで対応が可能である。以上の開発により、これまでの初期設定作業および運用の作業量からすれば格段に負担が軽減されるようになった。

4. その他の新機能の開発

本章では、2, 3章で述べた以外の、新機能の開発や機能の改良について述べる。4.1節では利用者から要望の高かった大量データ請求機能の開発について、4.2節では座標による検索の際の視野加算方式の改良について、4.3から4.6節では、フィールドモニター画像、シーイングモニターのデータ、サミットログなど、品質評価やデータの取捨選択・解析に役立つ情報の新規公開やその表示機能について、4.7節では論文4で開発した全文検索機能の改良について述べる。

4.1 大量データ請求機能

4.1.1 電子メールによるデータ請求機能

各データ検索機能を用いたデータ請求方法では、HTML表示の関係上一度に請求できるフレーム数を限っているため、大量にデータを請求する場合は請求操作を繰り返す必要がある。そこで我々は、利用者の利便性向上を目的として、大量のデータ請求を可能にする本機能を開発した。従来はハードウェア性能の限界もあり、実現が困難であったが、計算機システム更新によるハードウェア性能の強化で運用面での障壁が取り除かれたことで可能となった。

図12は本機能を用いたデータ請求の概要を示したものである。本機能は観測データを請求する機構の一つであり、図中のStep.1-3を担う。従来のデータ請求機能では、Web経由でフレームIDを選択して請求するのに対して、本請求機能では、フレームIDのリストを規定の書式に則って記載したメールを送付して請求する。利用者がメールを送付すると、Data Copy Manager (DCM, 論文1) がFTPサーバーへ請求データを展開し、処理終了後、FTPによってデータを取得できるようになる。データ請求過程でメールを受け付けて処理するという部分以外は、従来のシステムと同じである。

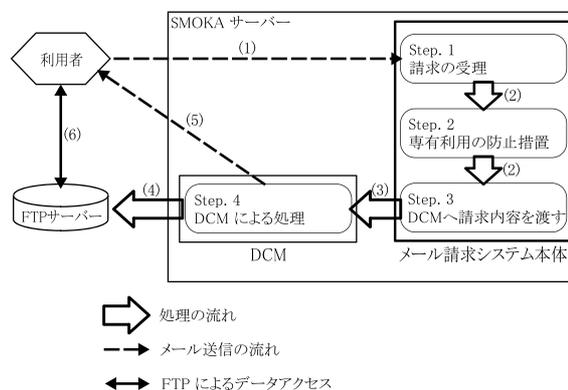


図12. メールによる大量データ請求機能を用いたデータ請求の流れ。図中、白抜き太矢印は処理の流れ、破線矢印はメール送付の流れ、実線矢印はFTPによるデータへのアクセスおよび転送を示す。大まかな流れは次のとおり: (1) 利用者からのメールによる請求, (2) メール請求機能における処理 (Step.1 - Step.3), (3) Data Copy Manager (DCM) へ請求内容を渡す, (4) DCMによるFTPサーバーへの請求データ展開, (5) 利用者宛に処理開始/終了の通知メール送信, (6) 最後に、処理終了のメールを受け取った利用者がFTPサーバーにアクセスし、データを取得する。今回開発したのは、太線枠内である。

本機能が備えるべき要件として以下の4点を設定した。

- (1) 従来のデータ検索機能を用いる場合の上限数 (1000フレーム) 以上の請求に対応できること。
- (2) SMOKA利用登録者以外からの請求は受け付けないこと。
- (3) 特定利用者による専有的使用を防止し、利用者ができるだけ公平にデータを取得できること。
- (4) 従来のデータ検索機能を用いた請求をこれまでと変わりなく行えること。

(1), (2) を備えるため、データ請求の操作として、「規定の書式で記述した電子メールの送信」を採用した。このメールには、件名にパスワードを記載する必要がある。規定の書式とは、SMOKA利用者ID、データ取得の目的、そして請求するデータのフレームIDであり、簡素なものとなっている。また、(3), (4) を満たすため、各利用者の請求件数とデータ量を監視し、同じ利用者の請求を同時に多数流さないようにした。

特定利用者の専有的使用を防止するため、現状では一通あたりの請求フレーム数に上限を設けざ

*1 <https://stars.naoj.org/>

*2 <https://www.mastars.nao.ac.jp/mastars/>

*3 <http://www.postgresql.jp/>

るを得ない。それでも従来のデータ検索機能を用いた場合に比べて、最大10倍(10000)のフレームを一度に請求することが可能となった。2008年09月の機能公開以来、本論文執筆時点まで大きな問題は起きていない。ただし、本機能を利用する利用者が当初の予想よりも少なく、公開から一年程度の時点で、機能の存在が周知されているか、使い辛いと受け取られていないか議論・調査が必要である。また、本機能によるデータ請求に要する時間はDCMが行っているFTPサーバーへのデータコピーにかかる時間が主である。今後の課題として、DCMの高性能化を検討する必要がある。

4.1.2 検索結果テキスト表示と全フレームリスト

従来のシステムでは検索結果をHTMLに整形し、テーブルとして提供していた。しかし、メールによるデータ請求のためには、ASCIIテキストで検索結果を提供する必要がある。そこで、以下の2つの機能を新たに追加した。

- (1) 検索結果をHTMLのテーブルに整形せず、ASCIIテキストとして出力する。
- (2) 観測された全フレームをASCIIテキストとして観測年毎(装置によっては月毎)に公開する。

これらの出力を元に利用者はメールによるデータ請求を容易に行うことができる。以下ではそれぞれの機能の詳細について述べる。

(1) 検索結果の出力：検索画面において、出力形式の選択肢として、HTMLテーブル形式とASCIIテキスト形式がある(図13(a))。テーブル形式を選択すると、従来通り、ショット画像やサムネイル、気象情報などへのリンクがあり、検索結果についての多くの情報を得ることができる。ASCIIテキスト形式を選択すると、利用者が選択した出力内容(例えば、FRAME_ID, DATE_OBS, OBSERVER, FILTERなど)の値が羅列されたASCIIテキストが出力される(図13(b))。利用者はこの出力をそのままメールでのデータ請求に使用することができる。尚、HTML出力では表示時間の都合上、最大で1000フレームまでしか出力しない設定にしているが、ASCIIテキスト出力では、表示にかからないことと、メールでの一回の最大請求数が1000であることを考慮して最大20000行まで出力することが

できるようにした。

(2) 観測フレームの全リスト表示：現在公開されている全ての観測フレームをリストにし、観測年毎に表示される(図14(a), (b))。ただし、MITSuMEだけはデータ量が多いので観測月毎に表示する。この観測フレームの全リストは週一回自動的に更新される。観測日他必要な情報は更新時にSMOKAのデータベースより取得され、年毎あるいは月毎にASCIIテキストファイルとして保存されており、利用者が年月を指定すると表示される仕組みとした。

