

光学赤外線観測データアーカイブシステム¹ SMOKA : 20年間の開発と運用、そして将来

中島 康, 樋口あや², 格和 純, 小野里宏樹, 野田祥代³,
古澤順子, 本間英智⁴, 高田唯史, 市川伸一

(2021 年 7 月 8 日受付 ; 2021 年 8 月 17 日受理)

Development and Operation of SMOKA: Past 20 Years and the Future

Yasushi NAKAJIMA, Aya HIGUCHI², Jun KAKUWA, Hiroki ONOZATO, Sachiyo NODA³,
Junko FURUSAWA, Hidetomo HOMMA⁴, Tadafumi TAKATA, Shin-ichi ICHIKAWA

Abstract

Twenty years have passed since we started the operation of SMOKA (Subaru-Mitaka-Okayama-Kiso Archive System). In this paper, we present an overview of SMOKA operations including its history, progress, various problems encountered and their solution, and statistics and achievements over the past twenty years. Furthermore, we present a discussion of issues to be considered for the next-generation system. We discuss the trends of data archive systems and instruments of astronomy in the world and make a proposal on the future astronomical data archive system.

概要

光学赤外線観測データアーカイブシステム SMOKA の運用を開始して20年が経過した。本論文では、SMOKA の運用の概要に加え、20年にわたる、SMOKA の進展の概要、運用と開発で直面した様々な問題とそれらへの対処、SMOKA の利用状況と成果について報告する。この中で、次世代の観測データアーカイブシステムに向けての検討課題についても論ずる。さらに、国内外の観測データアーカイブシステムや観測装置の進化の動向について論じ、将来の観測データアーカイブシステムへの提言を行う。

¹ 過去の論文1から6 (1章を参照) では、論文の日本語タイトルに「すばる望遠鏡公開データアーカイブシステム」を用いた。本論文でもすばる望遠鏡の観測データに重点を置いていることに変わりはないが、それ以外の望遠鏡の観測データが多くなったこと、「データアーカイブシステム」が一般的になって何のデータであるのかを明示する必要が出てきたこと、などによって日本語タイトルの表記を過去の論文から変更した。

² 東京電機大学 (Tokyo Denki University)

³ あいプラネット (Aiplanet & Company)

⁴ 日本科学未来館 (The National Museum of Emerging Science and Innovation)

1 はじめに

天体観測は、他の自然現象の観測と同様に、(厳密には) 同じ条件で遂行することのできない科学測定である。それぞれの観測データはある時刻のある方向の姿をある方法で捉えた唯一無二のものである。観測データは天文学研究を支える基盤であり、観測データを保存・管理して公開することにより次のことが実現される。

- (I) 観測立案者とは異なった発想・目的や較正・解析方法、観測立案者が対象としなかった天体や波長、あるいは、複数の時間・波長・天体のデータの組み合わせ等による新たな研究成果の創出
- (II) 第三者等による研究成果の検証
言うまでもなく、研究成果の再現性の検証は科学の根幹を成すものである。画期的・斬新な発見ほど第三者等による検証がなおさら必要である。
- (III) 観測計画の立案、研究テーマの発案、ソフトウェアの開発・試験、教材の開発、データ解析の実習、演習や自由研究、など様々な活用による研究・教育活動の推進

上記 (I), (II), (III) を目的とし、我々 (国立天文台天文データセンターを中心とするグループ) は光学赤外線観測データアーカイブシステム SMOKA¹⁾ (論文 [1-6]) を開発・運用 (運用については3章で定義する) し、世界中の研究者に「できるだけ多くの観測データをできるだけ早く」公開・提供している。SMOKAは、2001年6月から運用を続けており、本論文執筆時に運用開始から20年を超えた (以下、本論文執筆時現在とは2021年7月をさす)。

上記 (I) と (II) の目的のためには、提供する観測データは生データであることが必要であると考え、SMOKAでは生データを公開・提供している。ただし、SMOKAは、単に観測所からの観測データのコピーをサーバに配置してそれらを公開・提供しているだけではない。公開される観測データのあるべき姿、観測データアーカイブシステムとしてもつべき要件を検討のうえ、観測データを天文学で利用しやすいように、開発および運用を行っている。その背景には、観測データ (特にメタデータ) が観測所によって十分に整備されていないことが多い、メタデータに欠損や誤りがある (論文2) という事情があり、そのままでは、利用者が目的とする観測データを効率的に見つけるこ

とができない、較正・解析に必要な情報を得ることができないという問題が存在する。また、膨大な数の観測データがそれぞれ独立して並んでいるだけでは、利用者が迅速かつ的確に目的とする観測データを選択できないため、的確な絞り込みを行えるよう、利用者にとって使いやすい仕組みとともに、品質の情報も提供する必要がある。これらを、限られた予算・人員の中で、実現し継続していくための技術上および運用上の解は自明ではなく、観測データの数が増え、天文学の進歩に伴って、多様な観測装置が現れ、利用者の要求も変わる中、その時々で最適と思われる選択をしなければならない。

観測データを公開することは、観測所にとってはより多くの成果を得て観測所の意義を高め、観測者にとっては自らの研究成果が検証されることでより確固たるものになるため、観測所と観測者の両方に有益である。また、研究の公正さを確保する取り組みは世界的に行われており [7], 研究結果の検証の必要性は年々高まっていると思われる。これらのことから、観測データアーカイブシステムは当該コミュニティにおいて今後よりいっそう重要となると考えられる。

本論文では、運用開始から20年の節目に際してこれまで直面してきた様々な問題とそれらへの対応、さらにSMOKAの次世代の観測データアーカイブシステムに生かされるべき検討課題について論ずる。また、観測データアーカイブシステムの実現にとって重要なことは何か、そのために重視すべきことは具体的に何かについて、国内外の観測データアーカイブや観測装置の動向とともに論じ、本論文が、新しい観測データアーカイブシステムの構築においても参考となるようにした。

本論文では、第2章で、まず20年間の新規観測装置と新機能についてまとめるとともに、主に論文6以降のSMOKAの進展について述べる。第3章では、以降の議論の下地としてSMOKAの運用の概要について述べる。第4章ではSMOKAがこれまで直面してきた技術的な種々の問題とそれに対する選択、今後の観測データアーカイブシステムに向けての検討課題について論ずる。第5章ではSMOKAがこれまで対峙してきた運用面での諸問題について述べる。第6章ではSMOKAの利用状況と成果をまとめ、第7章では世界や観測装置の動向について論じたうえでSMOKAのこれまでの経験に基づいて今後の観測データアーカイブに向けての提言を行う。第8章では本論文のまとめを行う。

1) <https://smoka.nao.ac.jp/>

2 SMOKAの20年

SMOKAは、できるだけ多くの観測データをできるだけ早く公開するとともに、利用者の要望や新たな観測装置の登場などを受けて様々な機能を開発し運用に供してきた。

本論文執筆時現在、SMOKAでは七つの観測所の合計34の観測装置の観測データを公開している。それら観測装置の一覧を表1にまとめた（各観測装置で公開したフレーム数については表5を参照）。表1では、それぞれの観測装置のSMOKAにおける観測装置コード（観測装置名を表す3文字の英文字）およびフレームID（個々のフレームを特定するためのID）変換の有無を記載している。本論文では、光学赤外線天文学の慣習にならない、フレームとはFITSファイルを指すもの、またはFITSファイルの助数詞とする（一般用語としてのフレームは動画を構成する個々の静止画をさすが、FITSファイルにおいては、一つのFITSファイルに、拡張形式や3次元配列として、多くの画像データが含まれている場合でも、FITSファイル本体のことをフレームとよぶ）。各観測所で取得された観測データに元々付与されているフレームIDの書式は、観測所ごとに異なる。SMOKAでのフレームIDは、

- ・観測装置コードで始まること
- ・観測装置ごとに重複のない通し番号をもち、その通し番号は取得日時昇順であること
- ・複数フィルターでの同時観測が可能な観測装置においては（フィルターごとにフレームIDの系列をもつことが多いが）、フィルターの識別子を最後にもつこと

を前提としている。その前提に合わせるため、一部の観測装置について、SMOKAでは元々のフレームIDをSMOKA内部で取り扱いやすい書式に変換している。

SMOKAでは観測データの品質確認等のために環境データ（3.2節）も公開しており、各観測所の環境データについて表4にまとめた。

この20年の間に、観測データを公開対象に追加した新規観測装置、および、開発した新機能のうち主なものを表2にまとめた。新機能は、検索、データ請求、位置較正、その他に分類した。また、SMOKAに関係する観測所の主な動向も記した。ただし、論文[6]までの期間については、特に重要な事項のみ記載し、新規観測装置と新機能は省略した。それぞれの論文を参照されたい。

表2の新機能のうち、超新星検索機能、HSC専用検索（品質評価指標による検索）機能、HSC専用データ請求機能、の三つについて以下で詳しく述べる。

2.1 超新星検索機能

超新星は出現から100日程度の比較的短時間で減光して観測不可能となる天体である。超新星の減光の時間変化を示す光度曲線は超新星の性質に強く結びつく指標であり、観測不可能となるまでの短い期間になるべく多くの時刻の観測データを取得することが、超新星の研究で重要となる。

そこでSMOKAで公開している観測データから、超新星が写っている可能性のある観測データを効率的に検索し、利用者に提供する機能を2017年5月から公開した（2019年11月に閉鎖した。4.7節参照）。長期にわたる観測データの蓄積と、複数の観測施設における観測データを横断的に検索できることがSMOKAの大きな特徴の一つであるが、この特徴を生かし、既知の超新星が写っている可能性のある観測データをSMOKAのアーカイブデータから検索できるようにした。

我々が開発した検索機能は、既知の超新星の位置座標をもとに、SMOKAで公開している観測データのうち、その位置座標を含むものを検索する機能である（SMOKAで公開している観測データから新たな超新星を検出するものではない）。既知の超新星については、Transient Name Server（以下、TNS）²⁾とAsiago Supernova Catalog（以下、ASC）[8]³⁾でカタログ化されている天体を対象とし、それらに登録されている位置座標を参照した。SMOKAで公開されている観測データのうち、検索結果の信頼性に重点を置くため、位置較正が行われたうえでSMOKAデータベースのHEALPixテーブル（論文4）に登録されている観測データ（開発当時は、KCC、KCD、MTA、MTO）を検索対象とした。

以下、データベーステーブル（以下、テーブルとよぶ）と検索インターフェースについて述べる。テーブルと運用の流れおよび検索との関連について図1にまとめた。

(1) テーブル

超新星検索機能のために作成したテーブルについて説明する。作成したテーブルは、(a) 超新星カタログテーブル、(b) 日付テーブル、(c) フレーム数テーブルである。

2) <https://www.wis-tns.org/>

3) <https://heasarc.gsfc.nasa.gov/W3Browse/all/asiagosn.html>

表1：SMOKA で観測データを公開している観測装置一覧.

観測装置名 (正式名称)	観測装置 コード	フレーム ID 変換
すばる望遠鏡		
CAC ^a (Cassegrain Alignment Camera)	CAC	
CIAO ^a (Coronagraphic Imager with Adaptive Optics)	CIA	
COMICS ^a (Cooled Mid-Infrared Camera and Spectrometer)	COM	
CHARIS (Coronagraphic High Angular Resolution Imaging Spectrograph) [9][10][11]	CRS	
FOCAS ^a (Faint Object Camera and Spectrograph)	FCS	
FMOS ^a (Fiber Multi-Object Spectrograph)	FMS	
HDS ^a (High Dispersion Spectrograph)	HDS	
HiCIAO ^a (High-Contrast Coronagraphic Imager for Adaptive Optics)	HIC	
Hyper Suprime-Cam [12][13][14][15]	HSC	
IRCS ^a (Infrared Camera and Spectrograph)	IRC	
IRD (Infrared Doppler) [16][17][18]	IRD	
Kyoto-3DII ^a	K3D	
Suprime-Cam ^a	SUP	
OHS/CISO ^a (OH-Airglow Suppressor/ Cooled Infrared Spectrograph and Camera for OHS)	OHS	
MIRTOS ^a (Mid InfraRed Test Observation System)	MIR	
MOIRCS ^a (Multi-Object Infrared Camera and Spectrograph)	MCS	
SWIMS (Simultaneous-color Wide-field Infrared Multi-object Spectrograph) [19][20][21][22]	SWS	
MIMIZUKU (Mid-Infrared Multi-field Imager for gaZing at the UnKnown Universe) [23][24][25][26] [27][28]	MMZ	
岡山天体物理観測所188 cm 望遠鏡		
HIDES ^a (High Dispersion Echelle Spectrograph)	HID	有
ISLE ^a	ISL	
KOOLS ^a (Kyoto Okayama Optical Low-dispersion Spectrograph)	KLS	
OASIS ^a (Okayama Astrophysical System for Infrared imaging & Spectroscopy)	OAS	
SNG ^a (Spectro-NebularGraph)	CSD	
MuSCAT (Multi-color Simultaneous Camera for studying Atmospheres of Transiting planets) [29]	MCT	有
東京大学木曽観測所シュミット望遠鏡		
1kCCD [30]	KCC	
2kCCD [31]	KCD	
木曽シュミット乾板	KSP/KSQ ^b	有
KWFC (Kiso Wide Field Camera) [32]	KWF	
東広島天文台かなた望遠鏡		
HONIR (Hiroshima Opt/NIR Camera) [33][34]	HNR	有
HOWPol ^a (Hiroshima One-shot Wide-field Polarimeter)	HWP	有
東京工業大学 MITSuME 望遠鏡		
MITSuME 明野 [35]	MTA	有
MITSuME 岡山 [35]	MTO	有
西はりま天文台なゆた望遠鏡		
NIC (Nishiharima Infrared Camera) [36][37]	NIC	有
京都大学せいめい望遠鏡		
KOOLS-IFU (Kyoto Okayama Optical Low-dispersion Spectrograph with optical-fiber Integral Field Unit) [38]	KIF	

^a これら観測装置の参照論文については, 論文1, 2, 3, 4, 5, 6を参照のこと.^b KSQは, より精度の高い位置較正を施した観測データのフレームIDに使用される.
(https://ppplate.nao.ac.jp/KSQ_astrometric_calib.jsp)

表2：新規観測装置の観測データの公開開始と主な新機能の運用開始、SMOKAに係る観測所の動向の年表.

年月	新観測装置名	新検索機能	データ請求	位置校正	その他	観測所の動向
2001年06月				SMOKA 運用開始		
2001年08月					MOKA [39] を統合	
2002年06月				すばる望遠鏡共同利用観測データ公開開始		
2002年09月				論文1		
2003年09月				論文2		
2004年10月				論文3		
2005年10月				論文4		
2006年04月					改組 ^a	
2008年03月					SMOKA システム移行 ^b	
2009年10月				論文5		
2012年04月				論文6		
2012年07月		重複領域 ^c (更新)				
2013年03月					SMOKA システム移行 ^b	
2013年10月	KWF					
2014年02月	HSC (ENG ^d)					
2014年03月					全天モニタ画像システム ^e	
2014年06月		フレーム数 ^f				
2015年03月		KCD 移動天体 ^g				
2015年05月					全天モニタ画像 ^e 明野追加	
2015年06月		KCD ピンポイント ^h				
2015年07月		KCC 移動天体				
2015年08月		KCC ピンポイント				
2015年09月	HNR, HSC					
2016年01月				MTA, MTO		
2016年04月		MTA/MTO ピンポイント				
2016年06月		MTA 移動天体				
2016年07月		MTO 移動天体				
2017年03月						木曾共同利用終了
2017年04月			消去要請 ⁱ			
2017年05月		超新星				
2017年11月			MCS-MEF ^j			

表は次ページに続く

年月	新観測装置名	新検索機能	データ 請求	位置校正	その他	観測所の動向
2017年12月						岡山共同利用終了
2018年03月					SMOKA システム移行 ^b	
2018年04月						ハワイ観測所岡山分室発足
2018年08月						せいめい望遠鏡完成
2018年09月	MCT, CRS					
2019年09月					木曽写真乾板システム ^k	
2019年11月				木曽写真乾板		
2019年12月	NIC					
2020年01月	IRD					
2020年02月	SWS					
2020年07月	MMZ					
2020年09月		HSC 専用1	HSC 専用1			
2020年10月				木曽写真乾板 (更新)	全天モニタ 画像 ^e	木曽追加
2020年12月	KIF					
2021年03月	HIDES-F [40] ^m					
2021年04月				KWF		

^a 天文学データ解析計算センターが天文データセンターと天文シミュレーションプロジェクトに改組された.

^b 計算機システム更新 (リプレース) に伴い, SMOKA の移行を行った. リプレースについては4.8節で述べる.

^c 重複領域検索: 論文6参照

^d エンジンニアリングデータ

^e SMOKA の環境情報表示では縮小されている画像をそのまま公開するシステム [41] (<https://ozskymon.nao.ac.jp>).

^f フレーム数検索: 天体座標からフレーム数とフレーム ID を検索することができる.

^g 移動天体検索: 論文6参照

^h ピンポイント検索: 論文5参照

ⁱ 消去要請: データ 請求を行うと, そのデータはダウンロード用のFTP ファイルサーバに7日間保持されるが, この機能を使うと, 保持期限よりも前にそのデータを消去する請求を利用者が行うことができる.

^j MCS-MEF (Multi-Extension FITS) [42][43]: MOIRCS の FITS データのプライマリ部だけが, Multiple Extension 部もダウンロードするかをデータ 請求時に選択できる (表6参照).

^k 木曽写真乾板のデジタル化データの公開システム [44] (<https://pplate.nao.ac.jp>).

^l 2.2, 2.3節で述べる.

^m HIDES-F は, SMOKA ではデータベースを共用するなど HIDES と一体として運用している.