以上のように、ASCIIテキストでフレームIDや観測日などの情報を容易に得ることができる。利用者はこれをそのまま、あるいは任意に編集し

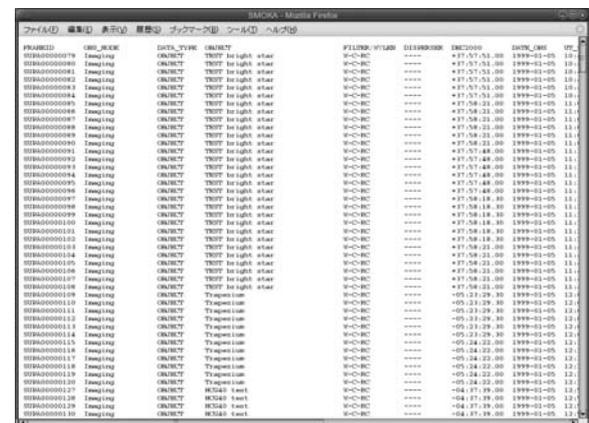
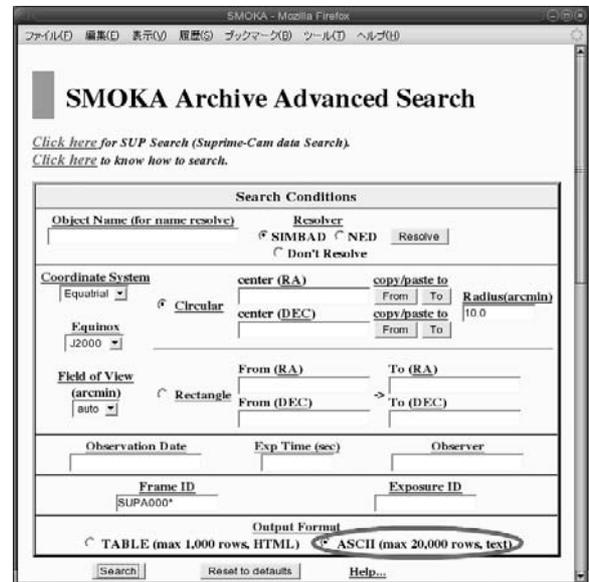


図13 (a), (b). (a) 検索画面における、テーブル表示とASCIIテキスト表示の選択ボタン。
(b) ASCIIテキストを選択した場合に出てくる、検索結果のASCIIテキスト表示画面。

て、メールにより大量に一括データ請求を行うことができるようになった。

4.2 検索時における視野分加算方式の改良

SMOKAには、天球座標の情報から該当するデータを検索する機能がある。例えば、赤経赤緯の値（中心座標）と半径を入力することで、その円領域に登録されているフレームを検索することができる。また、四角領域も検索可能で、赤経赤緯のそれぞれの最小値と最大値を入力することで、

囲まれた球面四角領域内に登録されているデータを検索することもできる。このように、領域を指定して検索を行う場合、データベースに登録されているフレームの座標はフレームの中心座標なので、視野の端に写っている天体は検索結果に反映されない可能性がある。従来SMOKAでは、このような取りこぼしが生じないように、利用者が指定した領域に自動的に観測装置固有の視野を加算し、より広い領域で検索を行っていた。視野分を加算することで検索時に取りこぼしを防止することができる一方で、検索条件に合致するフレーム数がいたずらに増大してしまう。登録されているフレーム数は年々増加しており、得られた検索結果の中から必要なフレームを探すのが困難になってきた。また、加算される視野の大きさは公開されておらず、利用者の理解の上で指定した領域と実際に検索している領域が一致していなかった。

このような問題点を解決するために、今回我々は視野分を加算するかどうかを利用者が選択できる方式にした。選択肢は次の4通りである。(1) 加算しない、(2) 装置毎の視野を加算する、(3) 0.5 - 100分角の固定値を加算する、(4) 複数装置選択の場合は最も広い視野を持つ装置の視野を加算する。このような選択方式を取ることで、利用者はより効率的に目的とするデータに辿り着くことができるようになった。

次に、円領域検索に関する改良点を述べる。従来、中心座標と半径から検索を行う場合、半径は極方向と方位角方向の幅と見なしていた。つまり、中心座標±半径の球面四角領域が検索領域と読みかえて検索を行っていた。もし半径が小さく、赤道付近であれば、これは円領域と殆ど違いがないが、極付近や半径が大きくなるとその差異が顕著になる。従来このような手法を取っていたのは、円領域での検索を実現するためには三角関数を使用する必要があり、検索速度が低下する懸念があったためである。今回我々は、検索速度を保ちながら円領域を検索するアルゴリズムを開発した。全データの中から直接円領域を検索すると、前述の通り三角関数を多用するので検索速度は低下する。そこで円領域は中心座標±半径の四角領域に内接することに着目し、まず全データの中から四角領域に入るフレームだけを探し出し、その後四角領域内部で円領域に入るフレームを検索するようにした。そうすることで、三角関数を用いて検索する回数を大幅に減らし、検索速度を低下させ

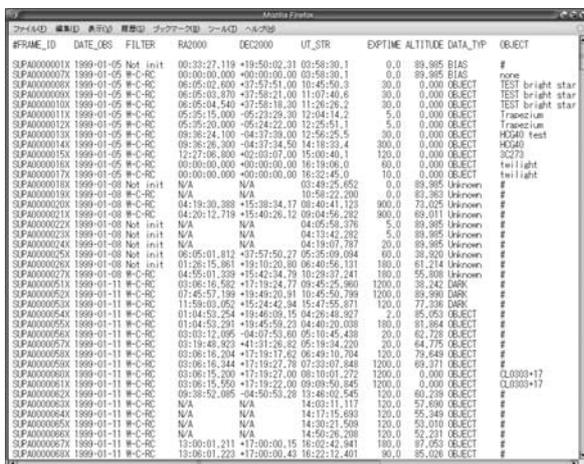
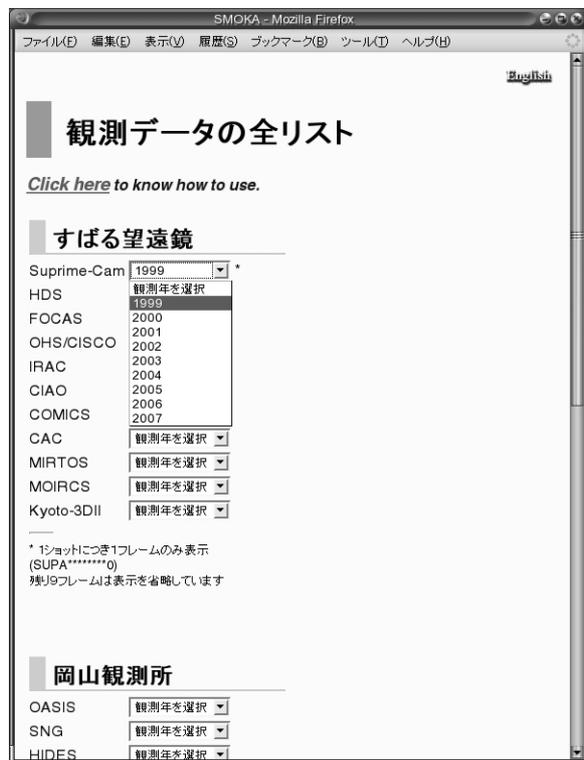


図14 (a), (b). 全フレームリストの選択画面と、その結果表示。選択画面で装置と年を選択すると、その年にその装置で取得されたデータの全フレームのリストがASCIIテキストで表示される。

ることなく円領域内に含まれるフレームを検索することができるようになった。

以上の機能を改良・追加することで、利用者が理解しうる検索条件通りの結果を得ることができるようになり、検索条件をより詳細に設定できるようになった。なお、検索範囲に対する視野の加算については、図15にその概念図を示した。

4.3 環境データ表示機能の再編

環境データとは、観測時の気温・湿度や風向・風速などの数値化された気象データならびに雲モニターなどの画像データである。直接の観測者でない、SMOKAからデータを取得してデータ解析しようとする利用者の場合、観測当夜の状況は、こういった気象データを利用する以外に知ることができない。よってSMOKAでは、解析結果の質や信頼性を判断する上で重要な情報として、環境データを順次公開してきた。また、そのユーザーインターフェースとして、すばる望遠鏡気象モニター表示機能（論文2）、木曾観測所および岡山天体物理観測所の天候データ表示機能（論文3）を開発してきた。しかし、環境データの閲覧時のユーザーインターフェースが観測所毎に異なり、観測所間の比較ができなかった。また表示形式（MPEGおよびJPEG）が違うので、描画に使用するソフトウェアが異なり閲覧環境によっては描画できないなどの利用者側での問題も生じていた。

そこで、ハワイ観測所、木曾観測所および岡山天体物理観測所の環境データの表示を統合した。統合は、木曾観測所および岡山天体物理観測所の天候データ表示機能と同様な閲覧形式となるよう、ハワイ観測所の気象モニター表示機能を変更するものである（図16）。これは多くのWebブラウザの本体機能のみで表示ができることを重要視したためである。このようにして、ハワイ観測所、木曾観測所および岡山天体物理観測所の環境データ表示は、同じユーザーインターフェースとなり、比較が容易となった。

4.4 AGシーイング、DIMMシーイングの表示

観測データを扱う者にとっては、時々刻々と変わるシーイングの情報は、観測時の大気の状態を知る手がかりとして重要である。ハワイ観測所では、観測当夜の状況監視や観測方針の決定のために、2つのシーイング測定機を設置してい

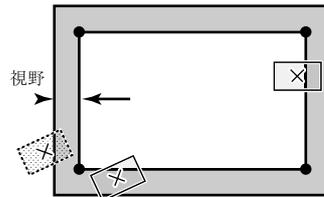
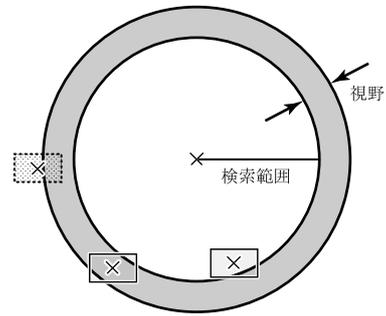


図15. 検索時における、視野分加算方式。内側の「円」「四角」領域は検索時に「半径」「範囲」で指定された領域である。小四角形は装置の1フレームの視野領域である。外側の円・四角部分は装置の視野分に相当する領域で検索時に利用者の判断で付加される。検索はフレームの中心座標（×印）で行われるので、内側領域に視野が全く含まれていない図中の一番左の破線網掛けで示したフレームは除外される。図中中央のフレームは視野分を付加しないで検索した場合は除外されるが、実際には視野の一部は指定された内側領域を映している。視野分を付加することで、このようなフレームの取りこぼしを防ぐことができる。

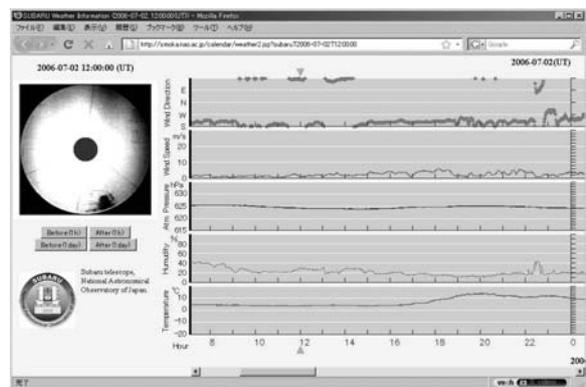


図16. すばる望遠鏡の気象モニター表示機能。左には雲モニター、右には上から風向、風速、気圧、湿度、気温を時間（UT、横軸）に対して表す。横軸の上下に示された三角印（実際の画面上ではオレンジ色）が指定した時刻を示す。JavaScriptを用いており、前後1時間、前後1日へ切り換えが可能である。

るが、これらの情報は今までSMOKAでは公開されておらず、公開が急務となっていた。

そこでSMOKAでは、AG (オートガイダー)^{*1)} ならびにキャットウォークに設置されたDIMM (Differential Image Motion Monitor)^{*1) 24)} が測定する、大気の影響により影響された星像の大きさを示すシーイング値 (それぞれAGシーイング・DIMMシーイング) を提供することにした。これらの情報を見やすくするため、各観測日のAGシーイングならびにDIMMシーイングの時間変化をグラフ化して表示させるようにした (図17)。

4.5 サミットログの表示と検索機能

サミットログとは、すばる望遠鏡の望遠鏡オペレータが書いている、毎日の観測の概略である。このレポートはその日の気象、用いた観測装置・望遠鏡とその運用状況を含んでいる。このレポートから、その日の観測が順調であったか否か、どのようなタイムスケジュールで観測が行われたのか、天候が急激に変化したかどうか、観測機器のエラーなど、観測データのFITSヘッダや環境データなどの数値からだけでは窺い知れない、観測当夜の状況を知ることができる。SMOKAでも気象・天候の状況とともに、このサミットログを、

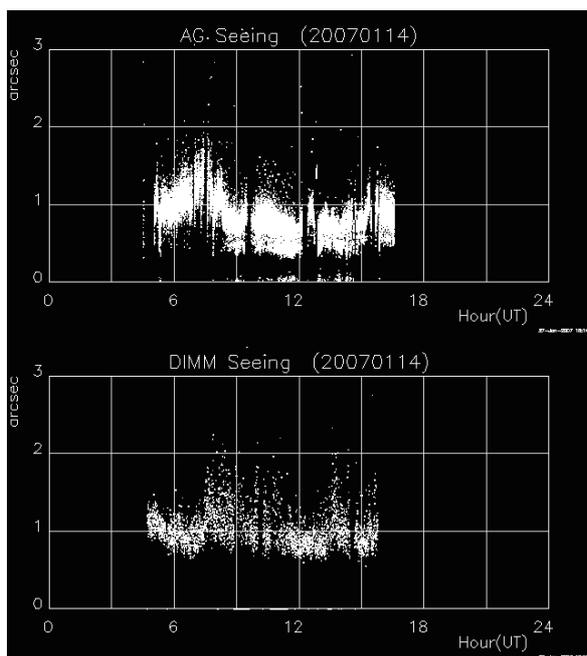


図17. AGシーイング, DIMMシーイングのグラフ。上にAGシーイング, 下にDIMMシーイングを横軸を時刻(UT)として1晩分を示す。

カレンダー検索からのリンクとして公開することとした。具体的には、カレンダー検索で目的の月のページを表示すると、サミットログのカラムがあり、サミットログが存在する日にはリンクが存在し、クリックすることで別ウィンドウで閲覧できるようになっている。

サミットログは、ハワイ観測所側で個人情報の削除などの加工を施した後に、ハワイ観測所の運用側からメールでSMOKAに送られてくる。そのメールのヘッダを除いた必要部分を切り分け、日毎のファイルとする。カレンダー検索画面を生成するプログラムは、ある日のサミットログのファイルが存在すればリンクを張るように設定されている。

また、利用者がサミットログからある状況を抜き出したい場合、例えば、降雪した日を知りたい、ある観測装置の稼働状況を連続的に知りたい、といった要求に応えるため、サミットログの文字列検索機能「サミットログ検索」を実装した。これは任意の文字列で検索し、その全文を閲覧できることを目的としたものである。

その詳細は以下の通りである。まず、観測夜毎のテキストファイルであるサミットログの全文を、SMOKAのMySQLデータベースに観測夜毎にテキスト形式にて登録する。検索は、文字列と期間を条件として、その都度データベーステーブルへ検索を行い結果を得る。検索結果は、該当する日付毎、行毎にリストとして表示される (図18 (a))。ここで左側の日付のボタンを選ぶと、対象となる日のサミットログ全文が表示される (図18 (b))。検索結果のリストならびに全文表示において、検索対象の文字列は赤字で表示される。サミットログは2004年11月から現在まで存在 (検索できるのは1年半前の2008年1月まで) し、全体で2.5MByte程度である。データベースを用いる検索としては量が少ないが、より複雑な条件で全文検索を行うなど、将来の拡張性のため、MySQLデータベースを用いることにした。現在サミットログ検索は検索文字列として、1単語のみにしか対応していない。今後の改良点として、検索文字列を複数語、さらにAND/OR/NOT検索へ対応することが挙げられる。