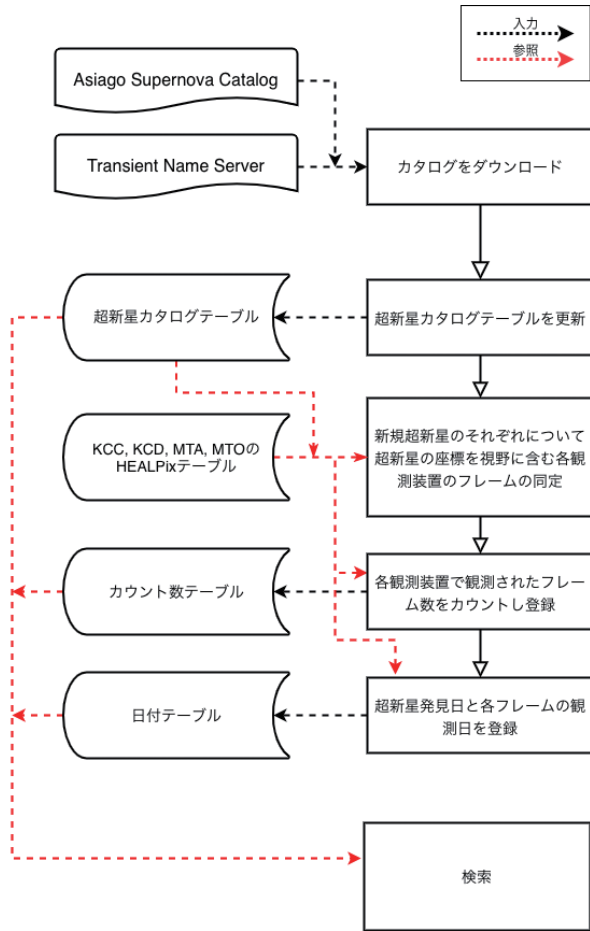


図1：超新星検索のためのテーブル更新の運用の流れとテーブルおよび検索の相関図。

(a) **超新星カタログテーブル** TNSおよびASCに登録されている超新星のカタログ情報を記録するためのテーブルである。TNSとASCから超新星のカタログ情報をCSVファイル形式でダウンロードし、その内容を読み込んでテーブルに記録する。TNSは日々更新されており、また新しいカラムが追加される場合もあるため、超新星カタログテーブルは更新のたびに消去と再作成を行う。ただし登録されている超新星の数が少ない（1万天体程度）ため、再作成にかかる時間は非常に短い。

このテーブルは、検索する超新星の位置座標を参照する際、絞り込み検索において超新星のタイプ（II型、Ia型、など）を分ける際にも使用する。このテーブルに記録されている各超新星のカタログ情報は、検索結果表示画面のリンクから確認できるようにした。

(b) **日付テーブル** 超新星検索機能は、超新星の発見日を基準として、観測データの観測日を絞り込む検索を提供する。この機能を実装するために、各超新星の位置座標を含んでいる観測データを検索し、その観測日と超新星の発見日を日付テーブルに事前に登録する。

これにより、絞り込み検索を行ってから結果を表示するまでの時間が大幅に短縮された。

(c) **フレーム数テーブル** 超新星検索機能では、各超新星の位置座標を含んでいる観測データがどれくらい存在するかすぐに分かるように、各超新星に対応した観測データのフレーム数の一覧を表示する。フレーム数テーブルには、各超新星に対して、各観測装置の観測データにおける超新星の位置座標を含むフレーム数を登録している。

(2) 検索インターフェース

超新星検索機能では、2段階で検索を行うように設計した。まず1段階目で超新星のタイプや検索したい観測装置などを選択し、その条件に当てはまる超新星のリストを表示する。2段階目でリストから超新星を選択することで、その超新星の位置座標を含んだ観測データのリストを検索できる仕組みとした。

1段階目の検索では、超新星のタイプと観測装置、観測データの条件を指定する。TNSおよびASCに登録されているタイプの中から超新星のタイプを選択でき、複数選択も可能とした。観測装置については、KCC、KCD、MTA、MTOを選択することができ、複数選択も可能とした。ただし、検索対象となる観測データは位置較正済みのものに限った。観測データの条件の指定では、SMOKAでアーカイブされている観測データに超新星の位置座標が含まれない場合にも「ない」という結果を表示するかどうかの選択と、超新星発見日からの前後の日数で観測日での絞り込みを可能とした。

2段階目の検索では、1段階目で指定した条件に該当する超新星のリストを表示し、その中から検索したい超新星を選択する。超新星のリストには検索の参考となるように、超新星の名前、タイプ、発見日、該当する観測データのフレーム数、超新星の最大光度を表示するようにした。

2段階目のリストから超新星を選択すると、その超新星の位置座標を含んだ観測データのリストが表示される。超新星のカタログ情報なども確認できるようにした。また外部リンクとしてThe Open Supernova Catalog [45]のリンク⁴⁾を貼り、このサイトに登録されている超新星の情報や、光度曲線を確認できるようにした。さらに、観測データ上の超新星の位置を画像ファイルとして確認できるようにした。

SMOKAの他の検索機能と同様に、観測データを選択した後にデータ請求画面またはメールによるデータ

4) <https://sne.space/>

請求で観測データを取得できるようにした。データ請求画面で観測データを請求した場合は、取得した観測データのフレームのバイナリテーブルエクステンション (BTE) に超新星の座標を書き込み、DS9などのソフトで観測データを開いた際に画像上で超新星の位置を確認できるようにした。

(3) 運用

定期的なデータベーステーブルの更新が必要である。特に TNS の超新星の情報は随時更新されていくため、それに合わせて SMOKA の超新星カタログテーブルを更新する必要がある。また日付テーブルやフレーム数テーブルも定期的に更新する必要がある (4.7節)。

2.2 HSC 専用検索機能

可視光広視野撮像装置の Suprime-Cam (SUP) は2002年から2016年ごろにかけて、SMOKA で扱う全ての観測装置の中でその観測データの請求量が最も多い装置であった。さらに、SMOKA から取得した観測データを用いた査読論文の数は装置別では SUP が最も多い (図11)。これは、SUP の観測データのアーカイブとしての利用価値が高いことを意味しているであろう。しかしながら、その観測データのフレーム総数は膨大であり (校正用フレームを除いた天体フレームだけでも約83万フレーム)、利用者が必要とする観測データを選び出すのはその時点の検索機能だけでは困難であった。そのため、SMOKA では SUP 専用高度検索機能を開発し2006年5月に公開した。本機能により、品質評

価指標 (シーイングサイズやバックグラウンドレベル) で絞り込み検索を行えるようにし、より品質の高いデータを選びすぐることを可能とした (論文4)。

Hyper Suprime-Cam (HSC) は SUP の後継の可視光広視野撮像装置であり、2017年以降では全ての観測装置の中でその観測データの請求量が最も多い装置である (図10)。HSC のデータアーカイブについては、SUP に続き多くの成果に繋がることが予想される。フレーム数については SUP 以上に膨大となるので、SUP と同様に、HSC についても専用高度検索機能の提供を2020年9月に始めた。

HSC 高度検索機能では、品質評価指標として、シーイングサイズ、星像の伸び、スカイレベル、限界等級を使用する。HSC 高度検索機能を提供するにあたり、SMOKA にアーカイブされている HSC の生データに対してこれら品質評価指標の測定を行った。以下、その測定について記述する。

HSC は104枚のサイエンス用 CCD、8枚のフォーカス用 CCD および4枚のオートガイダー用 CCD から成るモザイクカメラであり、そのうちオートガイダー用を除いた112枚が公開対象になっている。1度の露出で得られる公開対象の112枚のフレームの集合を、本論文ではショットと呼ぶ。品質評価量はショットごとにその代表値を求めた。ショットの視野中心付近の数フレームでの各品質評価量の平均値を求め、そのショットでの品質評価量の代表値とした。スカイレベルは中央の2フレーム (図2の破線囲み部分) を用い、シーイングサイズ、星像の伸び、限界等級については中央の6フレーム (図2の実線囲み部分) を用いた。

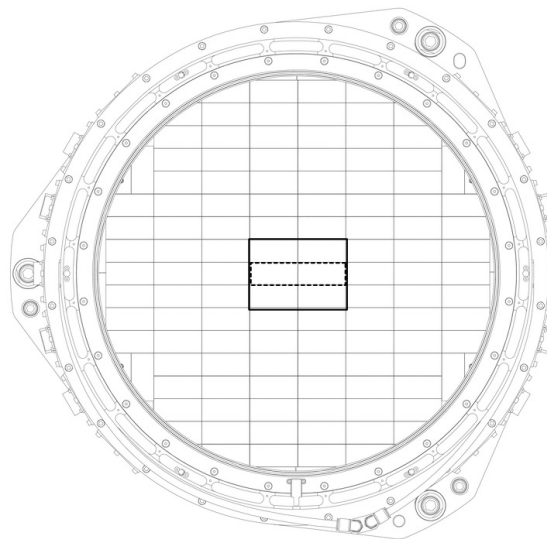


図2: HSC の焦点面の CCD の配置図 (Komiyama ら [13] の fig.1 より) に HSC データの品質評価量の測定に使用した CCD を示す。破線で囲まれた2枚の CCD をスカイレベルの算出に使い、実線で囲まれた6枚の CCD をシーイングサイズ、星像の伸び、限界等級の算出に使った。

(1) スカイレベル

HSCの各CCDは四つの読み出しチャンネルから成る⁵⁾。チャンネルごとに、オーバースキャン領域のADUカウント値のメジアンをバイアス値として求め、天体が写る領域のADUカウント値からバイアス値を引いたあとのメジアン値を求めた。2フレームの合計八つのチャンネルのメジアン値をそれぞれのチャンネルのゲイン (FITSヘッダに記載) から電子数に換算した値の平均値をそのショットのスカイレベル (電子数) の代表値とした。

(2) 星像, 限界等級

HSCの生データに対して、各チップごとに汎用天体検出アプリケーションソフト SExtractor [46] による天体の検出および星像測定 (シーイングサイズ, 扁平率) と開口測光を行った。星像測定には正しく測定された点光源のみを選出する必要があるが、SExtractorの検出した天体には広がった天体や宇宙線、サチュレートした星も含まれるため、それらを取り除く作業を行った。具体的には、SExtractorの出力のFLUX_MAX, FWHM, FLAGS, CLASS_STARの値を利用して、

- ・ $2000 < \text{FLUX_MAX} < 25000$ (FLUX_MAXはADUカウントのピーク値)
- ・ $1 < \text{FWHM (pixels)}$
- ・ $\text{FLAGS} = 0$
- ・ $\text{CLASS_STAR} \geq 0.9$

のものを正しく測光された点光源とした。SExtractorの出力のFWHMとELLIPTICITYがシーイングサイズと扁平率に対応する値である。6フレームで正しく測定された点光源のFWHMとELLIPTICITYのモード値をそのショットのシーイングサイズと扁平率の代表値とした。

各フレームでの正しく測定された点光源をPan-STARRS1カタログ[47]の星とマッチングを行い、測光零点を求めたうえで10シグマ限界等級 (AB等級) を求めた。6フレームでの平均値をそのショットの10シグマ限界等級とした。Pan-STARRS1カタログはHSCと同じSDSSフィルターシステムを使用しており、赤緯-30度以北の全天をカバーしているため、HSCの等級較正に最も適している。ただし、赤緯-30度以南の観測データ (本論文執筆時現在, 全体の1%程度) については限界等級を求めることができない。

2.3 HSC専用データ請求

データ請求量の多いHSCの観測データを利用者がより効率よく取得できるよう、HSC専用データ請求システムを開発し、2020年9月から公開している。

SMOKAは利用者からのデータ請求を受けた後に、FITSヘッダの修正、FITSファイルの圧縮等 (tarコマンドで固めることも含む場合に圧縮等と表現する)、ハッシュ値 (MD5) の計算を行ったうえで、ダウンロード専用FTPサーバにファイルを配置し、そのFTPサーバのURLを請求元の利用者にメールで連絡する (5.1節でも記述するように、観測データのFITSヘッダ情報は誤っていることがあり、SMOKAでは、利用者のデータ請求時に、その時点での修正情報に基づいてFITSヘッダの修正を行う。利用者はデータ請求時にファイルをtarまたはzipで固めるか、およびgzipまたはfpackで圧縮するかを選択する。SMOKAは、それら選択に応じて圧縮等を行う)。

利用者はそのメール連絡を受け取ってからではじめてデータをダウンロードできるが、データ容量が大きくファイル数が多いと請求してからメールを受け取るまでの待ち時間が長くなってしまうのが問題であった。

新たに開発したHSC専用データ請求システムでは、事前にFITSヘッダの修正、ショットごとにFITSファイル群をtar+bzip2またはfpack+tarを使用して圧縮等を行い、ハッシュ値 (ヘッダ修正後のFITSファイルと、ショット単位にまとめたファイルの両方について算出している) の計算を行ったものを用意しておくことで、利用者からのデータ請求を受けた後にすぐにダウンロード専用FTPサイトにファイルを配置できるようにしたため、利用者の待ち時間が非常に短くなった。SMOKAでデータ請求する方法はウェブページのフォームを用いる方法と定型フォーマットで書きこんだメールを送る方法の二種類を用意している (4.6節) が、現在のところ、HSC専用データ請求はメールを送る方法のみに対応している。

実際のデータ請求の例として、HSCの122ショット (13,664フレーム、フレームごとにfpackを適用) が請求された場合の所要時間を挙げる。旧来の請求システムでは、データ請求プロセスが走り始めてからFITSヘッダ修正が終了するまで27分間、そこからデータをfpackで圧縮する処理の終了まで25分間、そしてそこからハッシュ値計算の処理の終了まで43分間、合計で95分間であった。なお、圧縮にfpackではなく、gzipを選択をするとはるかに長い時間 (実測値では336分間; そのうち圧縮処理に280分間要している) がかかる。一方、今回開発の請求システム (フレームごとのfpackの後ショットごとのtarファイルを提供) では所要時間24分間であった。

⁵⁾ https://hsc.mtk.nao.ac.jp/pipedoc/pipedoc_7/hsc_info_j/index.html#hsc-info

今回開発の請求システムでは、磁気ディスクを節約するとともに、次世代観測データアーカイブシステムを考へて磁気テープライブラリのLTFS⁶⁾の活用をはかり、あらかじめ作成したショットごとにまとめたデータを磁気テープライブラリ上に置いた。また、磁気テープライブラリ上にあるデータを効率的に読み出すために、請求を磁気テープごとに分割し、それぞれの中での記録順に並べなおすソフトウェアを開発した。前述のデータ請求例を使って、データが磁気テープ上にある場合と磁気ディスク上にある場合の所要時間（データ本体を所定のFTP領域にコピーする時間のみ）を測定した結果、磁気テープ上の場合には20分間（請求データが含まれる磁気テープが、偶然、既にドライブにかかった状態の場合には、17分間であった）、磁気ディスク上の場合には10分間であった。

前述のように当初は、tar+bzip2およびfpack+tarを作成したが、おおむね、前者は1ショットあたり3分間、後者は4ショットで1分間の所要時間であることが判明し、ショット数が月あたり数百から2000であることから、前者（bzip2+tarにしても所要時間は同じ）は多大な時間がかかるため、現実的ではないと考えて作成をやめている。

HSCの場合、観測装置グループの尽力の結果として、観測データのFITSヘッダの修正が非常に少なく、そのため事前に修正、圧縮、ハッシュ値算出を用意しておくことが可能であり、このような専用データ請求システムが実現できた。同様に、FITSヘッダの修正が生じないような場合、例えば、しばらく前に観測が終了しているような観測装置について、今回の請求システムに組み入れることが検討に値する。

その場合、消費電力の少なさ、トータルコストの低さ、寿命の長さ⁷⁾から磁気テープライブラリ（LTFS）を活用していきたいが、現システムではテープドライブが2台しかないために読み出しを1多重にしており、その拡充（テープドライブの数を増やす）が必須である。しかしながら、今回開発した請求システムを2多重以上に対応させるためには、追加のソフトウェア開発が必要となる。また、磁気テープライブラリと磁気ディスクの読み出し速度を比較して考慮に入れ、観測データの量や請求頻度に応じて磁気ディスク上に置くか、磁気テープライブラリ上に置くか、をそれぞれで検討する必要がある。次世代の観測データアーカイブシステムにおいては、LTFSテープライブラリを重要要素として検討する価値があろう。

3 SMOKAの運用

4章以降で技術上および運用上の問題点等を議論する前に、本章で運用の概要について述べる。SMOKAは、単に、観測所からの観測データのコピーをサーバに配置してそれらを公開しているだけではない。観測データを天文学で利用しやすいように提供するために作り込まれた機能および日々の運用のうえに成り立っていることを本章で示す。

観測所で取得された観測データをSMOKAサーバ上に転送・公開し、SMOKA利用者に検索および請求機能を提供するのがSMOKAの主たる任務である。それらに加えて、環境データ（気象データと全天モニタ画像等）の公開と検索・請求機能の提供、観測データのクイックルック画像などの補助情報（4.5節）の提供も行っている。これらの任務を遂行するための各種機能を開発しSMOKAに組み込んでいる。それら機能を働かせるための一連の作業に加え、利用者からの質問への対応やSMOKAの利用状況の調査（6章）を本論文では運用とよぶ。

以下では、主たる運用である1. 観測データの公開、2. 環境データの公開、3. 利用者登録・管理の三つについて述べる。

「観測データ公開」の前段階である「新規観測装置の観測データを公開するための準備」も重要な作業であるが、これについては4.1節で述べる。本章では、運用の全貌を示すべく、観測データの流れについても述べる。

3.1 観測データの公開

各観測装置で取得された観測データが公開されるまでの運用の流れは以下のとおりである。

(0) 公開スケジュールの決定

SMOKAでは、それぞれの観測データについて、公開の可否、および、観測者占有期間が過ぎデータ公開が可能となる日付、の二つの属性が非常に重要である。本論文ではそれぞれ、PUBLICFLAGおよびPUBLICTIMEとよぶ。なお、PUBLICFLAGとPUBLICTIMEは観測所が決定し、観測所ごとに決められた手段でSMOKAに通知される（表3）。SMOKAは、通知されたPUBLICFLAGとPUBLICTIMEに従って、PUBLICFLAGが「可」のものだけについて、PUBLICTIME以降になるべく早く観測データを公開する。

すばる望遠鏡の場合は、観測装置別にどの日付のデータをいつ公開するか（以降、データ公開日とよぶ）をPUBLICTIMEに基づいてSMOKA運用要員が決

6) <https://lufs.jp/>

7) JEITA テープストレージ専門委員会: テープストレージ活用による省エネ貢献, https://home.jeita.or.jp/upload_file/20160303102225_Zf9QEChmUV.pdf

定する。公開作業は自動ではないため、PUBLICTIMEの日付に観測データが公開されるとは限らない。1日分の観測データ数が少ない場合には公開作業の効率が悪くなるため、数日分のデータを一つのデータ公開日にまとめることがある。また、PUBLICTIMEが休日の場合には公開日はPUBLICTIME後直近の営業日とする場合が多い。

すばる望遠鏡以外の場合は、(観測装置ごとではなく)観測所ごとに公開頻度(例えば二週間に一度)を定めてデータ公開日を設定し、データ公開日の時点で公開可能な未公開データを公開している(3.5節)。

(1) 観測所から SMOKA への観測データ転送

すばる望遠鏡の観測データはハワイ観測所ヒロオフィスのSTARS2および国立天文台三鷹キャンパスのMASTARS2というデータアーカイブシステムで保存される。そのため、すばる望遠鏡の観測データについては、データ保存が二重化されているためSMOKAへの転送を急ぐ必要がないこと、二重のデータアーカイブシステムであることによるデータ管理方法の特殊性があること、これらデータアーカイブシステムに観測後しばらくしてから入力されたりするなどの事情があることにより、すばる望遠鏡とそれ以外で、観測データ転送の時期と方法を異なるものとしている。

すばる望遠鏡の観測装置の場合、データ公開日にSMOKA運用要員が、処理ごとに分けられたプログラムを動作させるためのコマンドを手動で入力することで、観測所からSMOKAの公開データ領域(3.4節)へ観測データを転送する(例外として、HSCはcronを用いてPUBLICTIMEに自動的にこの転送処理を行う。データ量が多いため手動による公開作業開始までに転送処理を行っておくことで作業効率をあげている)。SMOKA運用要員は、観測データ転送前に、(ファイルステータス等の公開直前の修正・変更を考慮し、その情報の更新が最も早い)STARS2のデータベースで最新のPUBLICTIMEとPUBLICFLAGを確認したうえで、転送する観測データのリストを作成しMASTARS2のストレージからSMOKAへの転送を行う(STARS2、MASTARS2とSMOKAの観測データの整合性の確認が目的である)。SMOKA運用要員は転送処理の各段階(STARS2の確認とMASTARS2からの転送)の結果を確認し、ネットワーク等の不具合でSTARS2あるいはMASTARS2に接続できず処理が失敗する場合には、その時点で処理を中断し、失敗したフレームについては、不具合復旧後に一連の処理をやりなおす対応が必要となる。

すばる望遠鏡以外の観測装置の場合は、観測後すみ

やかに(表3)、SMOKAの自動処理によって、観測所サーバからSMOKAサーバへの観測データの転送を行う。転送後はまず、検索処理機能とは切り離れたファイルサーバの未公開データ領域(3.4節)に観測データを配置する。転送処理をいつ行うかは観測所によって異なる(4.2節)。ある回の転送処理が失敗しても次回に未転送分を転送するように転送処理機能を導入している(4.2節(4))。データ公開日よりも十分前に転送を行えるので、このような自動回復機能による運用の省力化が可能となっている。

(2) SMOKA への観測データの入力

全ての観測所の観測データに共通して、データ公開日にSMOKA運用要員が以下の作業を行うことで、公開作業対象の観測データに対してSMOKA上で検索およびデータ請求が可能となる。この一連の作業をSMOKAへの観測データ入力とよぶ。なお、以下の作業工程において、SMOKA運用要員がある特定の処理を行うプログラムを動作させるためのコマンドをコマンドラインに一つずつ入力のうえ、各工程の処理が正しく終了しているかを確認する必要がある。観測データのFITSヘッダで重要な情報が誤っていることなどにより、処理が正しく終わらない場合があり、人手による修正作業などの対応が必要となるからである(4.3節)。

公開作業において主要となる作業はテーブルへの情報入力である。公開作業での主要なテーブルは、登録用テーブル、検索用テーブル、ファイル管理テーブル、ヘッダ修正テーブルである。それぞれ観測装置ごとにひとつ、あるいはヘッダ修正テーブルについてはふたつ用意されている。図3にテーブルと観測データ公開手順工程の関連図を示す。

- ・登録用テーブル：観測データのFITSヘッダ情報を抽出して入力するテーブル。主に検索用テーブルを作成するための中間テーブルであるが、全項目検索機能の参照先としても利用する。
- ・検索用テーブル：SMOKAの検索機能の中核となるテーブルである。検索用テーブル等のSQL検索をウェブインターフェースで提供することで利用者に扱いやすくしたものがSMOKAの検索機能である。
- ・ファイル管理テーブル：観測データのFITSファイルや観測データ入力作業で作成されたファイルのファイルサーバでの所在や、観測データの公開ステータスを管理するテーブル。観測データ入力作業の各工程が正しく行われたかの確認にも利用する。

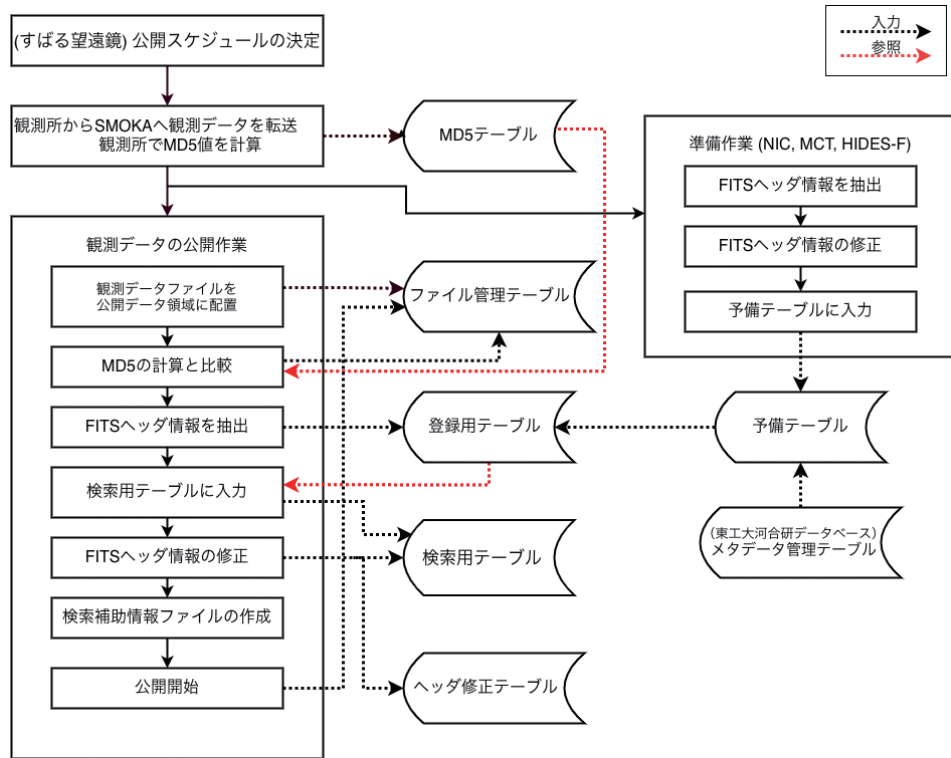


図3：観測データ公開作業工程。すばる望遠鏡の観測装置の場合のみ、公開スケジュールの決定の工程が加わる。NIC, MCT, HIDES-Fについては、データ公開日より前に、ヘッダ情報修正のための準備作業を行い、ヘッダ情報を予備テーブルに入力しておき、データ公開日に予備テーブルから登録用テーブルに情報を入力する。MITSuMEについては、転送後にMD5値を計算して東工大河合研データベースのものと比較し、FITSヘッダ情報抽出を行う代わりに東工大河合研データベースのメタデータ管理テーブルをコピーしているSMOKA上の予備テーブルから登録用テーブルにFITSヘッダ情報をコピーする。

- ・ヘッダ修正テーブル：観測データのFITSヘッダの修正情報を記録したテーブル。どの観測データのどのキーワードの値をどのように修正するか、および、どのキーワードを新しく追加するか2種類の（4.6節）のテーブルを用意している。

ここで、登録用テーブルと検索用テーブルには、FITSヘッダ情報が入力される（一部、FITSヘッダ情報から算出した値も含まれる）が、各カラムの値はFITSヘッダで定義されているデータ型および有効数字に合わせて定義しておく。その理由は次の二つである。(I) 科学データを取り扱っているので、検索結果で利用者が値の型や有効数字を把握できることが重要であるため。(II) 入力されてくる値が妥当なものであるかをSMOKA運用要員が確認できるようにするため（FITSヘッダに誤った値が入っている場合、テーブル入力時にそれを検知することができる）。

大部分の観測装置については、観測データ入力は以下の(a)～(g)の手順に従う。なお、NIC, MCT, HIDES-Fについては例外であり、以下の工程と異なる。FITSヘッダ情報の誤りを事前に検知し修正するために、データ公開日より前に、以下の作業工程の(c)

と(d)を行い、予備テーブルに修正情報などを入力しておく。そして、データ公開日には予備テーブルから登録用テーブルへコピーを行う(4.3節(1))。MTA, MTOについては、SMOKA側で東京工業大学河合研究室（以下、東工大河合研）データベースのコピーを持ち、入力作業ではそれを用いているので、以下の工程内容で異なる点がある。

(a) 観測データファイルの公開データ領域への配置とフレームID変換 すばる望遠鏡以外の場合は、データ公開日に観測データファイルをファイルサーバの未公開データ領域から公開データ領域に移動する(3.4節)。すばる望遠鏡の観測装置の観測データの場合は、上記(1)の転送処理の際に、既に公開データ領域に配置されている。この配置の後に、ファイル管理テーブルに観測データファイルの所在を入力する。このときフレームID変換もおこなう。

(b) 観測データファイルの確認 公開データ領域に配置した観測データファイルが壊れていないかを確認するため、各観測データファイルのハッシュ値(MD5)を計算し、観測所で計算したハッシュ値(図3のMD5

テーブルに記録されている）と比較を行う。

(c) FITSヘッダ情報を抽出 SMOKAで公開された観測データを検索する際に必要となるメタデータ（観測日時、赤経・赤緯など）を、観測データのFITSヘッダから抽出し、登録用テーブルに入力する。このとき必要に応じてフレームID変換も行う。MITSuMEの場合にはこれらメタデータは東工大河合研データベースのメタデータ管理テーブルでデータ型が正しく定義されているので、当該テーブルからSMOKA上の予備テーブルにコピーのうえ、登録用テーブルにコピーする（すばる望遠鏡の過去のSTARSでも、メタデータのデータ型を正しく定義していたので、同様に登録用テーブルにコピーしていた。現在のSTARS2ではメタデータは正しくデータ型が定義されず、全て文字列として定義されている）。さらに、観測データごとにFITSヘッダをHDI（HeaDer Information）ファイル（FITSヘッダのキーワードレコードの末尾に改行コードを加えてテキストファイルとして保存したもの）として保存し、その所在をファイル管理テーブルに入力する。

(d) FITSヘッダ情報を修正 観測所で作成されたFITSヘッダの情報は誤っていることがある。

- ・ STARS2の修正情報を記録したテーブルを参照する。
- ・ 観測所スタッフ等に問い合わせて修正情報を収集する。
- ・ 前後のフレームから推測を行う。

などの手段を用いて修正作業を行う。なお、ここでいう修正作業とは、公開データ領域に配置した観測データのFITSヘッダを直接書き換えることではない。登録用テーブル上で修正を行うことである。このとき同時に、修正用テーブルに修正情報を入力し、利用者がデータ請求をした際にダウンロードされるFITSファイルに修正情報が書き込まれるようにしている（4.6節）。

(e) 検索用テーブルに入力 観測データの検索に必要なカラムを登録用テーブルから（一部のカラムは算出して）検索用テーブルに入力する。

(f) 公開開始 各種テーブルへの入力工程が全て正常に終了したのちに、当該観測データについて、ファイル管理テーブルで公開ステータスを‘可’にし、さらに、カレンダー検索（4.4節）画面へのリンクを作成して、全ての検索メニューから検索および請求が可能な状態とする。

(g) 検索補助情報のファイルを作成 検索補助情報としてのQuick Look Image（以降、QLI）（論文1）、サムネイル（論文3）、ショットイメージ（論文4）など（4.5節）のファイルを作成し、所定のファイルサーバ領域に配置し、検索結果画面からそれらを参照できるようにする。

3.2 環境データの公開

SMOKAでは、各観測所の気象データ（気温、湿度、気圧、風速、風向）と全天モニタ画像（木曽観測所、岡山天体物理観測所、東広島天文台、MITSuME 明野観測所）等を、観測データの品質確認等の目的のために、環境データとして提供している。

(1) 気象データ

(a) 転送とテーブル入力 観測所によって気象データの転送方法は異なる。気象データファイルを転送する場合と、気象データファイルがないため観測データのFITSファイルから抽出する必要がある場合である。気象データファイルまたはFITSファイルの転送後、SMOKAのデータベースに気象データを入力する。木曽観測所以外の観測所については、転送はSMOKAの自動処理で行い、データベースへの入力も自動的に行うものと手動で行うもの（例えば、すばる望遠鏡用のCFHT（Canada-France-Hawaii Telescope）の環境データ（表3、表4））がある。木曽観測所については、自動処理の手法が確立していないため、転送とデータベースへの入力の両方をSMOKA運用要員が手動で行っている。

(b) 公開 各観測所の気象データについては、時系列グラフ化されたものを公開している。PUBLICTIMEに到達した日の気象データを自動的にグラフ化してウェブページで公開する（PUBLICTIMEよりも前に準備しているものもある）。

(2) 全天モニタ画像

全天モニタ画像のファイルを観測所からSMOKAサーバに毎日自動的に転送している。全天モニタ画像の公開場所は2か所あり、全天モニタ画像公開システムのウェブページと気象データのウェブページである。全天モニタ画像公開システムのウェブページでは全天モニタ画像取得日の翌日に自動的に公開する。気象データのウェブページでは、気象データと同様にPUBLICTIMEに到達すると、サイズが縮小された画像を自動的に公開する。

3.3 利用者登録・管理

SMOKAを用いてデータを請求しダウンロードするには利用者登録が必要である（検索やQLIの閲覧などには不要である）。利用者登録とその管理の流れを図4に示す。

利用希望者はSMOKAウェブページの利用者登録用フォームに必要事項（アカウント、氏名、電子メールアドレス、郵送物送付先、電話番号、所属、身分、使用目的）を記入して申し込みを行う。フォームに記入された情報は、自動的にデータベースの利用者管理テーブルに仮登録される。仮登録された情報をSMOKA運用要員が確認する。必要事項が記載されていれば、利用者管理テーブルの当該利用者のステータスを更新して利用者として本登録する。不足情報があれば利用希望者に電子メールにて問い合わせ、必要情報が揃い次第、利用者の本登録を行う。このとき同時に、利用希望者の電子メールアドレスを迷惑メール防止のためのメールフィルターに登録し、利用者からの電子メールを受信できるようにする。

利用者登録は年度ごとに更新を行う。利用者のメールアドレスや所属などは変更されることがあり、それらの情報の正しさを保つために、一定の頻度で利用者

登録を更新する方法をSMOKAでは採用している。年度末には利用者に対して周知確認メールを送信し、新年度に登録更新を受け付ける。

3.4 観測データの流れ

発生源の観測装置から、各観測所のファイルサーバを経由して、SMOKAから利用者に渡るまでの観測データの流れを図5に示す。

各観測所のファイルサーバからSMOKAへの観測データの流れはすばる望遠鏡とその他の望遠鏡で異なる。すばる望遠鏡では、PUBLICTIME以降にSMOKAへ転送される。その他の望遠鏡では、観測日（DATE-OBS）の翌日あるいは数日後にSMOKAへ転送される。

すばる望遠鏡の観測データはハワイ観測所と国立天文台三鷹キャンパスに存在する二つのファイルサーバで保存されているが、他の望遠鏡の場合は必ずしもそうではない。過去に観測所のサーバでデータを損失したことがあったが、SMOKAのサーバから復旧したことがあった。観測後すみやかにSMOKAへ転送することで、災害および事故等によるデータ損失のリスクが軽減される。

SMOKAにはPUBLICTIMEを迎える前の観測データを保存する未公開データ用ファイルサーバ領域と、PUBLICTIMEを迎えた観測データを保存する公開データ用ファイルサーバ領域が存在し、両者を分離して保存している。PUBLICTIMEを迎える前の観測データの公開を未然に防ぐしくみである。すばる望遠鏡以外の観測データは観測後数日程度でSMOKAに転送されるため、一旦、ファイルサーバの未公開データ領域に配置され、データ公開日に公開データ領域に移動される。すばる望遠鏡の観測データはPUBLICTIMEを過ぎてから転送されるので、直接、ファイルサーバの公開データ領域に配置される。

利用者から請求を受けた観測データを、Data Copy Manager（DCM）を使ってヘッダ修正、ファイル圧縮等、ハッシュ値算出などの処理を行った後にFTPサーバに配置する。利用者はこのFTPサーバから観測データをダウンロードする（4.6節）。

3.5 定期的な運用作業

SMOKAの運用の中で観測データの入力には重要な作業の一つである。前述の通りこの作業は、（観測装置によって手順が異なるが少なくとも核心部分は）cronで自動起動されるのではなく、人間が手動で起動している。大部分の観測装置ではPUBLICTIMEに達した後には随時、あるいは、ある程度まとめて作業を行うが、

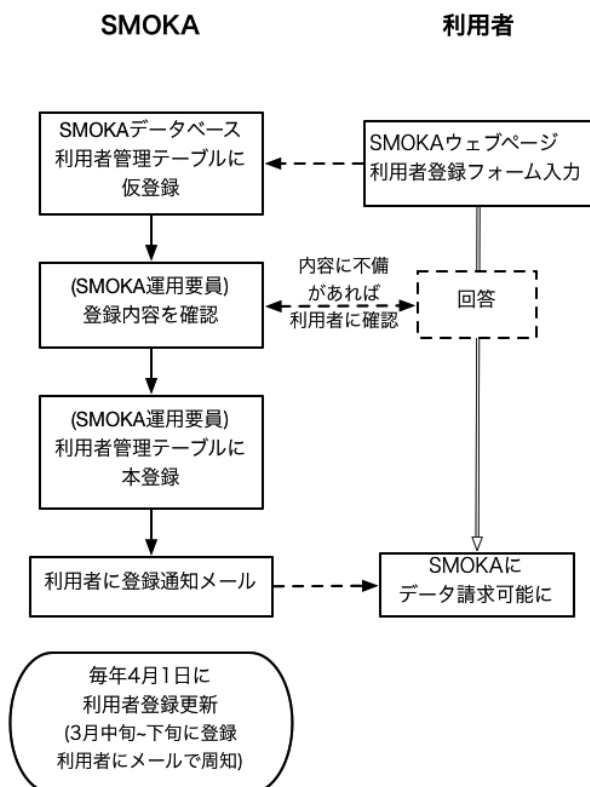


図4：利用者登録・管理の流れ。実線はSMOKAでの処理の流れ。破線は情報の流れ。白抜き線は利用者の作業および状態の流れ。

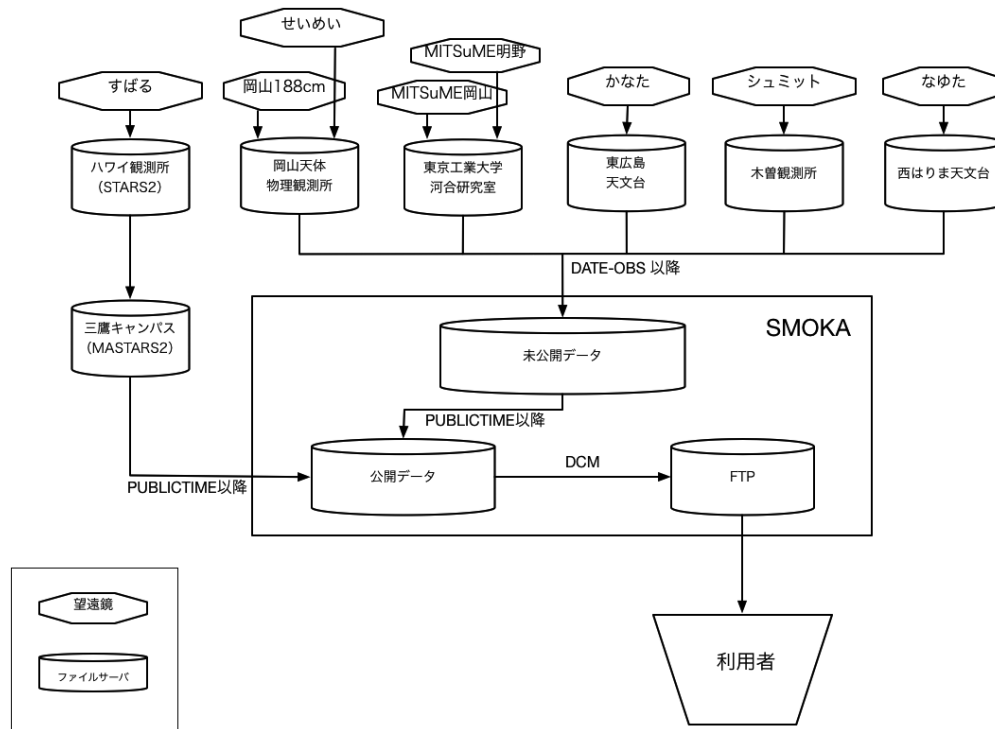


図5：観測データの流れ。

一部の観測装置では定期的に作業を行っている。それらを含めて人間が手動で起動する定期的作業を以下に挙げる。

- ・毎月1日
 - SMOKAの前月1か月の利用状況の統計調査
 - 木曾観測所と岡山天体物理観測所には前月のSMOKAデータ請求記録を送付
 - MITSuMEの前月公開分データを一括公開
 - 西はりま天文台NICの前月分公開データを一括公開
 - 計算機のシステム領域のバックアップ（磁気テープライブラリへの書き込み）
 - 木曾観測所の気象データの転送とデータベース入力
- ・毎週月曜日
 - STARS2のPUBLICTIME, PUBLICFLAG, ヘッダ情報修正記録をSMOKAテーブルに同期
- ・毎年4月1日
 - SMOKA利用者登録の更新