4.6 フィールドモニター画像の表示

すばる望遠鏡トッピングには、LI.CCDを用

いたフィールドカメラ*²⁾がある。このカメラは40×30度の視野を持ち、肉眼に近い感度で視野を捕らえている。この画像からは観測中の視野を横切る雲の様子など、観測天域のより詳細な状態が確認できる。環境データ・サミットログと並んで、このようなフィールドモニター画像は観測当夜の空の状況を知る手がかりとして重要である。そこでSMOKAでは、この画像を観測中の天候状態などの確認を目的として見やすい形で公開する

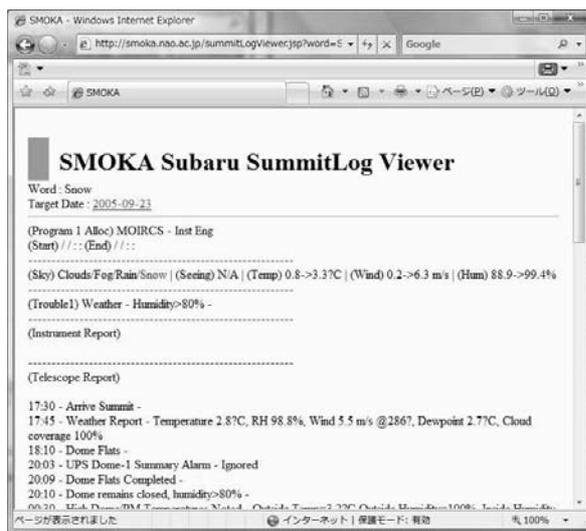
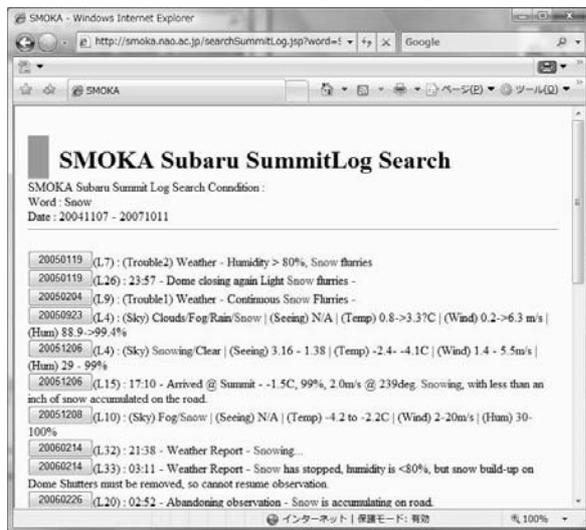


図18 (a), (b).

(a) サミットログ全文検索結果例。左側に年月日のボタンが並び、右側に対応する簡易結果（一行）を示す。検索に用いた文字列は実際の画面上には赤字で表示される。左側のボタンを押すと、全文が表示される。

(b) サミットログ表示例。上に、用いた文字列と日付を示す。日付からは、同日のデータ検索へとリンクが張られている。下にサミットログ全文が表示される。全文中の検索に用いた文字列は実際の画面上では赤字で示される。

こととした。

カメラから出力される画像は、パソコンによって取り込まれFITS形式で保存される。FITSのヘッダー部分は観測データ同様、公開時期に達したもののだけがMASTARSを経てSMOKAのデータベースに登録される。またFITSの画像部分は公開時期に達したもののだけがSMOKAに転送される。転送された画像は、Webブラウザで表示できるよう、SMOKAにおいてJPEG形式に変換される。

カレンダー検索からすばる望遠鏡を選んだ場合にリンクのカラムが表示され、利用者はフィールドモニター画像を検索できる。年月日毎のリンクを選択すると1晩分の毎正時の画像が右側に表示される。また各時刻のリンクを選ぶと毎分の画像が右側に表示される（図19）。

4.7 FITSヘッダ全文検索機能の改良

HISTORYやCOMMENTを含むFITSヘッダの全文を検索する機能としてSMOKAでは全文検索機能を実装した（論文4）。

この全文検索は全文検索システム *namazu* *³⁾を用いて実現されていた。しかし、*namazu*では検索の取りこぼしが存在し、その語を含む全てのフレームの検索結果を得ることができない問題があった。また、*mknmz*を用いた検索用インデックスファイルの作成は、更新の度に全てのフレームに対してインデックスを作り直さなければならないので、非常に長い時間（全装置に対して1ヶ月以上）が掛かり、日常的に運用するには不適當であった。このため、*namazu*による全文検索をやめ、データベース管理ソフトウェアMySQL

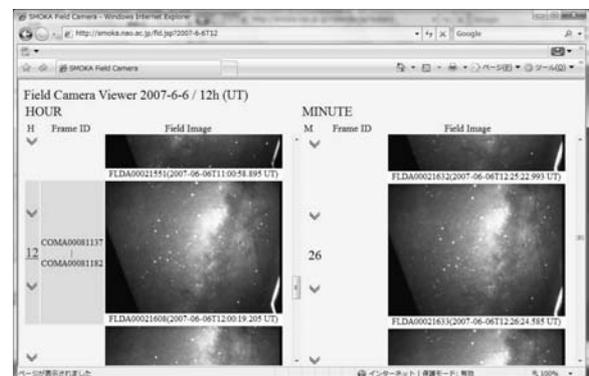


図19. フィールドモニター画像の表示例。画面左側に観測夜の1時間毎の画像を、右側に選択された時刻の1分毎の画像を縦スクロールで表示。実際の画面では、選択された時刻は背景色を赤（通常、水色）にして示される。

の全文検索機能を用いて、SMOKAのFITSヘッダ全文検索を実装することにした。現在、主要なデータベース管理ソフトウェアで、外部パッケージなどを必要としない標準機能として全文検索機能が実装されているものは、MySQLのみであるため、これを用いた。

MySQLによる全文検索は、データベースへのテーブル入力時にfulltextインデックスを張り、検索にはMATCH () ... AGAINSTシンタックスを用いて検索が行われる*4)。利用者が実際に検索する際は、検索対象装置を選択し検索キーワードを入力されている(図20)。検索にはAND, OR, NOT検索が利用できる。検索を行うと、SMOKAの他の検索と同様の結果ページが表示され、データ取得画面へと進むことができるようになっている。運用では、新たにSMOKAにアーカイブされたフレームのFITSヘッダをcronによって定期的にMySQLデータベースに入力する。対象となるFITSヘッダは、全ての観測装置合計で17.1 GByte(2009年4月30日現在)と膨大な量のテキストである。したがって、インデックステーブルの肥大化緩和と検索速度向上のため、観測装置毎にテーブルを分割している。全文検索用のインデックスの作成は、namazuを用いたシステムと違い、増えたフレームに対してだけおこなえばよく、所要時間が10000フレーム当たり数分から10分程度に短縮された。また、namazuによる検索で生じていた、検索結果の取りこぼしも解決された。

5. SMOKAの運用と利用状況

本章ではSMOKAの運用の現状と利用状況について論ずる。SMOKAが稼働している計算機システムの更新(リプレース)について5.1節で、SMOKAが公開しているデータ量について5.2節で、SMOKAの利用状況(データ請求量)について5.3節で、SMOKAから取得したデータを用いて得られた研究成果について5.4節でそれぞれ述べる。また、5.5節でSMOKAの運用に関する問題点、課題とその解決への方策について論ずる。