毎月1日に作業が集中しているのは、作業中に確認するログなどの目視性（確認容易度）が高いためである。

るのが第一であるが、集中させることによってやり忘れを防止するためでもある。しかし、SMOKA運用要員の負担を軽減すべく、それぞれの作業の手順を検討しなおす必要があるかもしれない。

4 SMOKAの技術面での諸問題

本章ではSMOKAがこれまでに直面してきた技術的な諸問題、そしてそれに対して行ってきた様々な技術的選択とその評価について論ずる。また、今後のSMOKAシステムに向けて改善すべき点、考慮すべき点を含めた検討課題についても論ずる。

SMOKAの運用開始からの20年間に様々な技術的課題を検討して解決してきた。それら解決の手法は、次世代観測データアーカイブシステムの検討の際にも大いに参考となるであろう。

4.1 新規観測装置の観測データ組み込み

新規観測装置の観測データの組み込みは、その観測装置が、初めから観測データの公開を前提として開発されているかどうかでその手間と難易度が大きく異なる。すばる望遠鏡の観測装置の場合、長年にわたる観測データ公開の歴史の中でSTARS2およびMASTARS2

を用いたデータ転送とデータ管理の仕組みが、観測装置によらず統一的に確立されており、今や新規観測装置の観測データに対して新たに仕組みを構築する必要がないが、それ以外の望遠鏡の観測装置の場合はそれらの仕組みを新規に構築する必要があることが多い。

しかしながら、すばる望遠鏡の新規観測装置の場合であっても以下で述べるような様々な問題に直面することがあり、SMOKAのデータベースや検索システムなどの改修に加えて以下のようないくつかの追加措置が生じることがある。

- (I) 特に観測装置の運用初期において、FITSヘッダのキーワードのデータ型の定義が固定されていないことがあり、観測データによってデータ型や文字数や桁数が変化し、テーブルのカラム値の定義を修正する必要が生じることがある。
- (II) 特に観測装置の運用初期において、FITSヘッダのキーワードが変更あるいは追加されることがあり、テーブルのカラムを修正あるいは追加する必要が生じることがある。

(1) 実際に発生した問題とその対処

以下では、最近組み込んだ観測装置の中から、せいめい望遠鏡のKIFとすばる望遠鏡のIRDを選び、その具体的な手順と発生した問題とその対処（対処の規模の指標として、対処に要した時間も含む）について述べる。これら2装置の観測データは比較的順調にSMOKAに組み込むことができたものであるが、それでも以下のような問題が発生した。

(a) KIF

- ・観測データのFITSヘッダのサンプルを調査し、キーワードや値に問題がないかを確認
FITS規約に違反した例は見つからなかったが、有効数字が正確でない値があった。KIFグループはPythonの汎用天文ライブラリの`astropy`⁸⁾ [48][49]を利用してFITSヘッダの編集を行っているが、現状の`astropy`にはFITSヘッダの値に浮動小数点数を書き込む際に、任意の有効数字の桁数を指定できずに丸め誤差を含む数値が書き込まれてしまう不具合が存在している。具体的には`astropy.io.fits.card._format_float`という関数に不具合が存在する。このことが今回の原因であった。SMOKA側が当該関数の修正プ

ログラムを1日程度の所要時間で開発し、KIFグループに提供して問題を解決した。

- ・観測所からSMOKAに観測データを転送するための仕組みを構築
観測所側とSMOKA側それぞれが観測データ転送用サーバを設置し、ファイヤウォール設定で双方のサーバが必要なプロトコルで通信できるようにし、観測データを保存するディレクトリ・ファイル構造とユーザ権限を設定した。そのうえで、観測所側は観測データが観測の翌日に観測データ転送用サーバに配置されるよう設定した。SMOKA側は、`rsync`を用いた転送用スクリプトを作成し、`cron`により自動的に転送用スクリプトを起動し観測所からSMOKAサーバの未公開データ領域（図5）へ観測データの転送を行うよう設定した。
- ・観測所側による公開データの管理の仕組みの構築
現在の観測データが公開可能であるかのリスト（以後、公開データリスト）を観測所側が作成し更新を行うこととした。公開データリストのファイルは観測所側のサーバに配置され、SMOKAはそのファイルを随時読み出すことができるよう設定した。SMOKAでは公開データリストに基づいて観測データの公開を行う。
- ・SMOKAでの公開手順の構築
公開データリストに記載されているもののうち未公開の観測データは新規公開対象の観測データである。新規公開対象の観測データに対して、ハッシュ値の比較確認、公開データ領域への移動（図5）、FITSヘッダからのメタデータ抽出、検索に必要なメタデータのエラー確認および検索必須項目の算出、検索テーブルへのメタデータ等の入力、検索補助情報の作成、検索画面への追加、といった公開手順をKIFに特化してSMOKA側が構築した。既存の他観測所の公開手順に使用されているプログラムを再利用することで効率的に構築を行うことができた。ただし、後述するように、公開から5か月後に観測装置間のFITSヘッダ内の書式等の違いによるエラーが生じた（4.3節）。
- ・観測データ入力時に生じた問題点
— ある日付の一部の観測データについて、FITSヘッダのDATE-OBSの値が誤っており修正が必要であった。観測装置での観測データ生成処理におけるバグが原因であった。2020年8月に発生し、FITSヘッダの修正に2時間程度を要した。

8) <https://www.astropy.org/>

— FITS ヘッダに想定しない桁数の値が入っており、テーブルへの入力に失敗した。具体的には、登録用テーブルにて望遠鏡の天頂角 z の $\sec z$ のデータ型の定義において 10^6 以上の数値を入力しようするとエラーが発生するように定義していた。そこに、 $1.6\text{E}+16$ の値をいれようとして失敗した。天頂角は最大で 90 度の値をとるので、 $\sec z$ の桁数が非常に大きくなりうることは想定できることであるので、SMOKA での処理の初期設定ミスである。データ型を桁数の指定なしで定義する、あるいは 10^6 以上の値をとる場合に例外処理を行うようにすべきであった。2020年12月に発生し、データベース入力プログラムに例外処理を追加して再処理を行った。所要時間は1日程度であった。

(b) **IRD** すばる望遠鏡の観測装置の場合、STARS2での公開データの管理とMASTARS2からの観測データ転送の手法が、既に他の観測装置で確立しているため、「観測所側による公開データの管理の仕組みの構築」および「観測所からSMOKAに観測データを転送するための仕組みの構築」は不要であった。

- ・観測データのFITSヘッダのサンプルを調査し、キーワードや値に問題がないかを確認
ハワイ観測所側が提示したサンプルデータには問題がなかった。

- ・SMOKAへのデータ入力作業手順の構築
SMOKAへのデータ入力作業手順をIRDに特化して構築した。すばる望遠鏡の既存の観測装置のデータ入力作業手順に使用されているプログラムを再利用することで効率的に構築を行うことができた。

- ・生じた問題点
初期の観測ではFITSヘッダのキーワードや値のデータ型が安定しておらず、登録用テーブルへの入力時にしばしばエラーが発生した。具体的には、キーワードのスペルミスや、RA2000, DEC2000の値に $\pm\text{DD:MM:SS.SS}$ のコロンが入っていなかった（おそらくFITSヘッダ生成時の単純なコーディングミスであろう）、FITSヘッダのキーワードの値のデータ型が観測日によって（極端なケースでは同じ観測日でも時間帯によって）異なったなどである。数十夜にわたる期間の観測データを集め、すばる望遠鏡のためのFITS規約（次の「(2) FITS規約に関する活動」を参照）で定義されていないキーワードが使われてないか、全てのキーワードの値がすばる望遠鏡のためのFITS規約での定義と違わないか、違うケースでは数

値の桁や文字列の長さがどの程度異なりうるかを調査したうえで、登録用テーブルのデータ型を定義する、および、例外処理をテーブル入力プログラムに組み込むという対応をとった。2019年12月、2020年7月、2020年11月実施のデータ入力時に問題が発覚し、対応に要した時間は合計で数日であった。

このように、新規観測装置の観測データの組み込みは、それぞれの観測データの特徴や特質を十分に把握したうえで臨機応変の姿勢で取り組む必要がある。今後、様々な観測装置が開発されていき、新たな望遠鏡・観測所も対象になっていくかもしれない。それらの観測データを可能な限り公開する任務を負ったSMOKAには、このような組み込み作業から逃れる術はなく、観測装置開発のできるだけ初期の段階から関与（例えば、FITSヘッダに起因する問題についてはなるべくデータ公開までに排除するよう協力）していく努力を続けることが唯一の方策であろう。

(2) FITS規約に関する活動

前述したように、観測データをアーカイブして公開するためには、観測データのFITSフォーマット（特にFITSヘッダ）の整備が非常に重要である。FITSの規約に従っていることはもちろんであるが、データを解析する際に必要となる望遠鏡や観測装置の設定・状態などの情報が漏れなく的確にヘッダに記載されていることが強く望まれる。

SMOKAでは、新しい観測装置の観測データを組み入れるごとに、その観測データのFITSフォーマットを検討し、観測所、あるいは、観測装置グループに助言して整備を促してきた。可能である場合は、観測装置の開発段階、あるいは、試験段階から積極的な関与を行ってきた。また、普段よりFITSへの理解と関心を高める活動も進めてきた。その一環としてFITS日本委員会やハワイ観測所の協力の下、「FITSの手引き」⁹⁾をおおむね3年に一度（IAU総会の開催年度に行う。FITSの規約が改訂される場合が多いからである）発行している。「FITSの手引き」にはFITSの標準規約に関する解説に加えて、すばる望遠鏡のためのFITS規約（標準規約では規定されていないハワイ観測所独自の事項などに関する細則）を掲載し、新規観測装置の開発時に参考になることを目的のひとつとしている¹⁰⁾。すばる望遠鏡のためのFITS規約については、「FITSの手引き」だけではなく、WEBページ¹¹⁾に、観測装置毎の

9) <http://jaipn.nao.ac.jp/jfits/>

10) 各地の観測所を訪れた際に実験室や制御室などに「FITSの手引き」が置かれているのを見ると普及の効果が現れているように思われる。

11) <https://www.naoj.org/observing/fits/header/regulation/>

FITS ヘッダキーワード辞書等が掲載されている。

また、ハワイ観測所では2020年11月にすばる FITS 委員会が設立され、観測データの自己記述性と標準化をさらに高めつつ、天文学研究の着実な発展を目指している。今後これまでにない新たな発想に基づく観測装置が登場することが期待され、FITS ヘッダに含まれるメタ情報も複雑になることが予想されるため、FITS についての理解を深めるだけでなく、目指す天文学研究の実現に必要なデータ構造を検討し、必要ならば FITS の標準規約の改訂や追加を IAU に提案していくことも視野に入れておく必要がある。

4.2 観測所から SMOKA への観測データの転送

観測データの管理方法や計算機環境、ネットワーク環境はそれぞれの観測所あるいは観測装置で異なっており、観測データ転送に関する様々な事項も観測所によって異なる。表3に各観測所からの観測データ転送の詳細をまとめた。すばる望遠鏡の観測データ管理システムは2013年3月に STARS/MASTARS から STARS2/MASTARS2へと更新された。観測データ転送に関する事項もかなり変更されているため、表3では STARS/MASTARS と STARS2/MASTARS2 を別々に記載した。

なお、岡山天体物理観測所と木曽観測所からは本論文執筆時現在は観測データの転送を行っていない。

(1) 観測データの転送の時期

すばる望遠鏡の観測装置については、PUBLICTIME の後に SMOKA サーバへ観測データが転送される。一方、他の観測所については、観測所側の SMOKA 転送用サーバ (MITSuME の場合は SMOKA 転送専用のサーバではない本来のデータ保管用サーバが転送に利用される) にデータが配置された後に、転送用の自動処理スクリプトが起動して三鷹の SMOKA サーバにデータが転送される。つまり、すばる望遠鏡以外の観測所の観測データは、PUBLICTIME よりもずいぶん前に SMOKA サーバに転送される。

(2) SMOKA が参照する観測所のデータベース

すばる望遠鏡および MITSuME の観測装置については、SMOKA が参照できる観測データのデータベースを持ち、

(I) PUBLICTIME と PUBLICFLAG

(II) FITS ファイルのハッシュ値 (MD5)

(III) 観測所サーバ上での観測データファイルの所在

(IV) 観測データのメタデータ

を管理している。(I) を SMOKA から参照することで、どのデータをいつダウンロードして公開すればよいかを知ることができる。(II) は SMOKA に転送した FITS ファイルのハッシュ値と比較してデータ破損がないかの確認に用いる。(III) は観測所から SMOKA へ観測データを転送する際に利用する。(IV) については、STARS と MITSuME ではメタデータにデータ型が正しく定義されているため、FITS ヘッダからメタデータを抽出する代わりに観測所のデータベースから登録用テーブルにコピーをすることができる (3.1節 (2) (c))。(STARS2はメタデータを保持するが、全て文字列として定義されており、データ型の情報が含まれないので SMOKA の登録用テーブルへのコピーには利用できない)

せいめい望遠鏡の KIF についても、データベースではないが、公開可能な観測データのリストを観測所側のサーバで管理しており、SMOKA から参照してどのデータをいつ公開すればよいかを知ることができる (KIF の観測データファイルは、観測日から一意に決まるディレクトリに配置されているため、ファイルの所在を参照する必要はない)。これら観測所の観測データについては、観測所が PUBLICTIME と PUBLICFLAG の管理をしているので、SMOKA の負担が少ないという SMOKA 側にとっての利点がある。

上記以外の観測所は観測所側で SMOKA が参照できるデータベースをもたない。PUBLICTIME は、基本的に、観測所ごとに制定されたルール (観測日からの日数など) で決定される。PUBLICFLAG の可否は個別に観測所より SMOKA へメールなどで連絡される場合が多いが、西はりま天文台の場合は観測所作成のルールに従う。なお、東広島天文台、岡山天体物理観測所、木曽観測所については SMOKA データベースの公開管理テーブルへの観測所側からのアクセスが可能であり、観測所側で個々の観測データの PUBLICTIME と PUBLICFLAG を編集することができる。ただし、現状の運用ではその機能が使われていない。

(3) 観測所側の作業

すばる望遠鏡の場合は、MASTARS2に観測データを転送し、STARS2に各観測データの PUBLICTIME, PUBLICFLAG の値および、FITS ファイルの MD5 値、ファイルの所在、メタデータ (データ型なし) を入力する。ただし、これら PUBLICTIME, PUBLICFLAG の値が誤っていることがある (5.2節)。MITSuME

表3：各観測所から SMOKA へのデータ転送詳細一覧。

	すばる望遠鏡		東京工業大学 MITSuME 望遠鏡	京都大学岡山天文台 せいめい望遠鏡	岡山天体物理観測所 188cm 望遠鏡	東京大学木曽観測所 シュミット望遠鏡	東広島天文台 かなた望遠鏡	西はりま天文台 なめた望遠鏡	
	STARS/MASTARS	STARS2/MASTARS2							
観測データの 転送の時期	データ公開日		観測翌日		観測所の SMOKA サーバ に転送された翌日		観測翌日		毎月10日と25日
SMOKA が参照する観 測所のデータベースの 有無と参照する情報	・ PUBLICFLAG と PUBLICTIME ・ ファイル MD5値 ・ ファイル所在 ・ メタデータ（データ型 有）	・ PUBLICFLAG と PUBLICTIME ・ ファイル MD5値 ・ ファイル所在 ・ メタデータ（データ型 無）	・ PUBLICFLAG と PUBLICTIME ・ ファイル MD5 値 ・ ファイル所在 ・ メタデータ（データ型 有）		なし				
PUBLICTIME と PUBLICFLAG の管理と 周知	観測所データベース		公開データリストの テキストファイル		基本的に観測者占有期間のルールに従う				公開 / 非公開は西はりま 天文台作成のドキュメン トに従う
観測所側の作業	・ STARS または STARS2に新規データの情報を入力 （修正も含む） ・ MASTARS または MASTARS2に観測データを転送		観測所データベースを管 理		観測所のSMOKA 転送用サーバに観測データを配置する				
SMOKA 側の作業	・ 新規データの PUBLICTIME を確認し転送スケ ジュールを決定 ・ MASTARS または MASTARS2からデータを転送		SMOKA サーバに存在し ないデータを観測所サー バで同定し、SMOKA サーバに転送、MD5 値 の計算。		SMOKA サーバに存在し ないデータを観測所サー バで同定し、SMOKA サーバに転送、MD5 値 の計算。		データを SMOKA サーバ に転送、MD5 値の計算。		
修正情報の流れ	・ STARS運用要員がデータベースに記録し SMOKA 側から定期的に確認 ・ 観測所担当者およびサポートサイエンティストと SMOKA とのメール連絡		観測所担当者と SMOKA 運用要員のメール連絡						定期的な修正情報は西は りま天文台作成のドキュ メントに従う
気象データ・ 全天スカイモニタ	・ すばる望遠鏡の TSC ステータスログ（気象データ を含む）を MASTARS または MASTARS2 より毎週 転送 ・ すばる望遠鏡の全天スカイモニタ画像を毎日転送 （2006年10月まで） ・ CFHTから、気象データを毎月転送、全天モニタを 毎日転送		・ 気象データ：観測デー タのヘッダから読み込 む ・ 全天スカイモニタ：毎 日転送		毎日転送		・ 気象データ：担当者間 でデータの授受とデー タベース読み出しと登 録 ・ 全天スカイモニタ：毎 日転送		・ 気象データ：観測デー タの FITS ヘッダから読 み出す ・ 全天スカイモニタ：提 供なし

の場合は、観測所データベースに各観測データの PUBLICTIME, PUBLICFLAG の値およびファイル所在、メタデータ（データ型あり）を入力する。その他の観測所の場合、各観測所の SMOKA 用サーバの所定のディレクトリに観測データを転送する。

(4) SMOKA 側の作業

すばる望遠鏡の観測装置については、どの観測データがいつ PUBLICTIME を迎えるかをあらかじめ情報を得ておき、公開スケジュールをたてる。PUBLICTIME は変更になることがあるので、その変更情報を定期的に確認する。データ公開日を迎えたら、ハワイ観測所の STARS2 で PUBLICTIME を確認したうえで、MASTARS2 から観測データを転送する。基本的にこれら確認および転送の作業は SMOKA 運用要員による手動で行う（HSC については自動スクリプトを使用する）。

他の観測所の観測装置については、自動スクリプトを用いて観測所サーバから観測データの転送を行う。このとき、SMOKA 側のサーバのデータベースの観測データ転送管理テーブルを参照して、過去に転送済みの観測データと重複しないように転送リストを作成する（例外として、NIC については観測データ転送管理テーブルを参照した転送リストの作成を行わず、rsync を利用することで重複のない転送を行っている）。同時に、観測所側のサーバで観測データファイルの MD5 値を計算し、その結果を SMOKA に転送する（ここで、MITSuME は例外であり、観測データ転送は SMOKA 側の予備テーブルで管理されており、観測データ転送直後に MD5 値を計算し東工大河合研データベースのものと比較する）。

(5) 修正情報の流れ

すばる望遠鏡の場合は STARS2 運用担当者によって、修正情報が STARS2 に入力される。STARS2 上の修正情報は定期的に SMOKA から確認する。これ以外に、ハワイ観測所の SMOKA 担当者やサポートサイエンティストと SMOKA 運用要因が直接連絡をとり修正情報を取得する。

他の観測所では、観測所の SMOKA 担当者と SMOKA 運用要員とのメール連絡を通して修正情報を取得する。ただし、西はりま天文台の NIC の場合は、突発的な事項以外は、観測所作成のドキュメントに基づいて修正を行う。