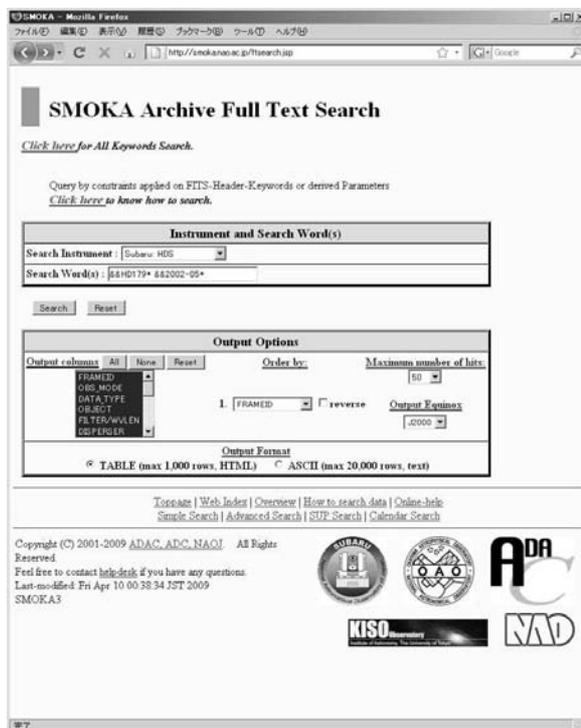


図20. 全文検索画面。FITSヘッダの全文検索ができる。検索は、FITSキーワードとその値だけでなく、COMMENT行や、HISTORY行等FITSヘッダ内の全てのテキストを対象としている。

5.1 計算機システムリプレース

SMOKAの運用の主要部分はレンタルの計算機システムの上で稼働しており、このシステムは契約の関係で数年に一度更新(リプレース)されている。現在の計算機システムは、2008年3月にリプレースされた。旧システムからの移行作業(設計など、準備作業を含む)は2007年11月から開始され、SMOKAの機能の大部分の移行が完了して安定運用に至った(SMOKA3.2版)のは2008年6月であった。この間、新システム上に応急移行して主要機能のみを動かした旧版を稼働させ、運用が途切れることがないようにした。

新システムの設計にあたっては、論文4(4.4節)で論じたサーバーの負荷分散をさらに進めて効率化をはかるとともに、運用に制限を加えていた磁気ディスク容量を大幅に増やすことに主眼をおいた。その結果構築された現在のSMOKA運用部分のシステム構成を図21に示す。磁気ディスク容量を150TB(ただし、ADSミラー、VizieRミラー、天体カタログなどの他の業務と共用)と大幅に増強したことを除く、旧計算機システム(論文4の図11参照)からの大きな改善点は、Web

*1 http://smoka.nao.ac.jp/help/env/acd_subaru.jsp

*2 http://www.naoj.org/staff/takato/IICCD_SkyMonitor.pdf

*3 <http://www.namazu.org/index.html>

*4 <http://dev.mysql.com/doc/refman/5.1/ja/fulltext-search.html>

サーバーとデータベースサーバーの分離である。旧システムでは、SMOKAはWebサーバーとデータベースサーバーを一台の計算機で運用してきたが、Webサーバー1台とデータベースサーバー3台で構成されるようにした。SMOKAのデータ検索は主にデータベースサーバー1（DBS1）で行われ、一部の検索がデータベースサーバー2（DBS2）で行われる。DBS2は、DBS1がダウンした場合に代替サーバーとなるように考慮されている。ハワイ観測所以外の観測所の観測データ（非公開分・未公開分を含む）の転送・管理は、専用のデータベースサーバー3（DBS3）を設け、公開部分であるDBS1、DBS2と運用体制上で隔離できるようにした。また、磁気テープライブラリは速度面と操作性の面で効率的運用になじまないため、新システムではバックアップの用途のみに用い、運用の主な流れからははずした。

また、計算機システムリプレースと、SMOKA新版化にともない、論文4（4.1節、4.2節）で論じた運用ドキュメント類の大規模な更新を行った。

5.2 SMOKAのデータ量

現在SMOKAで公開されているフレーム数、データ量を観測所毎に表2に示す。ここでは、観測データ本体（生データ）のみで、検索データベースに登録されているもの、すなわち、検索可能な

フレームのみを示している。表には合わせて観測後データが公開になるまでの標準期間（それまでは観測者にデータ占有が認められている）も示した。

どのデータ（フレーム）を公開し、どのデータ（フレーム）を公開しないかの判断はそれぞれの観測所が下すものであり、SMOKAは「できるだけ多くのデータを公開して欲しい」という原則を述べる以上の関与はしていない。試験観測などの観測装置が不安定な状態で得られた観測データを公開するかどうかは観測所によって異なっている。SMOKAでは、深刻な問題があるデータ、たとえば、致命的に壊れている、日時が不明（間違っている）、赤経赤緯が不明（間違っている）、といった場合は公開しない場合があるが、それ以外は公開時期に至ったすべてのデータを公開している。

5.3 SMOKAの利用状況

SMOKA運用開始2001年6月以来の月毎のデータ請求量を図22に示した。請求量は月によって大きく変動しているが、2003年9月のような特定の個人による大量取得は最近では見られなくなった。図には、SMOKAから取得したデータを用いた主要論文誌掲載論文の出版年月も示している（詳細は5.4節参照）。

SMOKAではデータ検索や早見画像の閲覧は自由にできるが、観測データを取得するには利用者登録を行ってもらうこと（毎年度末に更新）にしている。これは、個人による大量取得や商業利用を抑止することと、利用者を把握したいという観測所側の要請に基づくものである。磁気テープによるデータ送付を可能にするために利用者登録を

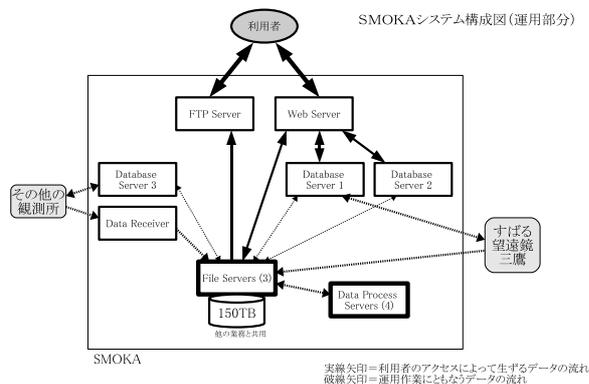


図21. SMOKAシステム構成図。SMOKAが稼働する計算機システムのうち、運用部分（開発・試験部分を除く）のシステム構成を示した図。四角一つが1台のハードウェア（サーバー）を示しているが、File Serverは3台、Data Process Serverは4台でそれぞれ構成されている。利用者のアクセスにともなって生ずるデータの流れ（通信）を実線矢印で、運用作業に際して生ずるデータの流れ（通信）の主なものを破線矢印で、それぞれ示している。

表2. SMOKAが公開している観測データ量。SMOKAが2009年7月16日現在で公開している（検索可能な）観測データのフレーム数とデータ量を観測所毎に示した。東広島天文台の観測データはSMOKAに蓄積されているが、最初のデータの公開時期（データ占有期間が切れる時期）に至っていないので、この表では0になっている。

観測所	データ占有期間	公開フレーム数	公開データ量
すばる望遠鏡	18ヶ月	1,344,287	14,598 GB
岡山天体物理観測所	2年	121,155	1,001 GB
木曾観測所	1年	144,710	819 GB
東工大 MITSuME	1年	702,096	1,551 GB
東広島天文台	18ヶ月	0	0

2009年7月16日現在

行ってもらおうという元々の目的は、ネットワークが発達し磁気テープによる請求がほとんどなくなった現在ではその意義が薄れている。2008年度末での登録利用者数は世界中で255名であった。