(6) 観測所から SMOKA への環境データの転送

MITSuME の MTA, MTO および西はりま天文台の NIC の場合、気象データは FITS ヘッダに記載されているものを用い、すばる望遠鏡の場合は、TSC（望遠鏡制御統括計算機）ステータスログに記載されているため、FITS ヘッダや TSC ステータスログから気象データを読み出す機能を用意している。

各観測所の環境データの一覧を表4に示す。

(7) PUBLICTIME と PUBLICFLAG を SMOKA に周知する仕組み

すばる望遠鏡および MITSuME については、それら観測所のデータベースを SMOKA から参照でき、そのデータベースを用いて PUBLICTIME と PUBLICFLAG の情報を SMOKA と共有している。せいめい望遠鏡は公開可能な観測データファイルのリストをテキストファイルで作成して毎日更新を行い、SMOKA と情報を共有している。

上記以外の観測所は観測所側で SMOKA が参照できるデータベース等をもたない。上記 (2) で述べたようにデータベース以外の手法で連絡などが行われる。

観測データは FITS という制約の下でそれぞれの特性を発揮すべく、多彩な構造やヘッダをもっている。観測所における観測データの管理も様々な事情によってそれぞれの方法や特徴をもっている。SMOKA の立場では、これから加わる観測所を含めた全ての観測所で統一された方法であるのが理想であり、観測所側の努力と SMOKA との協議によってある程度の統一化ははかれているもののそれには限界がある。今後も観測所側の事情を尊重しつつ、観測所と SMOKA の連携と協議を重ねていくしかないであろう。一方、7.2 節で述べるように膨大な観測データが生成されるようになってきているため、インターネットでは観測データを転送しきれない場合も既に想定されており、磁気ディスクユニットでの観測データ郵送など、本節で述べたこれまでのやり方では収まらない方法を取らざるを得ないかもしれない。これらを検討するのが今後の課題である。

4.3 SMOKA への観測データの入力

観測所から転送された観測データは PUBLICTIME に達した日、あるいは、作業や SMOKA 運用要員の都合によってそれ以降のできるだけ早い日に SMOKA への入力が行われる。入力の作業に際しても様々な問題が生じる場合がしばしばあるが、以下で述べるように

問題への対処が臨機応変で、かつ、できるだけ早期に容易に行えるような作業手順を構築している。

観測データの SMOKA への入力作業は、大別して

- ・ 検索や管理に必要となる情報のテーブルへの入力
- ・ 観測データの FITS ファイルのファイルサーバの所定の領域への配置
- ・ 検索や品質評価に役立つ情報やファイルの算出・作成および所定のデータベースやファイルサーバ領域への配置

から成っている。具体的な手順は4.2節で述べた観測所ごとの事情や仕組みによって異なっており、また、観測装置によっても少々の差異があるが、どの場合でも10–20程度の工程に分けてそれらを順に人手によって確認しつつ作業している。例えば、かなた望遠鏡 HNR では14, すばる望遠鏡 HDS では20, そしてすばる望遠鏡 IRD では16の工程に分けている。

全ての観測データにおいて、その FITS ヘッダに必要な情報が正しく記録されていれば、この工程は全て自動処理に任せることができる。実際、HSC では、観測装置グループの尽力により、FITS ヘッダ情報が整備されており、半分近くの工程を自動処理に任せている。また、残りの半分の工程でも2019年8月19日–2021年6月の期間にはエラーが発生したことはなく、結果的に自動処理に任せることができた状態であった（2019年8月18日に SMOKA への入力工程でエラーが発生したのが最後であるが、その時は FITS ヘッダが原因ではなく、MASTARS2 と SMOKA 間のネットワーク不調が原因であった）。

しかし、その他の装置では FITS ヘッダが十分に整備されておらず、FITS ヘッダ内で重要な情報が誤って記載されていることがあるため、人手による確認および修正作業が必要であり、全ての工程で、処理ごとに分けられたプログラムを動作させるためのコマンドを手動で入力するようにしている。1回の観測データ入力作業には、フレーム数や修正の有無により、典型的には数10分–1時間程度の時間がかかる。一方、手動でのコマンド入力のため誤入力をしてしまった例も存在する。また、手動でのコマンド入力を行うので PUBLICTIME が休日の場合には公開が数日遅れる場合がある。手動でのコマンド入力にはそのようなデメリットもあるが、全装置の全工程の自動処理化は困難であると判断し、現状では導入していない。

自動処理の工程でトラブルが発生した場合には処理の手戻り（問題発生以降の段階の結果を作業前に戻し、

そこからやり直す）を行う必要があるが、手戻りの作業は煩雑で時間がかかり、それによる誤りも懸念される。過去に自動処理を試行したが、自動処理と手戻り作業の得失を比較した場合に必ずしも自動処理が得ではないと結論し、現在の手動入力的方式で定着している。発生する可能性があるトラブルを念入りに洗い出し、それに対応する手順をあらかじめ組み込んでおく等の自動処理の開発には時間と労力がかかり、人手による入力の時間と労力を補って余ることはないと判断している。

1回の入力作業中にひとつでもトラブルが発生したケースは、最近5年間では、すばる望遠鏡の場合、HSC の場合は約0.8%であり、HSC 以外の観測装置の場合は平均約4%である。以下はトラブルの内容の例である。

- ・ FITS ヘッダのキーワードの値が誤っている（値が入っていない、定義されている型と異なる、想定していた範囲を逸脱しているなど）

この場合、登録用テーブルおよび検索用テーブルへの登録ができない（従って、当該フレームを公開できない）、または誤った情報を公開してしまうことになるため、SMOKA 運用要員が修正作業を行う。具体的には、観測所へ問い合わせたり、前後のフレームから推測するといった手段を用いてヘッダ情報を補足する。あるいは、当初の想定していた範囲を逸脱している場合には、SMOKA のテーブルでの値の定義を変更したり、例外処理を行うなどの対応をとる。修正ができない場合には公開しない。

- ・ ネットワークトラブル

これは厳密には SMOKA への入力時のトラブルではない。すばる望遠鏡の場合、データ公開日に MASTARS2 から観測データの転送を行いその直後に SMOKA への入力を行う。そのタイミングで MASTARS2 とのネットワーク接続に不具合があった場合、観測データ転送が行われず SMOKA への入力も失敗に終わる。また、SMOKA への入力時の最初に STARS2 のテーブルに接続するため、このタイミングでハワイとのネットワーク接続に不具合があると SMOKA への入力が失敗に終わる。

- ・ 入力のためのプログラムのバグ

KIF 用の SMOKA 入力プログラムは HNR/HWP 用の SMOKA 入力プログラムを参考に作成したが、FITS ヘッダのキーワードの構成およびキーワードの値の書式が KIF と HNR/HWP で異なることが原因でエラーが生じた。KIF の公開を開始して5か月目に初めてプログラム中で進んだ分岐でこの差異によるエ

ラーが発覚した。

このように、手作業中心のSMOKAへの観測データ入力は非効率で手間がかかるものであることは否めないが、問題の発生頻度と回復の時間と作業ミスを考えて現時点ではやむを得ないであろう。幸いなことに観測所や観測装置グループの努力によってHSCのように問題発生頻度が低い観測データもあり、他の観測データについてもそれを期待したい。また、観測データの量やフレーム数が激増している昨今の状況の下、問題が生じた観測データはSMOKAへの入力をあきらめる、あるいは、問題があることを明示して公開する、などという対策が必要となる可能性がある。

以下に、上で述べた作業手順には収まらない例として、一部の観測データに適用している入力作業の工夫と、MCTの入力手順改善について述べる。

(1) 予備テーブル

MTA/MTOのメタデータ管理のための予備テーブル(3.1節)のほかに、観測データのSMOKAへの入力作業の効率化をはかるために、一部の観測装置に対して、入力エラーの事前察知・対策や入力情報の事前修正を行うための予備テーブルを用意している。

- ・西はりま天文台のなゆた望遠鏡NICに対しては、非公開判定を西はりま天文台作成のドキュメントに基づきSMOKA側で行っている。また、修正事項が大量にあるため、3種類の予備テーブルを用意しており、公開開始よりも前にそれらへの入力を行い、公開に備えている。3種類の予備テーブルとは、(F) ファイル管理テーブル、(O) SMOKA登録用テーブルに対応したテーブル(FITSヘッダ抽出事項はそのままだがフレームID変換を施す)、(C) SMOKA登録用テーブルに対応したテーブル(Oから非公開フレームを除き、必要な修正を施したもの)である。SMOKA登録用テーブルへの入力はCのコピーを元に行われる。

- ・岡山天体物理観測所188cm望遠鏡のMCTに対しては、赤経・赤緯などの必須項目が欠落しているフレームを事前に検知すべく、予備テーブル1種類を用意している。公開開始より前にそこへの入力を行って欠落情報のあるフレームを検知して、SMOKA登録用テーブルへの入力の際に手戻りを行うことなく前後のフレームからの補間などによって補填を行っている。

- ・岡山天体物理観測所188cm望遠鏡のHIDES-Fに対しては、一体運用を行っているHIDESのテーブルへの入力の際にFITSヘッダキーワードの非互換などの問題が生じないかどうかを確認すべく3種類の予備テーブルを用意している。3種類の予備テーブルは、SMOKA登録用テーブル、SMOKA検索用テーブル、SMOKAファイル管理用テーブルと同じ構造のものである。

(2) MCTデータの入力作業での手戻りの改善

MCT(2017年12月に共同利用観測が終了した)では、赤経、赤緯、観測装置フランジ方位角、観測者の情報が欠落した観測データがわずかながら存在し、ほぼ毎回の入力作業でそれらを補填する必要があった。

初期の段階では、登録用テーブルの入力作業後にそれらの欠落情報がある観測データを検出し、以降の作業を優先して進めるのに支障をきたさないよう、それら観測データを登録用テーブルから削除し、FITSファイルを公開データ領域から移動し、ファイル管理テーブルでの存在フラグ変更を行ったうえでその他の観測データの入力作業をまず終了させていた。欠落情報のある観測データは、その後に、FITSファイルの公開データ領域への移動、ファイル管理テーブルでの存在フラグ設定、登録用テーブルへの入力を再度行い、欠落情報の(前後時刻の観測データからの)推定を行って登録用テーブルへ入力し、これらの観測データの以降の入力作業を完了させていた。

その後、非効率を改善し、手順の複雑さによって引き起される誤りを減らすべく、前述の予備テーブルを導入して入力作業を始める前に欠落情報がある観測データを検出し、手戻り部分を除去した手順とすることができた。

4.4 検索

SMOKAでは膨大なフレーム数の観測データ(表5)が公開されている。利用者が自分の求める観測データをいかに早く見つけ出せるかが利用者にとって重要である。利用者によってどのような観点や目的で観測データを探すかは千差万別である。SMOKAとしては様々な要求を想定して多種多様な検索機能を提供したいところであるが、開発の人手や時間と維持保守の手間を考慮して以下の検索方法を提供している。

(1) 検索メニュー

- ・シンプル検索：観測装置を選択すると、その観測装

置で観測されたことのある天体名（FITSヘッダキーワードのOBJECTの値。従って、実際の天体名ではなく当該観測者にしか理解できない名称の場合もある）のリストが表示され、その中から天体名を選択する。

- ・アドバンスド検索：赤経・赤緯、観測日、露出時間、観測装置、観測モード（OBS-MOD；撮像、分光、偏光など）、データタイプ（DATA-TYP；OBJECT, BIAS, DARK, FLATなど）、フィルターなどの様々な絞り込み条件を指定する。
- ・カレンダー検索：観測所と年と月を選ぶと、その月にその観測所のどの観測装置がどの日に観測に使われたが一覧できる（この一覧にはそれぞれの日の気象データへのリンクもある）。日付を選択するとその日の観測データ一覧が表示される。
- ・SUP専用検索：SUPに特化したアドバンスド検索。データ品質評価量（シーイングサイズ、星像の伸び、スカイレベル）を用いた絞り込み検索も行える。
- ・HSC専用検索：HSCに特化したアドバンスド検索。データ品質評価量（シーイングサイズ、星像の伸び、スカイレベル、限界等級）を用いた絞り込み検索も行える（2.2節）。
- ・全項目検索：観測装置を選択し、その観測装置のほぼ全てのFITSヘッダキーワードについて、その値や値の範囲を指定して検索が行える。
- ・エンジニアリング検索：HSCの観測データが対象。エンジニアリングの目的で取得された観測データを検索する。HSCの初期の観測データのFITSヘッダには検索に必要な情報が十分に入っていなかった。そのため、通常のSMOKAの検索システムには組み込めなかったが、HSCの初期のエンジニアリングデータは大きな価値をもっており、別途エンジニアリング検索を組み込んでアーカイブデータとして生かすこととした。

図6に各検索機能とテーブル等（シンプル検索と全項目検索には専用のリストが用意されている）の相関図を示した。検索用テーブルだけが検索に利用されるわけではない。

SMOKAの大きな特長として、一度に複数の観測装置を対象にした検索が可能なが挙げられる。例えば、観測装置を特定せずに、どの天域のどの波長域の

どの観測モードという探し方ができる。それを実現するために、共通項目を抽出した（一部の項目は換算・算出した）検索用テーブルと、複数の観測装置を対象にしたSQL文生成プログラムを用意している。また、観測装置によって少々の差異があるDATA-TYPとOBS-MODの値として使用されている文字列について標準形（IMAG, SPEC, など）を設定し、それら標準形に変換して検索することも行っている。利用者の利便性を高める方策ではあるが、標準形変換のためのテーブル（DATA-TYPテーブルとOBS-MODテーブル、図6）が必要になるため観測データのSMOKAへの入力時の手順が増えること、新規観測装置の観測データを組込む際にSQL文生成プログラムの改造の手間がかかることが欠点である。

データ請求画面には、較正フレーム検索機能が組み込まれている。データ請求の選択をした天体フレーム

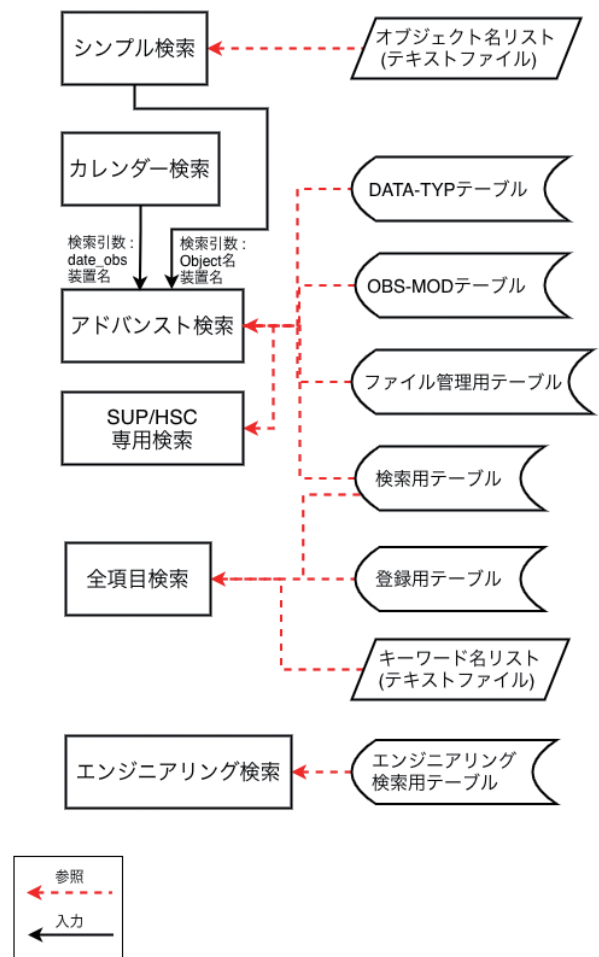


図6：各検索機能とテーブル等の相関図。テーブルのほかにテキストファイルのリストも参照する。シンプル検索が参照するオブジェクト名リストは、データ入力時等にFITSヘッダのキーワードOBJECTの値から作成される。全項目検索が参照するキーワード名リストは、観測装置ごとに用意されており、新規観測装置がSMOKAに追加される時に作成される。

の校正処理に必要な校正フレームの検索が行える機能である（論文2）。それぞれの観測装置の観測データに対応する校正フレームとその条件を定義し（論文2の表1；例えばSUPでは、校正にはFLATのみが必要であり、当該天体フレームと同じ検出器IDかつ同じフィルターでMJDが前後3.5日以内のFLAT校正フレームであること、という観測装置グループが定めた条件）、設定している。校正フレーム検索を行うことで、選択した天体フレームについての条件を満たした校正フレームがデータ請求リストに加えられるが、現在のSMOKAでは校正フレームの検索がこの場面のみであることが課題であり、データ請求に至る前、例えば、検索結果リストの段階で校正フレームの存在が表示されることが望ましい。

(2) 全フレームリスト

利用者の検索に対する要求は多彩である。それらの要求に全て答えるのは現実的ではない。それを補うために、全フレームリストを提供している。フレームIDと観測日時、赤経・赤緯、フィルター、積分時間、データタイプなどの主要メタデータのリストを、各観測装置について観測年ごとにテキストファイルとして用意している。利用者がそれぞれの目的に応じた検索プログラムを構築して、そのテキストファイルから必要なフレームを探すことを想定している。探し出したフレームのテキストファイルを使って、SMOKAのメール請求方式（4.6節）でデータ請求を行うことができる。

4.5 検索補助情報の提供

SMOKAの検索機能を用いて膨大な観測データの中から必要なものを選ぶ際に、その品質の情報が得られればより的確な選択ができる。SMOKAでは、その助けとなるものとしてクイックルック画像の類、（一部の観測装置で）品質評価指標、環境データ（表4）を用意している。

クイックルック画像の類としてQLI、サムネイル、ショットイメージが閲覧できる。SMOKA初期の段階ではQLIを品質評価の主要な道具と位置付けており、ネットワーク速度がさほど速くなかった時代に観測データ本体をダウンロードする代わりにQLIをダウンロードして利用者の手元で評価を行うという目論見であった（論文1）。しかしながら、ネットワーク速度が速くなったため、1フレームごとであれば、観測データ本体をダウンロードをして容易に評価を行えるようになった。一方でフレーム数が激増し、1フレームごとの評価の前にまず複数フレームの様子を一度に眺めた

いという要望が高まり、QLIから作成するサムネイルやショットイメージの意義が高まった。このようにしてQLIの意義が低下したことと、QLIには統計値などの品質評価情報を付加している（論文1）ため、その作成には相応の手間がかかることから、最近SMOKAに組み入れたIRD、SWS、MMZ、KIFでは、QLI作成を省略して、サムネイルをFITSファイルから直接作成するようにした。近年観測データがますます膨大な量になり、その構造も複雑化しており、MEFの観測データのクイックルック画像をどのようにすべきかが課題として残されている。

HSCとSUPにおいては、SMOKAで品質評価指標を算出し、それぞれの専用検索（SUPについては論文4の2.2節、HSCについては本論文の2.2節参照）から品質評価指標による検索ができるようにしているほか、検索結果表示画面でもその指標を表示している。

データ解析やその解釈の際に観測時の環境データを把握する必要がある場合も多いと思われる。SMOKAでは、気象データページを用意し、気象データページで気温、湿度、気圧、風速、風向といった気象データに加え、全天モニタ画像（縮小版）も閲覧できるようにしており、カレンダー検索のカレンダー画面に気象データページへのリンクを用意している。さらに、一部の観測所についてはDIMM（Differential Image Motion Monitor）やAG（Auto Guider）によるシーイングの情報も閲覧できるようにしている。すばる望遠鏡についてはSummit log（山頂での観測日誌）も検索し閲覧することが可能である。表4に各観測所ごとの環境データの一覧をまとめた。各種データの取得期間（機器の導入時期や故障などによる廃止時期によって様々である）も記載した。

4.6 データ請求

利用者は、SMOKAの検索機能などを利用して必要な観測データを決定したのちに、SMOKAのデータ請求機能を利用して観測データを取得することができる。利用者にとってはできるだけ簡単かつ迅速に観測データを入手したいところであるが、SMOKAの特長であるHeader Replacer（本節（1）で簡潔に述べる）（論文3、4）を介さなければならぬために少々の手順と待機時間が必要になっている。

SMOKAでは、データ請求方法として、ウェブ請求方式とメール請求方式（論文5）を用意している。ウェブブラウザの検索結果画面で観測データのフレームIDの選択を行うなど、グラフィカルインターフェースを利用した直感的・容易な請求を利用者に提供するためにウェブ請求方式を用意している。加えて、複数

表4：各観測所の環境データ一覧.

観測所・望遠鏡	環境データ種類	データ取得期間
ハワイ観測所	全天モニタ画像	2001年07月～2007年09月
	気象データ	2000年01月～
	シーイング (DIMM) ^a	2005年06月～2007年12月
	シーイング (AG) ^a	1998年12月～
	フィールドモニタ ^a	2007年04月～2010年03月
	summitlog ^a	2004年11月～
	CFHT 全天モニタ画像 ^b	2010年01月～
	CFHT 気象データ ^b	1999年01月～
	CFHT シーイング (DIMM) ^b	2009年09月～
岡山天体物理観測所	全天モニタ画像	1997年08月～
	気象データ	1996年04月.
	シーイング (DIMM)	2004年11月～2008年02月
	観測野帳 (スキャン)	2000年10月～2014年09月
東京大学木曽観測所	全天モニタ画像 ^c	2001年09月～
	気象データ ^c	1990年10月～
	天候表 (スキャン)	1993年01月～2014年11月
MITSuME 岡山	全天モニタ画像	(岡山のデータ)
	気象データ (FITS ヘッダ)	2006年08月～
MITSuME 明野	全天モニタ画像	2014年04月～
	気象データ (FITS ヘッダ)	2006年04月～
東広島天文台	全天モニタ画像	2008年12月～
	気象データ	2008年12月～
西はりま天文台	全天モニタ画像	(SMOKA には転送されない)
	気象データ (FITS ヘッダ)	2013年02月～
京都大学せいめい望遠鏡	全天モニタ画像	(岡山のデータ)
	気象データ	(岡山のデータ)

^a 論文5参照

^b CFHT (Canada-France-Hawaii Telescope) に許可を得て使用している.

^c 機器が何度か更新されており, 中断期間や重複期間あり.

の検索結果画面に表示された観測データを一括して請求すること, 4.4節で述べた全フレームリストを利用して定めた観測データを請求すること, そして, SUP, HSC および KWF の観測データをショット単位で請求することのためにメール請求方式を用意している.

請求情報は, SMOKA のデータ請求処理機能の中核機能である Data Copy Manager (以降, DCM; 論文1参照) に入力される. DCM は検索機能などの他の SMOKA の機能と分離されており, DCM への入力情報はファイル (請求ファイル) として渡される. システムを分離した目的は, 保守や開発の際に独立した試験が行え, それぞれの改造がお互いに影響を与えにくくすることにある.

データ請求処理の流れを図7に示す.

(1) Header Replacer

Header Replacer は DCM 中で最も主要な機能であり, SMOKA を特徴づけるものである. ここでは, その機能の背景にある方針と, 現状での課題を述べる.

観測データにおいて, その FITS ヘッダを修正あるいは補填する必要が生じることがしばしばある. 望遠鏡や観測装置の不具合により FITS ヘッダ情報が誤っているあるいは欠落していること (フィルターや赤経・赤緯が誤っているなど) があつたり, 位置較正済みの WCS 情報などの有用な情報を観測の後から加えることがあるためである.

SMOKA では, データ請求で起動される処理の中で Header Replacer を用いて, 利用者が観測データを請求する際に, その FITS ヘッダに対して修正・補填を施している. あらかじめ元の観測データを書き換えておく

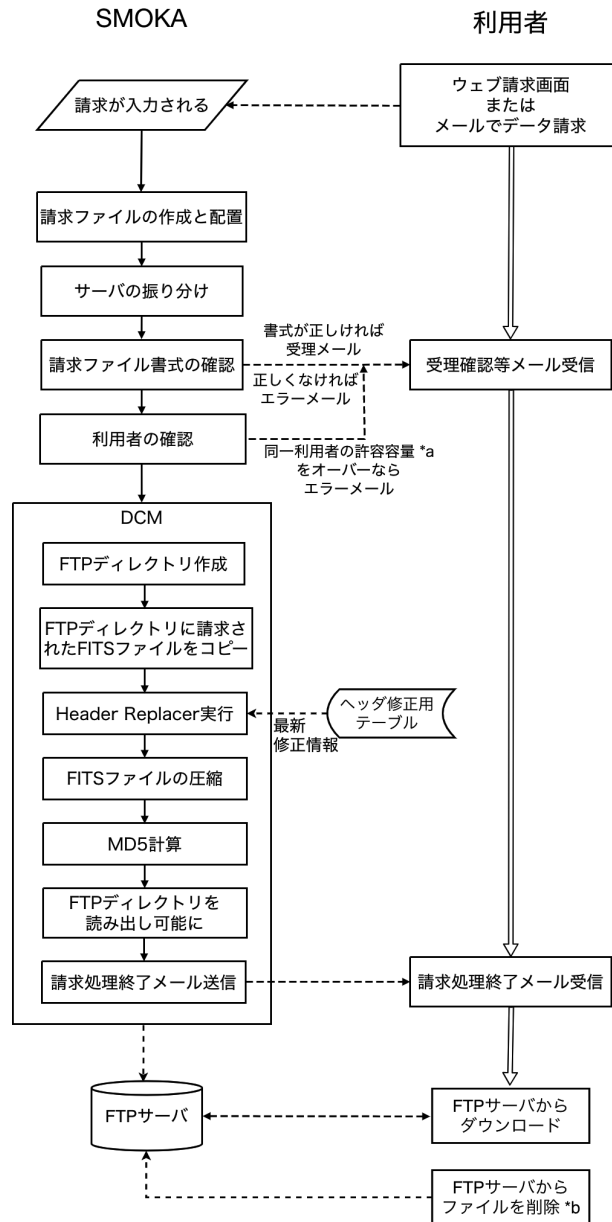


図7：データ請求処理の流れ：実線はSMOKAでの処理の流れ、破線は情報（ファイル含む）の流れ、白抜き線は利用者の作業の流れ。DCMは四つのサーバで独立に稼働する。利用者登録時に、利用者アカウントごとに対応するサーバがひとつずつ割り振られており、請求ファイルに記載の利用者アカウントに応じたサーバに振り分けられてDCMの処理が行われる。
 *a：特定の利用者が大量の請求を行うことで他の利用者の処理が圧迫されることを防ぐために、SMOKAでは、同一利用者が同時に利用できるFTPサーバ上でのファイル容量は合計4096 GBまでとしている。これを超えてデータ請求を行った利用者には受付拒否のエラーメールが送信される。
 *b：利用者には、FTPサーバの過去の自身の請求ファイルを消去できる機能を提供している。ファイル容量制限を超えている場合には、利用者が、ダウンロード済みのファイルをFTPサーバから消去することで、次のデータ請求を行うことができる。

のではなく、データ請求ごとにFITSヘッダの修正・補填を行う理由は以下である。

- ・元の観測データの書き換えの際に誤りが生じると取り返しがつかなくなる。
- ・元の観測データを書き換えてしまうと観測所側の観測データとの照合ができなくなる。

- ・元の観測データの書き換えには相応の時間と手間がかかるため、対象となるフレーム数が多い場合に迅速な修正・補填ができないことがある。また、修正・補填の必要性が複数回生じる場合に最新の修正・補填まで至っているかどうか懸念がある。

修正・補填は以下の2種類の作業で構成される。(I)
既にFITSヘッダに存在するキーワードの値を変更し

たりキーワードを消去する（変更前の行はコメント化する）作業、および、(II) FITSヘッダに存在しないキーワードを行単位で（あらかじめ設定された順序で）追加する作業である。この2種類の作業の際に入力する情報はその内容や形式が大きく異なるため、(I) (II) にそれぞれ対応する2種類のテーブルを観測装置ごとに用意している。修正・補填の準備作業として、修正・補填対象データのフレームIDおよび修正・補填後のキーワードとその値をテーブルに入力する。このとき、そのキーワードが対象データのFITSヘッダに存在するか否かを把握したうえで、(I) または (II) のどちらのテーブルに入力するかを判断する必要がある。

このように、現状では、Header Replacerの処理の方法に応じてテーブルの種類を上記のように二つに分けているため、修正・補填の準備作業が複雑になっている。テーブルの構造およびHeader Replacerのプログラムを改良して修正・補填の準備作業をより単純にすることが今後の課題である。また、現在のHeader Replacerには、FITSヘッダの各行後半のコメント部分を追加、あるいは、修正する機能が存在していないが、その機能が必要な場合が発生しており、それも今後の課題である。

別の課題として、データ請求の応答時間（利用者からのデータ請求受取からダウンロード準備完了通知を出すまでの時間）の改善が挙げられる。現在（2015年12月14日以降）では、データ請求の応答時間を短くするために、4台のサーバでDCMを稼働することで、同時処理可能なデータ請求処理の本数を増やしているが、応答時間を長くしている最大要因はHeader Replacerである（その所要時間のHSCの例については2.3節を参照のこと）。この対策として、新たな観測データの取得が既に終了しており、FITSヘッダの今後の修正・補填も見込まれない観測装置の場合には、あらかじめ修正・補填処理をしておき、Header Replacerを経ないようにすることが挙げられる。2.3節で述べたHSC専用データ請求機能は、HSCデータではFITSヘッダの修正頻度が少ないという特長を生かしてそれを実験的に実現したものである。

また、SMOKA側でFITSヘッダの修正・補填を行うのではなく、利用者に修正情報と修正プログラムを提供する、という方法も考えられる。SMOKAのHeader Replacerでは修正データベースに登録されている全ての修正・補填を適用するが、利用者の目的用途によっては不要な修正が含まれているかもしれない。利用者それぞれがそれぞれに必要な修正を施すことによって、SMOKAでのデータ請求待ち時間を短縮するとともに、修正内容の取捨選択を行い、場合によっては利用者独自の修正も加えることで相応の利点が想定できるであろう。

FITSヘッダ修正が必要な（データ請求時にHeader Replacerが起動する）フレームの数（2021年7月1日現在）を観測装置ごとに表5に示した。Header Replacerで修正、追加、削除するFITSヘッダのキーワードには、誤っているのを修正する、必要な情報なので補填する、不要な（誤解を招く）情報なので削除するだけでなく、便利な（有用な）情報なので追加するものがあり、典型的なものは位置校正情報である。表5では、修正フレーム数のカラムの（）内に、位置校正情報追加フレーム数を記した。

(2) FTPサーバからのダウンロード

データ請求された観測データはFTPサーバの所定の領域に置かれ、利用者はそこからダウンロードする。このことは、観測データを検索・請求するサーバとダウンロードするサーバを分離することによる役割分担・負荷分散（論文4）を実現するものであり、技術面および運用面で利点がある。

大量のデータ請求を行う場合には、メール請求方式が利用できる現状で利便性に問題はないように思えるが、数フレームのみを請求する場合には、データ請求とその後の待機時間、FTPサーバへのアクセスという手順は煩雑であり利便性に問題があると思える。そのような場合には検索結果表示画面で欲しいフレームをクリックすれば（必要な場合はHeader Replacerを経て）その場でダウンロードできるという機能が便利であろう。この機能のプロトタイプを試作は済ませているが、運用面での課題（利用者認証、フレーム数、データ量の制限やクリックの時間間隔など）の検討が進んでおらず、実装に至っていないという状況である。次世代観測データアーカイブシステムでは実現すべきであろう。

4.7 閉鎖した機能

SMOKAの運用や開発は発展ばかりではなく、行き詰まることもある。開発して運用の段階に至った機能でもその中のいくつかはその後、運用を続けることができなくなり閉鎖する破目に陥っている。その中の主なものは移動天体検索機能（論文6）、重複領域検索機能（論文5）、超新星検索機能（本論文2.1節）、全文検索（論文5）である。

移動天体検索機能では太陽系内移動小天体の軌道要素ファイル（Minor Planet Center¹²⁾ から提供されているMPC Orbit Databaseの全軌道要素ファイル¹³⁾、毎日

12) <https://minorplanetcenter.net/>

13) <https://www.minorplanetcenter.net/iau/MPCORB.html>

表5: FITS ヘッダ修正が必要なフレームの数 (2021年7月1日現在).

装置	修正フレーム数 ^{a, b}	総フレーム数	修正割合 (%) ^b
すばる望遠鏡			
CAC	0	5,916	0.0
CIA	0	153,748	0.0
COM	816	95,160	0.9
CRS	8,989	40,188	22.4
FCS	1,320	180,179	0.7
FMS	0	25,156	0.0
HDS	2	138,647	0.0
HIC	54	145,239	0.0
HSC	129,485	11,218,826	1.2
IRC	11,023	441,875	2.5
IRD	14,797	20,271	73.0
K3D	78	5,416	1.4
MCS	5,277	239,772	2.2
MIR	0	7,432	0.0
MMZ	0	967	0.0
OHS	0	175,548	0.0
SUP	101,671 (101,500)	1,181,473	8.6 (8.6)
SWS	0	7,801	0.0
岡山天体物理観測所188 cm 望遠鏡			
CSD ^c	0	40,734	0.0
ISL	221,646	221,646	100.0
HID	309,663	315,717	98.1
KLS ^d	0	26,523	0.0
OAS ^e	0	9,824	0.0
MCT	385,726	385,726	100.0
東京大学木曾観測所シュミット望遠鏡			
KCC ^c	62,212 (26,806)	62,354	99.8 (43.0)
KCD ^c	193,469 (112,635)	193,470	100.0 (58.2)
KSP/KSQ ^f	0	7,039	0.0
KWF	1,299,104 (699,535)	1,299,104	100.0 (53.8)
東広島天文台かなた望遠鏡			
HNR	484	314,089	0.2
HWP	9,072	510,154	1.8
東京工業大学 MITSuME 望遠鏡			
MTA	6,058,948 (3,403,667)	7,240,945	83.7 (47.0)
MTO	3,808,729 (3,432,560)	6,064,060	62.8 (56.6)
西はりま天文台なゆた望遠鏡			
NIC	195,793	213,049	91.9
京都大学せいめい望遠鏡			
KIF	118	27,575	0.4

^a 修正フレーム数が0である観測装置は、修正の必要がない場合（観測所側が整えた場合を含む）だけではなく、修正すべき箇所が判然としていない場合もあると思われる。

^b 修正フレーム数および修正割合のカラムで () 内は位置較正情報追加フレームの数および割合である。

^c 観測データ入力前に SMOKA がデータ本体をある程度修正した。

^d 観測所側が FITS ヘッダ情報を整えて観測データ本体を修正したうえで SMOKA に転送した。

^e どう修正すべきかわからないまま修正を断念し、そのまま公開したものと公開しなかったものがある。

^f TIFF から FITS への変換の際に FITS ヘッダ情報を整えており、Header Replacer を使わない。

更新される)を定期的に取得する必要がある、超新星検索機能では超新星のカatalog (IAUの超新星ワーキンググループが提供するTNSで超新星を含む突発天体のリストが毎日更新される)を定期的に取得する必要がある。

これらの情報収集と更新の作業は手間がかかるものであり、移動天体検索機能と超新星検索機能を再開するためには、これらの作業を自動化する必要がある。自動化する場合にはその開発が必要なだけでなく、運用においてその処理結果の監視が必要(過去に情報提供元のWebアドレスやデータ形式が変更されたことがある)であり、それら開発・運用のコストが人手による手間を補って余るものになるかを見極める必要がある。

また、移動天体検索機能でも超新星検索機能でも、そして、重複領域検索機能でも位置較正を行った観測データが対象であり、位置較正を経てからHEALPix(論文5)の計算を行ってそれをテーブルに入力することで成り立っている。しかしながら、位置較正はそれぞれの観測装置の観測データに適した手法の開発から取り組まねばならないために多くの労力と時間を要し、また、FITSヘッダに元々記載されている赤経・赤緯等の天球位置情報が不正確な観測装置から順に手がけてきたために、これらの機能として需要が高いと思われるSUPやHSCにまで手が届いていなかったのが最大の問題であった。

これらの検索機能はいずれも開発研究としては十分な成果を挙げたと言えるが、予算や人員などの現実の下での運用には無理があったと言えよう。SMOKAの様々な機能の中には、「なければならない」ものも多いが、「あったら便利」というものもある。後者についての利用者の要求や利用の頻度などは時代とともに変わっていくことが考えられ、先を見通すのは簡単ではない。タイムドメイン天文学が天文学の主流の一部となつて久しく、高い時間分解能をもつ観測装置も登場しており(7.2節)その勢いはまだ衰えていないように思える。この状況下では、重複領域検索機能は「あったら便利」よりは「なければならない」寄りの位置付けであるかもしれない。新しい技術(先進的なアルゴリズムや多様な機能をもつライブラリ)を導入して、位置較正について克服することも検討すべきであろう。

これまでSMOKAでは利用者や観測所などからの要望とともに、SMOKAグループ内外での天文学的興味や技術的挑戦度などを勘案して新たな機能の開発と実装を行ってきた。いずれにせよ、今後、新機能を追加する際には、予算や人員などを踏まえた実現可能性を十分に検討する必要がある。

4.8 計算機システムの更新

SMOKAの運用の中核を担う計算機システムは借入(レンタル)で導入されている。開発や作業を担う計算機は買取にても賄われているが、運用要員の数などを考えた場合にハードウェアや基本ソフトウェアの支援が伴ったレンタル計算機システムなしにはSMOKAの安定した運用は存在しえない。

そのレンタル計算機システムは、契約の都合上、たいていの場合、5年に1度更新(リプレース)される。SMOKA運用開始以降では、2008年3月(新システム稼動開始)、2013年3月(同じく)、2018年3月(同じく)、に計算機リプレースが行われた。いずれも政府調達(スーパーコンピュータ)であり、2018年3月の際の調達日程は以下の通りであった。

- ・資料提供招請公告：官報平成28年8月22日(政府調達第157号)
導入説明会9月1日
- ・意見招請公告：官報平成29年2月1日(政府調達第20号)
仕様書案説明会2月9日
- ・入札公告：官報平成29年5月26日(政府調達第98号)
入札説明会6月2日
入札書受領期限7月24日
開札9月5日
- ・落札公告：官報平成29年12月11日(政府調達第233号)
新システム稼動開始平成30年3月1日

検討を開始してから新システムが稼動を始めるまで、おおむね2年間かかる。また、稼動を開始してもすぐに完璧に運用できるわけではなく、様々な調整が必要となり、5年間の借入期間の半分以上は前後の更新に関わる作業を併せて進めてきた。

毎回の計算機リプレースでは、それまでシステムの問題点や弱点などを克服してさらなる発展を目指すような設計を行い、それを予算の範囲内で実現できるようにしてきた。膨大なデータを旧システムから新システムに移行(コピーとMD5算出・照合など)する手間と時間はやむを得ないが、プログラムの類はできるだけ手間をかけずに移行できるようリプレースの前後でOSやDataBase Management System(DBMS)などを同じ系列としてきた。しかしながら、2013年のリプレースは例外であり、大きな変革を迫られた。すなわち、大部分のマシンのOSをSolaris10からRedHat Enterprise Linux 6(RHEL6)へ、また、DBMSをSybase