5.4 SMOKAによる研究成果

SMOKAが2001年に公開されてから2009年7月までの間に、SMOKAのデータを利用した研究成果は増加し続けてきた。我々が把握している、SMOKAから取得した観測データにもとづいて執筆され主要論文誌に掲載された査読論文は、2009年7月現在、53本を数える。1年間で生産される主要論文誌掲載の査読論文数は年々増加しており、特に、2007年に出版された論文が8本であったのに対して2008年に入ってから出版された論文数は17本と倍増した。以下は、これまでのSMOKAの利用傾向を出版された論文から分析した結果である。

図23に、論文執筆に利用されたデータがどの観測装置で観測されたものかを示す。まず、一見してSuprime-Camデータが公開から一定して最も多く活用されているデータであることは明らかである。しかし、近年になるにつれて、さまざまな装置を活用した論文が出てきていることがわかる。特に、論文数が飛躍的に増加した2008年を

みると、2007年の出版数に対する増分はほぼSuprime-Cam以外の装置を利用した論文数に相当している。また、すばる望遠鏡のみならず、東京大学木曾観測所の1kCCDおよび2kCCDを利用した論文も出版されていること、公開したばかりのMOIRCSデータを利用した論文が出版されていることをみても、様々な観測装置のデータが活用されるようになってきていることが読み取れる。また、撮像データと比べれば数は少ないが分光データを利用した論文も増えてきている。

図24 (a) に、これまでに出版された査読論文53本で、SMOKAのデータがどのように使われたかを示す。

使われ方は、(1) SMOKAデータのみによる研究、(2) SMOKAデータを主要データのの一つとした研究、(3) SMOKAデータを補助的なデータとして利用した研究、(4) SMOKAデータを較正データとして利用した研究、(5) SMOKAデータは参照のみ、の5つに大別した。それぞれ、(1) 4本、(2) 13本、(3) 32本、(4) 3本、そして(5) 1本である。出版された主要論文誌掲載の査読論文は、SMOKAのデータを他の観測と組み合わせて活用した研究が多くを占める一方で、中にはSMOKAデータセットを仮想サーベイ領域として活用した論文^{25), 26), 27)}なども出版されるようになった。また、他の観測データとの組み合わせも、視野や波長の違うデータ、取得時期の異なるデータを組み合わせた研究など多岐に渡ってきている。

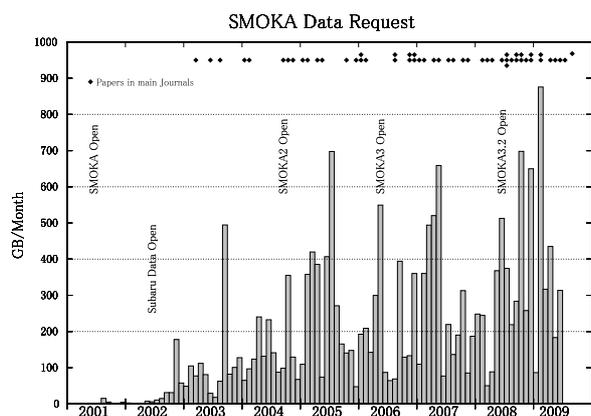


図22. SMOKAデータ請求量の推移。データ請求量(単位GB)をSMOKA稼働開始から月毎に示した図。上端にSMOKAから取得したデータを用いた主要論文誌掲載論文の出版年月を菱形で示してある。また、すばる望遠鏡のデータがはじめて公開された年月、SMOKAの各版が公開された年月も示してある。SMOKA3は論文4で述べた機能を実装して安定運用を実現した版、SMOKA3.2は計算機リプレースにともなう移行作業の主要部分が完了して安定運用に至った版である。

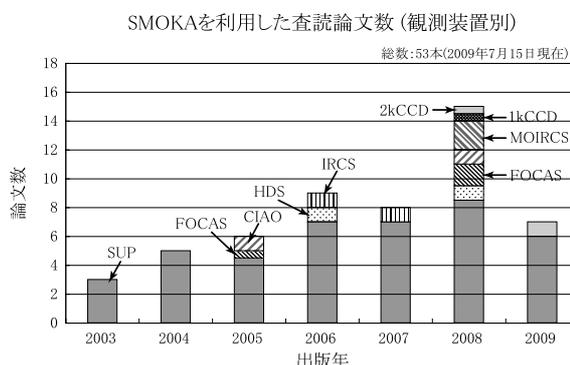


図23. SMOKAを利用し主要論文誌に掲載された主要論文誌掲載の査読論文数。観測装置別の積み上げ棒グラフとなっているので、出版年毎の総数は棒グラフの高さで示される。なお、1本の論文に複数装置のデータが利用されている場合は、合計で論文数が1となるように各装置の論文数を1/(装置数)とした。

図24 (b) には、分野ごとに分類したものを示した。分野は (A) 太陽系, (B) 恒星, (C) 銀河 (近傍・個別), (D) 銀河 (遠方), (E) 宇宙論, (F) その他, に大別した。それぞれ, (A) 3本, (B) 13本, (C) 7本, (D) 17本, (E) 11本, (F) 2本である。SMOKAを利用した研究の大半は遠方銀河や宇宙論的な研究であるが, 次いで恒星の論文も相当数あり, 多分野の研究者から利用されていることがうかがえる。

SMOKAが公開されてから8年を経て, 53本に及ぶ査読論文が生産されるまでになってきた。特に, 近年は様々な装置で取得されたデータをもとにした論文が増加している。また, データ蓄積量増加とともにSMOKAはデータベース天文学の重要な資源となってきた。今後も公開される観測データおよび観測装置が増えるに従って, 更なる研究成果の増加が期待される。

5.5 SMOKA運用の問題点と解決への方策

SMOKAは2001年6月から運用を続けており, 多くの天文学的成果の創出に貢献してきた。SMOKA運用の主要部分は国立天文台天文データセンターの5名の人員 (しばしば欠員あり) が, 各観測所の担当者などと連携を取りつつ担っている。5名の人員は日常の業務である, データベ-

スやデータの入力, データの加工処理, 利用者の登録や質問への対応, 観測所との連絡, システムの監視や利用統計算出, システム (ハードウェア, ソフトウェア) の障害への対応などに追われており, 新機能の開発に割ける時間・労力はあまり多くはない。また, 5名のうち4名は短時間契約職員で入れ代わりが多いので, 運用や開発のための教育や研修に費やされる時間・労力も少ない。しかし, 予算や人員枠の制限によって人員を簡単に増やすことはできず, 一部を外注に委ねるのも予算面で不可能である。このように厳しい状況下での運用や開発の効率を上げ, 省力化をはかるための方策について論ずることにする。

まず第一に, データ処理サーバー (群) の能力を向上させることである。統計値を求めたり, 2章で述べたHEALPixインデックスを算出したり, 早見画像を作成したり, 位置較正を施したり, などと運用の過程でデータ処理を行う場面が増えてきている。現在の4台のサーバー群では処理が滞ってしまう場合がしばしば生じており, 効率化をはかるためには処理能力の大幅な向上が望まれる。また, 新機能の開発・試験の際には試行錯誤を行わねばならないが, 処理能力がより高まればそのサイクルをより早めることができる。

第二に, 開発・試験用の環境 (ハードウェア, ソフトウェアの構成) を整えることである。運用環境にできるだけ近い開発・試験用環境があれば, 早期に問題点を把握でき, 効率的な開発・試験作業の実現が期待できる。また, 運用システムに組み込んだ際に運用停止や誤動作を招く危険性を低減することができ, 挑戦的な開発を進めやすくなる。

第三に, 障害発生時の代替環境 (ハードウェア, ソフトウェア) を整えることである。障害発生後に復旧させるのにかかる時間や労力はかなり大きく, それをできるだけ軽減することは効率化の面ではもちろん, 利用者の利便の面でも重要である。5.1節で述べたように, 現在のシステムでは, DBS2はDBS1がダウンした場合の代替を果たすものと位置づけられている。しかし, DBS2は負荷分散のために運用の一部を担っており, 通常稼働している状態になっている。ホスト名など個々のマシンに固有な設定を変更した上でなければDBS1の役割を果たすことはできず, その作業は現在手動で行わなければならない。よってこの作業の自動化あるいは省力化が望まれる。一方, 代

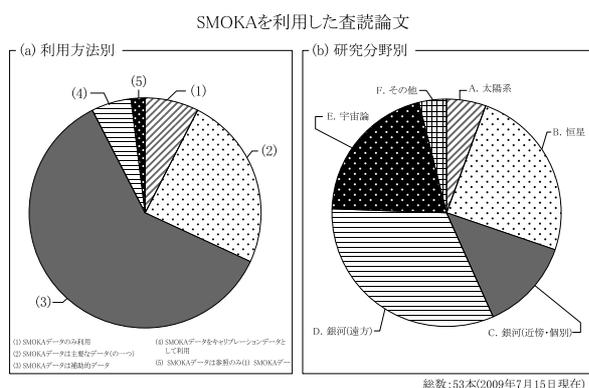


図24 (a), (b). SMOKAを利用した査読論文の分類。
 (a) 利用方法の分類。次の5つに大別した: (1) SMOKAデータのみ利用した研究, (2) SMOKAデータを主要なデータの一つとした研究, (3) SMOKAデータを補助的データとして利用した研究, (4) SMOKAデータを較正データとして利用した研究, (5) SMOKAデータを参照としてのみ利用した研究。
 (b) 研究分野での分類。分野は, A. 太陽系, B. 恒星, C. 銀河 (近傍・個別), D. 銀河 (遠方), E. 宇宙論, F. その他, である。いずれも2009年7月までに出版された53本について分類した。

替環境を用意するために多大な労力・時間をかけなくてはならないのであれば、運用の効率化という面ではむしろマイナスである。これらの兼ね合いを考慮しつつ代替環境の整備を進めることが重要である。

第四に、観測所側に働きかけて実現してもらうこととして、観測データの取得後できるだけ早い時期にデータチェックを行って観測にフィードバックをかけることがある。観測装置の不調や、制御計算機、通信経路の障害などによって観測データにエラー（データの欠損や付随情報の欠落）が生ずる場合がしばしばある。しかし、後日SMOKA側でエラーを発見しても、観測から時間が経っているために観測担当者の記憶が薄れてしまい、エラーの解明・修正に多くの手間と時間が掛かる。このため、少なくとも観測所側が観測後直ちにそれを検知して観測者に通知し、解明・修正した事象に関するログを共有できるようにすれば、SMOKA側の労力はかなり軽減される。また、SMOKAへのデータや付随情報の入力の際のエラー発生率が減り、自動化省力化へ一歩踏み出すことが可能になる。観測直後のデータチェックを行うことで、観測データの総合的な品質向上や観測効率の向上が図れ、データアーカイブとしての価値も高まる。ハワイ観測所のデータチェックシステムとしてはNAQATA²⁸⁾があるが、まだ機能的に不十分であり、高機能化・安定運用が待たれる。

上記以外にも細かな方策は考えられるが、重要なこの四点を実現できれば、運用や開発の効率の大幅な向上や省力化をはかることができよう。

6. まとめ

我々は、国立天文台ハワイ観測所すばる望遠鏡・岡山天体物理観測所188cm望遠鏡・東京大学木曾観測所105cmシュミット望遠鏡によって取得された観測データを公開するアーカイブシステムSMOKAの運用を行っている。また、運用だけでなく、SMOKAの利用者がより効率的に観測データを取得し、天文学的成果を得られるよう、開発・改良を続けており、現在までに様々な検索機能を開発してきた。本論文では新たに、空間分割法HEALPixを用いて、「何回以上観測された領域を検索する」といったことを可能にする重複領域検索を開発した。また、Suprime-Cam専用検索

に実装されているピンポイント検索をHEALPixを用いた検索に変更し、運用上の問題を解決した。また、データ請求システムについても機能を追加した。これは、検索結果をテキスト表示で得られるようにし、請求するフレームのリストを電子メールでSMOKAに送付することで大量のデータ請求を行うことができるようにするものである。

すばる望遠鏡のKyoto3DII, MOIRCSデータのSMOKAへの組み込みを行った。また、岡山と明野に設置されている東京工業大学のMITSuME望遠鏡の観測データ、広島大学東広島天文台かなた望遠鏡のHOWPolデータの組み込みも行った。今まで各観測所からのデータ転送方式が観測所毎に異なっていて、煩雑さの元となっていたが、新望遠鏡・新観測装置データの組み込みを機に、新方式に統一し、簡素化を図った。データの品質評価・取捨選択・解析の補助のための環境データ・シーイングモニター・サミットログの公開とその表示機能について開発を行い、より観測時の状況がわかるようになった。

SMOKAに蓄積されるデータも年々増加し、利用者も増加しており、計算機システムの負荷が高まっている。そのような需要に応えるべく、SMOKAを運用している計算機群は、2008年春の計算機システムの更新を機に、サーバーの分割など、負荷の分散をし、運用・開発環境を強化した。SMOKA利用者数の増加とともに、SMOKAから取得されたデータを用いた学術論文も50本を超え、その内容・使用用途・使用観測装置も多岐に渡ってきている。

天文学的成果を上げてきているSMOKAではあるが、各章で述べた通り、開発課題・運用上の問題がまだ山積しており、SMOKAによる天文学的成果をますます増やすためには、これらの開発・改善が継続的に求められている。

謝辞

本研究の遂行にあたっては、国立天文台天文データセンターDB/DAプロジェクトの計算機資源を活用した。

ハワイ観測所の高遠徳尚氏、浦口史寛氏、八木雅文氏、仲田史明氏、東京大学の安田直樹氏、日本スペースガード協会の奥村真一郎氏、国立科学博物館の洞口俊博氏の本研究への助力助言に感謝する。また、SMOKAの運用を連携して進めると

ともに、開発を支援応援して下さい、ハワイ観測所、岡山天体物理観測所、東京大学木曾観測所、東京工業大学河合研究室、広島大学東広島天文台の皆様にも感謝する。貴重なご意見を下さった利用者の方々に感謝する。

本研究では、NASAで開発されたHEALPixパッケージを用いた。本システムの運用にあたっては、仏CDSによって運用されるSIMBADデータベース、VizieRシステム、および、米NEDデータベースを利用している。有用な各種フリーソフトウェアを提供している開発者の方々にも感謝したい。