から PostgreSQL へ全て変更することになったのである。DBMS については予算の面が大きい、OS についてはレンタル業者の都合と各種フリーソフトウェアが動作するかどうかによるところが大きい。また、大部分のマシンでは CPU チップが SPARC 系から Intel 系に替ったことによってビッグエンディアンからリトルエンディアンへの変更が生じたが、扱うバイナリデータのほぼ全てが FITS (ビッグエンディアンと規定されている) であり、影響は一部にとどまった。このようにこの回の移行は以上のような非互換項目が多く、作業量が大きくなった。

一方、移行作業についての基本的考え方である、できるだけ運用停止期間を短くして利用者が被る不便を最小限にすることと、移行作業をできるだけ単純かつ容易にすることを実現すべく、旧システムと新システムを一時期共存させて移行作業を進めることにしているが、そのために旧システムと新システムではサーバなどの名前を異なるものにすることや磁気ディスク領域、特に NFS 領域の名前を異なるものにすることがある。つまり、移行の前後でサーバ名や領域名を変える必要があり、それによって種々の設定やプログラムの変更が必要になる。また、運用を続けながらの移行作業であるため、コピーした後に MD5 を算出して確認している間にファイルはどんどん増えていき、コピーや MD5 算出を逐次継続しなくてはならない。アクセスログなどの記録も同様である。また、リプレースの経費をできるだけ削減すべく、移行作業はレンタル業者に委ねるのではなく、SMOKA グループの自力に依ることになり、計算機リプレースの際の作業量は調査や設計、調達などの事前作業だけでなく、移行作業そのものも膨大になる。なお、予算削減にともない、移行作業だけでなくその後の運用においても業者に支払う経費を削減すべく自前の作業の比率が高くなっている。

また、同じレンタル契約に属するために上流の MASTARS2 と同じ時期のリプレースとなること、さらに上流の STARS2 は別のレンタル契約であるが、MASTARS2 と連携しているために、引き続いた時期のリプレースとなることも移行作業に時間的制限を加え、作業の負担を増やしている。

SMOKA グループでは、SMOKA だけでなく ADS ミラー (現在は運用終了) や Vizier ミラーなどの運用も行っており、それらのサーバの移行も必要であった。

直近の 2018 年 3 月のリプレースの移行作業の主要部分を以下に具体的に示す。この回は OS や DBMS について大きな変革はなく、比較的平穏な移行作業であった。

まず、観測データや付随データ (環境データなど)、派生データ (QLI や統計値など SMOKA で作成算出

したもの) などのデータの移行を 2017 年 11 月 1 日に旧システム上のファイルの MD5 を計算してそれをデータベースに記録されている値と比較する作業から始めた。続いて、旧システムから新システムへのデータコピーを特設経路 (SMOKA 運用とは別のネットワーク) を用いて 2017 年 12 月 8 日に開始し、2018 年 1 月 10 日に第一段階 (約 410 TB) が終了して、MD5 の新旧比較を開始した。それ以降は運用に伴って増加・改変されたデータを継続的に随時移行した。そして、ファイルサーバ 4 台の切り替えを 2018 年 2 月 8 日から 9 日に順次行い、データ移行作業は完了となった。次に、ソフトウェア (レンタル業者担当の OS や主要フリーソフトウェアを除く) の移行は、2018 年 1 月 16 日のマシン引き渡し (それ以降は国立天文台側が使って良い) から始まり、各種設定や httpd, tomcat, などフリーソフトウェアのインストール・設定から SMOKA 本体の移行作業を行った。DBMS (PostgreSQL) のデータ移行・切り替え (2018 年 2 月 1 日から) と、ファイルサーバの切り替え (2018 年 2 月 8 日から 9 日) を経て旧システムから新システムへの段階的移行を進め、2018 年 2 月 27 日の DNS の CNAME の切り替えによって新システムへの移行が完了した。しかし、その後も同年 3 月末の旧システム完全停止・撤去まで各部の調整や不具合修正を続けた。さらに、移行作業に携った SMOKA グループ総員 4 名は、移行期間の間は旧新両システムの運用作業も担っていたためにどうしても少々確認漏れが残し、しばらくの間、不具合が判明することに対応を繰り返した。

この移行作業にあたっては以下のような重大な問題が生じた。

まず、磁気テープライブラリ上のデータの移行中に、旧システムの磁気テープ (LTO5) を新システムのドライブ (LTO7) で読み出す際のエラーが多発した。当初作業を行った磁気テープ 114 本中 61 本でエラーが発生して移行作業が頓挫した。この障害は同年 6 月上旬のテープドライブのファームウェアの更新で解決したが、移行すべき磁気テープ 135 本の移行作業は同年 7 月下旬までかかり、多大な時間と労力を費した。また、OS が RHEL6 から RHEL7 に変更になったことに伴い、cron は初期設定で `crontab` がインストールされていて¹⁴⁾、その仕様上、指定日時に確実にジョブを実行させることができない状況であった。そのため、`crontab` ¹⁵⁾ に入れ替える必要があった。さらに、NFS サーバとして Solaris11 (ファイルシステム

14) https://access.redhat.com/documentation/ja-jp/red_hat_enterprise_linux/7/html/system_administrators_guide/ch-automating_system_tasks/#s1-Scheduling_a_Recurring_Asynchronous_Job_Using_Anacron

15) https://access.redhat.com/documentation/ja-jp/red_hat_enterprise_linux/7/pdf/migration_planning_guide/Red_Hat_Enterprise_Linux-7-Migration_Planning_Guide-ja-JP.pdf

ZFS), NFS クライアントとして RHEL7 という構成が導入設定されたのだが, ファイルサーバ上のファイルがクライアントから見えない, という事象が発生した。例えば, ファイルサーバ上の数千~数十万ファイルを読み出してクライアント上で処理する場合に相当数(数百~数万)のファイルの処理が行われず, OS 側のエラーも発出されない, というたいへん厄介な現象であった。ソフトウェアの改変と運用で急場を凌ぎ, 同年6月中旬の OS の修正作業の後に発生しなくなったが, 多大な労力と時間を費やすこととなった。

計算機システムのリプレースに際して技術の進歩や市場の動向に左右されることは逃れられない。また, 移行作業を軽減して安定した運用を継続しつつ, より効率的なシステムに進歩していくように十分に検討しておかねばならない。そのための根本的対策は, 十分な人員と予算の確保であるが, それは現在の状況ではきわめて困難である。工夫や創意にも限界があり, SMOKA による成果の拡大と存在意義の主張を通じて説得力を増していくしかないだろう。

4.9 未解決の課題

以上で述べたように SMOKA では長年にわたって様々な技術的課題に取り組んできた。しかし, SMOKA にはまだ未解決の問題が多く残されている。解決すべき課題は天文学の進歩や観測装置の発展などにともなう増えていくであろう。さらに, 計算機関連技術の発展によってこれまで最良だと考えてきた解決策よりも優れた方策が現れるかもしれない。天文学や計算機関連技術の動向には常に注目していかなければならないであろう。

本章の各節でもそれぞれについて残された課題や今後の検討課題について述べたが, 現時点で重要だと考えている課題をまとめると以下の通りである。

(I) Multi-Extension FITS (MEF) の情報

SMOKA では観測データの情報の表示や検索には FITS の Primary ヘッダの情報だけが使われている。しかしながら, 7.2 節で述べるように MEF の観測データが増えてきており, Primary 以降に単なる付加的な情報ではなく, 重要な情報が記載されている場合が多くなってきている。MEF の2番目以降の情報をどう表現するかが今後の検討課題である。とりわけ, フレーム ID を主キーとしたテーブルを構築している SMOKA では, MEF の2番目以降の情報を検索にどのように取り込んでいくのかを検討する必要がある。さらに, 有用な画像が納められた全ての Extension に対応した QLI やサムネイル等も検討課題

である。

(II) 天球位置情報の表現 (WCS)

HSC や SUP (2008 年7月の検出器入れ替え以降) のように一つのフレームの画像中にギャップ (オーバースキャン領域等) がある場合には, 天球位置情報を表現する WCS は離散した複数 (HSC や SUP の場合は四つ) のものになる。これを FITS ヘッダ中に正しく記述する方法は現時点では存在せず,

- ・複数の WCS を FITS ヘッダに記載することは可能なので, 画像上でのそれぞれの対象範囲を何らかのキーワードで表現し, 利用者が判断して切り換える。
- ・オーバースキャン領域等を画像の外側に移動してギャップがない画像を Image Extension にし, そこに統合した WCS を加える。
- ・画像を四つの部分に分けてそれぞれ Image Extension にし, それぞれに WCS を加える。

などが考えられる。時間と手間を要するが, FITS の新たな規約を IAU に提案するのが正攻法であろう。

(III) 観測データ取扱の単位 (フレーム)

SMOKA ではフレーム単位で観測データを扱っている。SUP, HSC および KWF ではショット単位の検索やデータ請求が可能であるし, ショットイメージの提供もしているが, 基本の単位はフレームであり, テーブルの主キーもフレーム ID である。(I) で述べた MEF の情報の表現や, (IV) で述べる関連付けられた観測データの組, 7.2 節で述べる大量データを産出する観測装置の場合, フレーム単位では数が多くなりすぎるなど, この枠組みだけでは無理が生じる場合が想定され, 観測データを扱う単位を柔軟に拡張していく必要がある。また, 位置校正の結果得られる WCS については, 現在, 各フレームのものをそれぞれのフレームに書き込むだけであるが, ショット全体の WCS (各フレームとショットとの座標変換情報を含む) もあればより便利だと思われる。それをどのように FITS ヘッダに記載するのかが検討に値するであろう。

(IV) データ間の関連付け

モザイク CCD, 多色カメラなど, 複数の検出装置が同等の観測データを同時 (検出装置それぞれが独立して観測データを生成する場合もある。これら

は、EXP-IDで結び付けられる場合が多く、しばしばショットと呼ばれ、本論文のショットとは定義が異なる）に生成する場合だけでなく、7.2節で述べるように異質のデータを組として利用すべき場合もあり、フレーム間の関連付け、あるいは、セット化が必要になってきている。

また、SMOKAでは較正フレーム検索機能（論文2）を提供しているが、現時点ではデータ請求の際に初めてその機能が起動されるものであり、検索結果で得た観測データに対して標準的な較正データが存在するかどうかは検索結果画面ではわからない。較正データだけでなく、分光データに対するスリットビュー画像のような関連付けを含めて、標準的なデータ解析を行うために必要なフレームのセットとしての検索が必要であろう。さらに、気象データ、全天モニタカメラ画像などの観測条件を記録した種々のデータとの関連付けをより強化する必要もある。

(V) 手動作業と自動化の選択

機器の不調は皆無にはできないし、ある程度自然を相手にしなければならない天文観測では不測の事態は避けられず、その際の人間の判断に基づく対処は頼りになることが多い。SMOKAではそうした考えと経験に基づき、手動作業を多く採用している。しかし、人間の操作には必ずやミスがともなう。人間のミスを除去するとともに作業の効率化を目指して自動化を進めることも重要である。だが、プログラムには必ずバグがあると言っても過言ではなく、想定外の事態への対処には限界がある。それらを克服して自動化するための労力と時間に対し、それぞれの作業の実行頻度と重要性に鑑みて定めることになろう。AI（人工知能）などの新しい技術の導入もその得失を良く理解したうえで検討を進める価値がある。

以上の検討課題の他に、利用者に求められながらも取捨選択の結果、あるいは、人手と時間をかけられずに実現できていない機能も多々ある。SMOKAが立ち向かうべき技術的課題は尽きることがないと言えよう。

5 SMOKAの運用面での諸問題

本章ではSMOKAがこれまでに直面してきた運用面での諸問題について論ずる。SMOKAが立ち向かうべき課題、とりわけ運用面での問題の根本的解決は人員や予算に依るところが大きい。また、観測データを公

開するという方針をゆるぎなく持ち続けるためには組織や体制が重要である。20年の間の人員体制はおおむね、承継職員（法人化前は正職員）でSMOKA主務の者1名、SMOKAを主務とする契約職員3ないし4名、外部の者最大5名、であった。時とともに扱う観測データの量も数も観測装置の種類も増えてきており、日常の運用にかかる手間も増えてきている。人員、予算、組織、および体制についての検討や議論や交渉や主張は途切れることなく継続していかなければならないが、本論文の主旨ではないのでそれらについては省略する。

5.1 観測データの公開に至るまでの問題

観測データアーカイブを運用するにあたっては、まず観測データが公開されねばならないが、それは一筋縄ではいかない場合が少なくない。以下のような事情が観測データの公開を妨げていると考えられる。

- ・ 観測データをSMOKAで公開するためには、SMOKAと観測所との連携が必須であるが、観測所側に人手が足りず、そこに労力を注げない。
- ・ 観測データの公開の方針やルールが観測所側で定まっておらず、まずそれらを定めなければならない。
- ・ データフォーマット、特にFITSヘッダの整備が不十分で、データ解析に必要な情報が網羅されていないなど、先にその整備が必要である。
- ・ データ解析などのドキュメントの整備が不十分で、まずそれらを整備しないと公開された観測データの十分な活用が期待できない。

7.1節で述べるように、研究成果（論文）の発表や資金の獲得などに際して観測データの公開が求められる場合が増えており、観測データ公開の機運は高まってきているが、上記の事情に様々な心情的要素が加わって消極的な姿勢になっている場合が多いと思われる。

SMOKA側が観測所側に観測データ公開を強引に迫ることはないが、反復的な宣伝活動と上記の阻害要因を少しでも減らすためのSMOKA側の支援と助言によって少しでも多くの観測データが公開されるように努力してきている。

図8に観測データの公開に至るまでの手順を示す。

5.2 SMOKAの運用にともなって発生した問題

20年間のSMOKAの運用では様々な問題が発生した。

しばしば起きている問題は以下のような一般的なものが多い。

- ・ネットワーク不通による転送失敗
- ・磁気ディスク領域あふれ
- ・データベース領域あふれ
- ・サーバ障害（故障）による運用停止
- ・プログラムのバグ（特に、分岐の頻度が低い箇所に）

観測所側のテーブルに記載されているファイルパスが誤っている（あるいは、そこにファイルがない）場合も時々生じる問題である。また、誤りの修正や欠落情報の補填などFITSヘッダに関する問題はしばしば発生する。新規観測装置からの観測データをSMOKAに組み込んでから暫くは様々な問題が生じることが多い（4.1節）。

この20年間に発生した特に重大な問題は以下の（1）から（5）である。これらの回復にはいずれも多大な労力を要した。

（1）2006年4月観測データ損傷

MD5値が上流（MASTARS）と一致しない観測データが、FCS、COM、HDS、SUP、IRC、COMQ（COMの簡易処理済みデータ、論文4）、IRCQ（IRCの簡易処理済みデータ）あわせて530フレーム発見された。MASTARS上で壊れたものと推測され、これらの観測データはSTARSから再コピーをして修復した。

（2）2008年5-6月観測データ損傷

すばる望遠鏡の観測データ全フレーム（2006年9月以前の1,302,197フレーム）について、SMOKAファイルサーバ上のFITSファイルのMD5ハッシュ値とSTARSに記録されているものとの照合を行った。その結果、SUPの299フレームが不整合であった。全てSMOKAへの入力後に変わったと推測されるため、MASTARSより再転送して入れ替えた。

（3）2013年7-8月IRCデータ公開取下げ

SEEDS [50] のプロポーザルIDで観測されたIRCの観測データの観測者占有期間を18か月から30か月に延長するため、既に公開されている298フレームの公開を停止する旨の依頼がハワイ観測所からあり、SMOKA側ですぐに対処した。しかし、STARS2での当該観測データのPUBLICTIMEの変更が行われるまでに20日間ほど要した。同じプロポーザルIDでIRCおよびHICの2観測装置で観測が行われたが、IRCのPUBLICTIMEの変更の際に、同じプロポーザルIDのHICの観測データ（15,858フレーム）の観測者占有期間

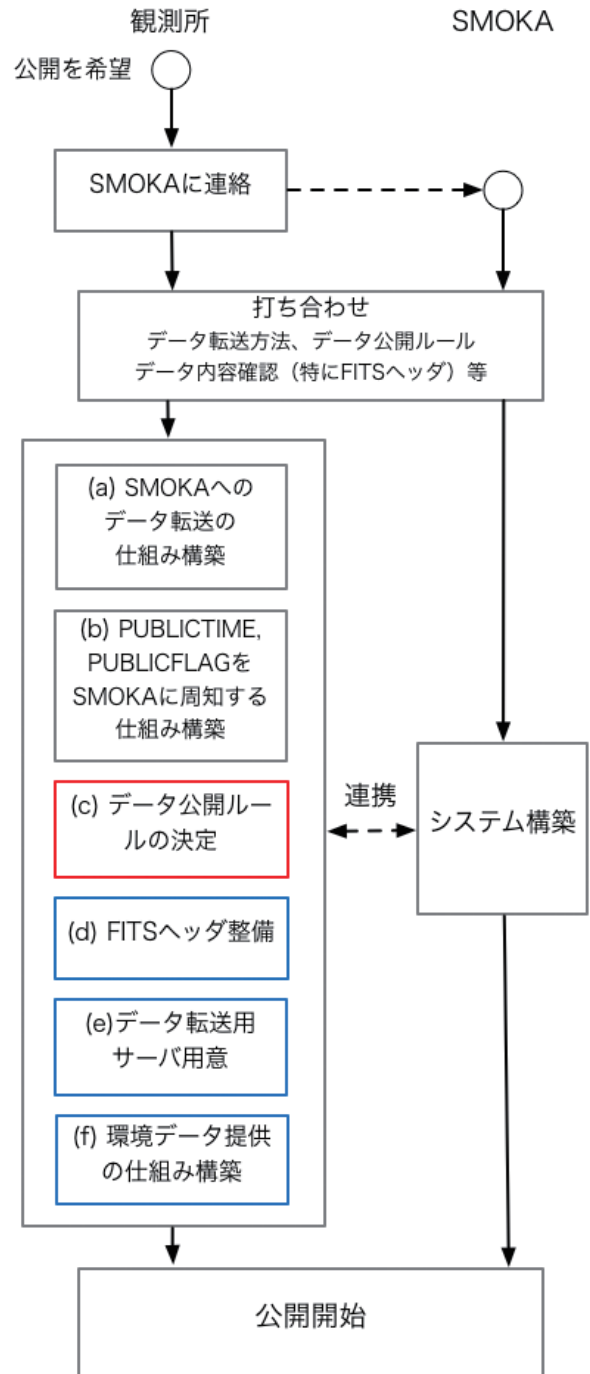


図8：観測所が観測データの公開を希望してからSMOKAでの公開を始めるまでの流れ。共同利用観測ではない大学等の観測所の場合には、（c）データ公開のルールの決定は省略可能である（赤囲みの部分）。共同利用観測の有無にかかわらず、（d）、（e）、（f）の手順は場合によって省略可能である（青囲みの部分）。（d）較正・解析に必要な情報がヘッダに含まれていない場合には、電子的に読み取り可能な形のリストが観測所から提供されれば、SMOKA側でFITSヘッダに埋め込みを行う。FITS規約に従わない場合には、SMOKA側で正しいFITSに変換を行う。（e）データ転送用サーバが用意できない場合には、定期的にハードディスクドライブを郵送すること等で対応が可能である。（f）環境データがあるのが望ましいが必須ではない。