最後に、有益な助言を与えて下さった査読者にも感謝する。

本研究は、国立天文台天文データセンター開発経費の援助を得て行ったものである。

参考文献

- 1) 馬場肇, 安田直樹, 市川伸一, 八木雅文, 岩本信之, 高田唯史, 洞口俊博, 多賀正敏, 渡邊大, 奥村真一郎, 小澤友彦, 山本直孝, 濱部勝: すばる望遠鏡公開データアーカイブシステムの開発, 国立天文台報, **6**, 23–36 (2002).
- 2) 山本直孝, 野田祥代, 多賀正敏, 小澤友彦, 洞口俊博, 奥村真一郎, 古荘玲子, 馬場肇, 八木雅文, 安田直樹, 高田唯史, 市川伸一: すばる望遠鏡公開データアーカイブシステムの開発 2, 国立天文台報, **6**, 79–100 (2003).
- 3) 榎基宏, 多賀正敏, 小澤友彦, 野田祥代, 奥村真一郎, 吉野彰, 古荘玲子, 馬場肇, 洞口俊博, 高田唯史, 市川伸一: すばる望遠鏡公開データアーカイブシステムの開発 3, 国立天文台報, **7**, 57–84 (2004).
- 4) 出田誠, 榎基宏, 小澤友彦, 吉野彰, 仲田史明, 奥村慎一郎, 山本直孝, 古荘玲子, 矢治健太郎, 山田善彦, 八木雅文, 洞口俊博, 高田唯史, 市川伸一: すばる望遠鏡公開データアーカイブシステムの開発 4, 国立天文台報, **8**, 59–84 (2005).
- 5) Gwyn, S.D.J: MegaPipe: the MegaCam image stacking pipeline, arXiv:0811.1993 submitted (2009).
- 6) 原 正, 五島正光, 洞口俊博, 縣 秀彦, 矢治健太郎, 古荘玲子, 金光 理: Ia型超新星を使った銀河の距離測定の指導, 地学教育, **61**, 4, 113–122 (2008).
- 7) S. Miyazaki, Y. Komiyama, M. Sekiguchi, S. Okamura, M. Doi, H. Furusawa, M. Hamabe, K. Imi, M. Kimura, F. Nakata, N. Okada, M. Ouchi, K. Shimasaku, M. Yagi, N. Yasuda: Subaru Prime Focus Camera – Suprime-Cam, *PASJ*, **54**, 833 (2002).
- 8) K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, M. Bartelmann: HEALPix: A Framework for High-resolution Discretization and Fast Analysis of Data Distributed on the Sphere, *ApJ*, **622**, 759–771 (2005).
- 9) 吉野彰, 山田善彦, 仲田史明, 榎基宏, 高田唯史, 市川伸一: すばる望遠鏡 Suprime-Cam データの位置較正, 国立天文台報, **10**, 19–37 (2007).
- 10) P. Z. Kunszt, A. S. Szalay, A. R. Thakar: The Hierarchical Triangular Mesh, Proc. of the MPA/ESO/MPE workshop, ed. A. J. Banday, S. Zaroubi, M. Bartelmann, 631 (2001).
- 11) J. E. Gunn, G. R. Knapp: The Sloan Digital Sky Survey, *ASP Conf. Proc.* **43**, Sky Surveys: Protostars to Protogalaxies, ed B. T. Soifer (San Francisco: ASP), 267 (1993).
- 12) C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolte, N. Odegard, H. V. Peiris, L. Verde, J. L. Weiland: First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results, *ApJS*, **148**, 1 (2003).
- 13) S. Miyazaki, Y. Komiyama, H. Nakaya, Y. Doi, H. Furusawa, P. Gillingham, Y. Kamata, K. Takeshi, K. Nariai: HyperSuprime: project overview, *Proc. SPIE*, **6269**, 9 (2006).
- 14) R. J. Hanisch, A. Farris, E. W. Greisen, W. D. Pence, B. M. Schlesinger, P. J. Teuben, R. W. Thompson, A. III. Warnock: Definition of the Flexible Image Transport System (FITS), *A&Ap*, **376**, 359–380 (2001).
- 15) W. D. Pence, et al.: DIVISION XII / COMMISSION 5 / WORKING GROUP: FITS, Proc. IAU

- XXVI General Assembly, Transactions IAU, Volume XXVIB, Cambridge University Press, 218(2008).
- 16) H. Sugai, T. Hattori, A. Kawai, S. Ozaki, G. Kosugi, H. Ohtani, T. Hayashi, T. Ishigaki, M. Ishii, M. Sasaki, A. Shimono, Y. Okita, J. Sudo, N. Takeyama: Test observations of the Kyoto Tridimensional Spectrograph II at the University of Hawaii 88-in and Subaru Telescopes, *Proc. SPIE*, **5492**, 651–660 (2004).
- 17) H. Sugai, A. Kawai, T. Hattori, S. Ozaki, G. Kosugi, A. Shimono, Y. Okita: The Kyoto tridimensional spectrograph II, *New Astronomy Reviews*, **50**(4 – 5), 358–361 (2006).
- 18) T. Ichikawa, R. Suzuki, C. Tokoku, Y. K. Uchimoto, M. Konishi, T. Yoshikawa, T. Yamada, I. Tanaka, K. Omata, T. Nishimura: MOIRCS: multi-object infrared camera and spectrograph for SUBARU, *Proc. SPIE*, **6269**, 38(2006).
- 19) R. Suzuki, C. Tokoku, T. Ichikawa, Y. K. Uchimoto, M. Konishi, T. Yoshikawa, I. Tanaka, T. Yamada, K. Omata, T. Nishimura: Multi-Object Infrared Camera and Spectrograph(MOIRCS)for the Subaru Telescope I. Imaging, *PASJ*, **60**(6) 1347–1362 (2008).
- 20) T. Kotani, N. Kawai, K. Yanagisawa, J. Watanabe, M. Arimoto, H. Fukushima, T. Hattori, M. Inata, H. Izumiura, J. Kataoka, H. Koyano, K. Kubota, D. Kuroda, M. Mori, S. Nagayama, K. Ohta, T. Okada, K. Okita, R. Sato, Y. Serino, Y. Shimizu, T. Shimokawabe, M. Suzuki, H. Toda, T. Ushiyama, Y. Yatsu, A. Yoshida, M. Yoshida: MITSuME - Multicolor Imaging Telescopes for Survey and Monstrous Explosions, *Il Nuovo Cimento C*, **28**, 755 (2005).
- 21) M. Yoshida, K. Yanagisawa, Y. Shimizu, K. Okita, S. Nagayama, H. Toda, K. Ohta, N. Kawai: OAOWFC: Okayama Astrophysical Observatory NIR Wide-Field Camera, *AIP Conf. Proc.*, **1000**, 596–599 (2008).
- 22) M. Uemura, A. Arai, T. Uehara: GRB 061121: optical observation at the KANATA 1.5m telescope, *GCN Circ.*, **5828** (2006).
- 23) K. S. Kawabata, O. Nagae, S. Chiyonobu, H. Tanaka, H. Nakaya, M. Suzuki, Y. Kamata, S. Miyazaki, K. Hiragi, H. Miyamoto, M. Yamanaka, Akira Arai, T. Yamashita, M. Uemura, T. Ohsugi, M. Isogai, Y. Ishitobi, S. Sato: Wide-field one-shot optical polarimeter: HOWPol, *Proc. SPIE*, **7014**, 151 (2008).
- 24) F. Uruguchi, N. Takato, A. Miyashita, T. Usuda: The DIMM Station at Subaru Telescope, *Proc. SPIE*, **6267**, 21 (2006).
- 25) C. I. Fuentes, M. J. Holman: A SUBARU Archival Search for Faint Trans-Neptunian Objects, *AJ*, **136**, 83–97 (2008).
- 26) A. Yoshino, T. Ichikawa: Colors and Mass-to-Light Ratios of Bulges and Disks of Nearby Spiral Galaxies, *PASJ*, **60**, 493 – 520 (2008).
- 27) C. I. Fuentes, M. R. George, M. J. Holman: A Subaru Pencil-Beam Search for $m_R \sim 27$ Trans-Neptunian Bodies, *ApJ*, **696**, 91 (2009).
- 28) 仲田史明, 出田誠, 八木雅文, 榎基宏, 吉野彰, 山田善彦, 高田唯史, 市川伸一: すばる望遠鏡観測データ品質評価システムの開発, 国立天文台報, **8**, 43–57 (2005).