がSTARS2上で一時誤って42か月に設定されてしまうなど、ハワイ観測所側で混乱をきたし、事態が収拾して安定した運用に至るまで1か月以上を費した。なお、公開を取り下げたIRC観測データに対するSMOKAでのデータ請求がなかったことをログで確認した。

(4) 2019年5月 MITSuME データ公開取下げ

既に公開されている2018年3月12日から2018年4月30日の天体M（本論文では明示しない）の観測データを非公開にすると東工大河合研から連絡があった。すぐに、MTAの81,144フレーム、MTOの6,531フレームの観測データとHDIおよびQLI、サムネイルなどの公開を取り下げた。2018年11月30日までの天体Mの観測データの公開は2019年12月1日とするとのことであった。しかし、東工大河合研側データベース（MITSuMEの観測データのPUBLICTIMEを設定している）での当該観測データのPUBLICTIMEの設定変更が行われるまでに1か月ほど要し、SMOKAではその期間は念のため当該天体以外の観測データについても公開作業を行わず、公開が遅れることとなった。PUBLICTIMEの延期はその後さらに2回行われ、2019年11月には2020年6月1日に延期（対象フレーム数は合計MTA 438,276フレーム、MTO 77,555フレーム）、2020年5月には2021年5月1日に延期（対象フレームは合計MTA 664,426フレーム、MTO 204,996フレーム）されたが、いずれも対象フレームが公開されるよりも前に連絡があり、深刻な事態にはならなかった。2021年4月には2018年3月12日以降の天体Mのデータは当分公開しないという決定が東工大河合研側でなされ、それらのPUBLICFLAGを‘否’にする作業が東工大河合研側で行われてSMOKA側はそれに追従して運用を進めている。なお、公開を取り下げた観測データに対するSMOKAでのデータ請求がなかったことをログで確認した。

(5) 2020年12月 IRD データ公開取下げ

IRDのすばる戦略枠プログラム（SSP）（以降、IRD-SSP [17]）の観測データの観測者占有期間が24か月になったとの連絡がハワイ観測所からあり、既に公開されているDATA-TYP = ‘OBJECT’の966フレームについて即刻応急処置を行ったが、その時点ではそれ以外のどのDATA-TYPのデータを24か月後にするかが定まっておらず、ようやく4日後に‘ACQUISITION’と‘DARK’あわせて214フレームも対象にするとの連絡があり、それらの対処も済ませた。しかし、STARS2（すばる望遠鏡の観測データのPUBLICTIMEが設定されている）でのすべてのIRD-SSPデータの設定変更が行われるまでに2か月間ほど要し、その後ようやく安定した運用に至ることができた。その間一時、IRD-

SSPデータ以外の一部のIRDデータの公開作業を控えるなどの影響があった。なお、公開を取り下げたIRD-SSPデータに対するSMOKAでのデータ請求がなかったことをログで確認した。

(3), (4), (5)の「データ公開取下げ」に関しては観測所内の連絡体制（公開時期延期という決定事項の周知徹底；現場まで伝わっているか？）と迅速な行動（データベースの設定など；現場がすぐに動くか？）が重要であるが、最近でも発生していることから見て、根本的な改善には至っていないものと思われる。

以上の他に以下のような問題も生じた。

(6) 公開されるべきデータの滞留

すばる望遠鏡の観測装置では、山頂で得られた観測データの全てがすぐにSTARSあるいはSTARS2に入るとは限らず、どこかで滞留していることもある。観測からかなりの期間が過ぎてからSTARSあるいはSTARS2に送られた観測データもあり、SMOKAでの公開も所定の期日よりかなり遅くなることがあった。例えば、CIAのDATE-OBS = ‘2001-07-06’のデータ402フレームは、209フレームが所定の2003年1月10日に公開されたが、残る193フレームは2005年6月13日に公開された。しばらくSTARSへ送られないままであったとの連絡を受けている。また、CIAのDATE-OBS = ‘2003-07-10’のデータ1,025フレームのうち、88フレームはSTARSへ送られたのが2005年7月21日であり、SMOKAでの公開は2005年8月6日となった。これらのフレームではフレームIDが取得日時順ではなく、SMOKA側ではさらなる混乱をきたした。

(7) フレームIDのずれ・誤り・重複・前後

SMOKAではフレームIDをテーブルの主キーとしており、取得日時昇順を前提としている。しかし、現実にはそれがしばしば崩れている。例えば、前項のCIAのデータのうち、DATE-OBS = ‘2001-07-06’の402フレームでは、最小フレームIDはCIAA00011928、最大フレームIDはCIAA00096860であり、その間に他のDATE-OBSのデータが複雑に混じり込んでいる。また、CIAのDATE-OBS = ‘2003-07-10’の1,025フレームでは、最小フレームIDがCIAA00085903、最大フレームIDがCIA00148188であり、その間に他のDATE-OBSのデータが複雑に混り込んでいる。これだけだとSMOKAでの扱いが不便なだけかもしれないが、DATE-OBS = ‘2001-07-06’の観測データのうち8フレームが重複（FITSヘッダ中のフレームIDが違うだけでデータそのものは同じ）していることが判明し、混迷を極めた。また、MCSでもDATE-OBS = ‘2005-08-12’と‘2005-08-

14'の観測データでフレームIDの昇順が乱れているが、それだけでなく、ファイル名とFITSヘッダ中のフレームIDが異なり、そのファイルはFITSヘッダ中のフレームIDと同じ名のファイルと中身が同一であることが判明している。これらの重複データは片方を公開しないことで対処している。さらに、SUPでは、DATE-OBS = '2004-08-11'の観測データのうち1,140フレームで、フレームIDの1の位がずれている事例があった。フレームIDはその日の最初の値から増分していく仕組みで、最初の値がずれていたことから生じた事例である。SMOKAでは1の位をCCDチップID、10の位より上が同じフレームをショットとして扱っていたために不都合が生じた。すばる望遠鏡の場合はこのような問題はそれぞれの観測装置の黎明期に多く発生しており、やむをえない面もあるかもしれない。一方、MITSuMEのMTAでは、フレームID重複問題が発生したことがある。DATE-OBS = '2018-01-14' ~ '2018-01-27'のデータがDATE-OBS = '2017-11-16' ~ '2017-11-29'のデータのフレームIDを上書きしてしまった事例である。東工大河合研側で新しい方のデータのフレームID（テーブルおよびFITSファイルで）をつけなおして対処したが、收拾するまでに3か月程度の期間を要した。さらに、木曽観測所のKCCとKCDではフレームIDの重複がそれぞれ39組と13組発生しており、後で取得された方のフレームIDの後ろにSMOKA側で「A」を付加している。

以上で述べた以外の重大な問題に、東日本大震災（2011年3月11日）直後のハワイ・三鷹間などのネットワーク断絶と三鷹地区計画停電（数回実施された。夏過ぎまで可能性あり、常に備えておく必要があった）にともなう問題があったが、詳細にわたるため本論文では省略する。

6 SMOKAによる成果

本章では、SMOKAの利用状況について述べたうえで、SMOKAの利用者による天文学研究や教育活動の成果、研究成果の検証の実際について論ずる。

6.1 利用統計

SMOKAでは観測データの取得には利用登録が必要であり、年度ごとに登録を更新している（3.3節）。図9は、2002年度から2020年度の利用登録者数の推移を示している。年度ごとに更新をしているため、その年度の実質的な利用登録者数を反映している。過去10年あまりにわたって250名前後がSMOKAを利用している。

図10は月ごとの請求ファイル数の推移を3か月平均で示している。総請求ファイル数は20年ほどにわたって増加傾向にあり、最近では毎月数十万ファイル、データ量でいうと数テラバイトが請求されている。装置別では2016年ごろまではSUPの請求ファイル数が最も多かったが、その後はHSCの請求ファイル数が最も多くなっている。このことから、可視光広視野撮像データのアーカイブとしての価値の高さがうかがえる。

HSCについては、データ公開はSMOKAからだけではなく、HSCプロジェクトからも行われている。SMOKAでは全観測の生データを公開しているが、HSCプロジェクトではSSP [51]のHSC観測データ¹⁶⁾および共同利用のHSC観測データ（本論文執筆時現在、2014年分のみが公開されている¹⁷⁾）が、それ以降の観測データも公開される予定である[52]の較正処理済みデータを公開している。SMOKAに請求されたHSCデータのうち82%がSSP以外のデータである。残り

16) <https://hsc-release.mtk.nao.ac.jp/doc/>

17) <https://hscla.mtk.nao.ac.jp/doc/>

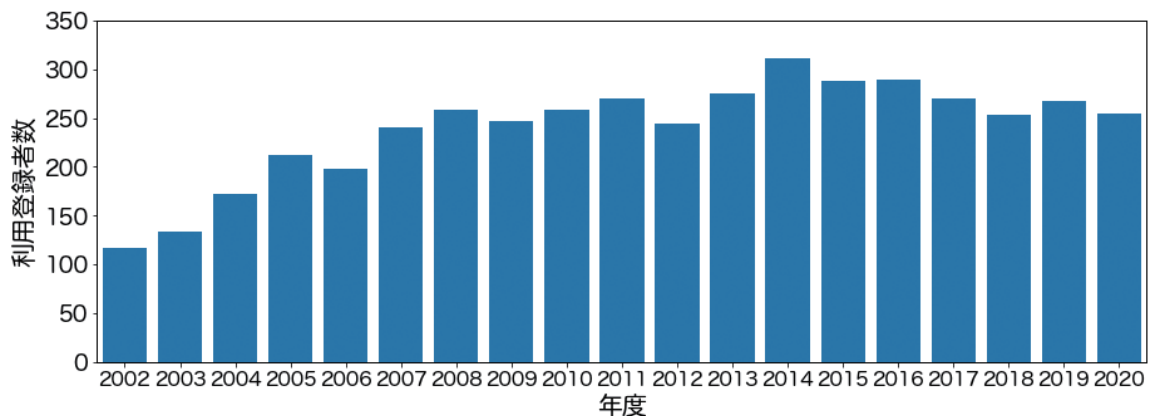


図9：各年度ごとのSMOKA利用登録者数の推移。

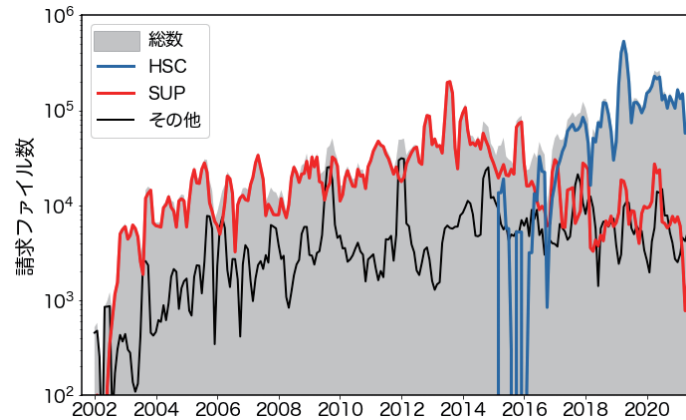


図10：請求ファイル数（3か月平均）の推移（縦軸は対数）。灰色の塗り潰しが請求ファイル数の総数であり，実線が観測装置別の請求ファイル数であり，赤がSUP，青がHSC，黒がその他の観測装置を示す。

の18%の約109万フレームはSSPのデータである。較正処理済みではない生データの需要も相当あることがうかがえる。

6.2 SMOKAによる成果

(1) 研究

SMOKA から取得した観測データを利用した研究論文（主要論文誌の査読論文）の年ごとの発表数の推移を図11に示した。総論文数は2021年7月1日現在で263本である。直近5年では平均的に毎年15本程度の論文が出版されている。装置別では，SUPの観測データを使った論文が最も多く，全体の71%を占めている。HSCの観測データ使用の論文も加えると74%にのぼり，可視光広視野撮像装置の観測データのアーカイブとしての価値の高さを示している。これら論文の分類（論文中でのSMOKA から取得した観測データの位置付けおよび研究分野）を図12に示した。SMOKA から取得した観測データが論文中で補助的に使われている場合が50%と最多であるが，単独あるいは主要な観測データとして用いられている場合も42%もの相当数があり，新規観測を遂行することなく（あるいはあまりそれに依存せず），アーカイブされた観測生データから科学的成果を産出できることを実証していること，および，観測データアーカイブ天文学にSMOKAが貢献していることが言える。また，研究分野としては銀河（近傍および遠方）と宇宙論で69%もの割合を占める。これはSUP/HSCの観測データを用いた論文数が多いことと相関している（図12の銀河（近傍および遠方）および宇宙論の論文の92%がSUP/HSCの観測データを使用しており，SMOKA から取得したSUP/HSCの観測データを使用した論文の86%が銀河（近傍および遠方）および宇宙論の分野の研究である）。

(2) 教育

高校や大学の教育活動においてSMOKA から取得した観測データが活用されている。伊藤ほか[53]で詳しく述べられているが，三重大学教育学部では，教員養成のための実験・実習の科目にて，恒星の温度推定（東京大学木曽観測所シュミット望遠鏡KCDの観測データ使用）や渦巻銀河の回転速度の推定（国立天文台岡山天体物理観測所188 cm望遠鏡SNGの観測データ使用）などの目的でSMOKA から取得した観測データが使われている。その他の活用例についても伊藤ほか[53]掲載の参考文献で述べられている。また，論文誌などに掲載されていない活用例を，インターネット上で「SMOKA 教育活動」などで検索して多数見出すことができる（例えば山田隆文¹⁸⁾）。

(3) 検証

SMOKA の役割の一つに研究成果の検証のための生データの提供があり，研究成果が画期的なものであればあるほどその重要性は高まると思われる。しかし，実際に行われた検証を論文などで目にすることはあまりない。SMOKA から取得した観測データを利用した検証（再解析）の例として，Ishiguroほか[54]がある。この研究は，Fuseほか[55]で用いたすばる望遠鏡の観測データをSMOKA から取得して再解析を行ったものである。

6.3 観測効率の向上

すばる望遠鏡の観測提案（プロポーザル）では，提案する天体が過去にすばる望遠鏡で観測されたかどうか

18) 天文分野におけるアーカイブデータの活用，<http://www.e-net.nara.jp/kyouka/index.cfm/20,400,c,html/400/20200123-110156.pdf>

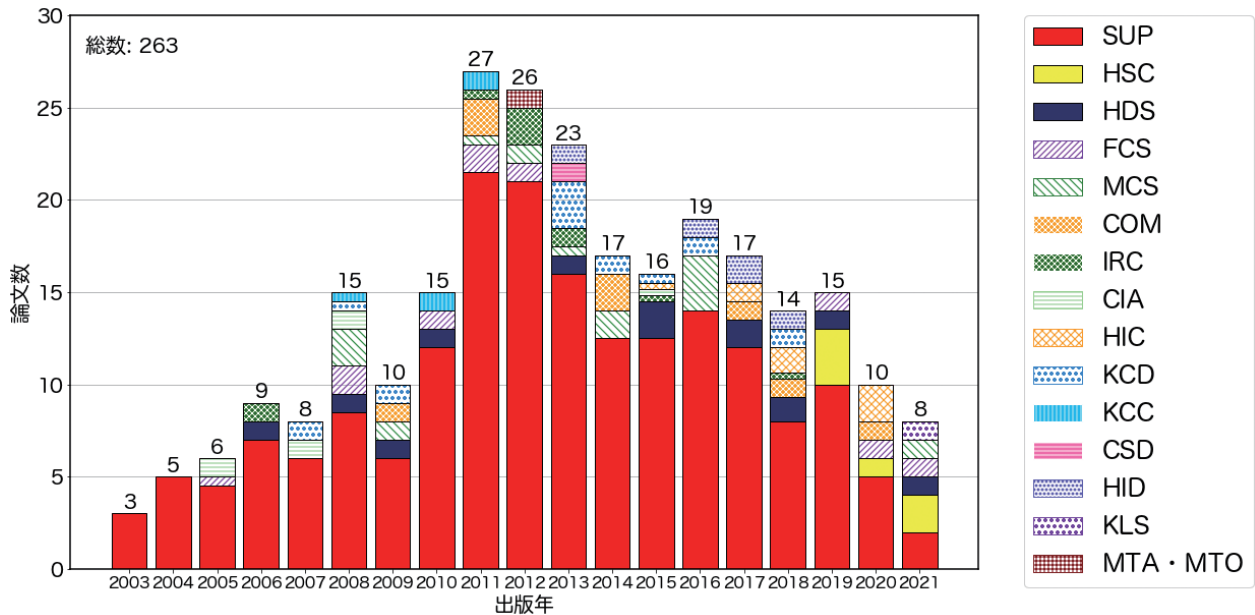


図11：SMOKA から取得した観測データを利用した主要査読論文数の推移（2021年7月1日現在）。各出版年のビンの上部にその年の総数を記した。観測装置別に色・パターンを分けて表示した。1本の論文で複数の観測装置の観測データが利用された場合には、その観測装置数の逆数を各観測装置の寄与とした。

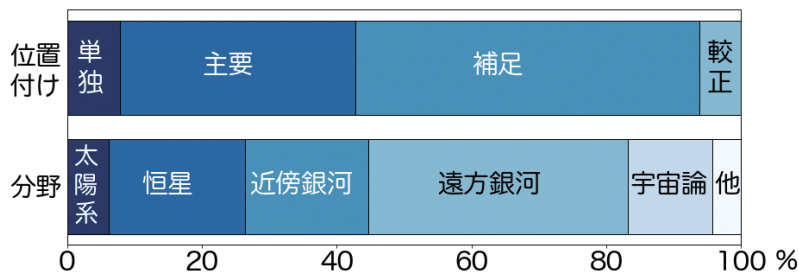


図12：SMOKA 論文の分類。SMOKA から取得した観測データを論文中でどのような位置づけで利用しているかの割合（上）と、論文の研究分野の割合（下）を示す。

かを SMOKA で確認し、過去に観測された場合には再度観測する必要性を記述することが義務づけられている¹⁹⁾。

これによってすばる望遠鏡の観測効率が向上しているものと思われるが、その効果は数値などで表すことができないのは残念である。

7 これからの観測データアーカイブに向けて

本章では観測データアーカイブに関する世界および日本の動向を概観し、今後のあり方について論ずる。次に、天文学の発展にともなう観測装置の進化について述べ、観測データアーカイブがそれに対応するため

にはどのようなことを考慮しておくべきかについて論ずる。そのうえで、これまで20年にわたる SMOKA の運用と開発を通じて得られた、これからの観測データアーカイブシステムのための提言を行う。

7.1 観測データアーカイブに関する世界の動向

ここでは、近年の観測データアーカイブに関する国外および国内の動向について概観する。

近年の観測データアーカイブを取り巻く環境は、SMOKA が運用を開始した頃に比べると大きく変化してきていると言える。

一つめは、データアーカイブからデータを取得して何かしらの科学的活動を行うのは、ごく普通の行為となってきたということである。これは1990年代より様々なデータアーカイブシステムが構築され、それらが保有するデータを利用することが一般化されてき

19) <https://www.subarutelescope.org/Observing/Proposals/policy.html>

たことを意味する。特にハッブル宇宙望遠鏡 (HST) やヨーロッパ南天天文台 (ESO) などの競争力の高い観測データについては、観測データアーカイブからのデータダウンロード量が当該観測提案グループのダウンロード量の数倍以上であることは10年以上前から指摘されており [56], また、何らかの形でアーカイブデータを用いた成果論文も全体の25%以上を占める (ハッブル宇宙望遠鏡の場合には50%以上)²⁰⁾ ²¹⁾ など、科学研究を進めるうえでなくてはならない重要な役割を担うようになってきている。国内の望遠鏡に目を移してみれば、SMOKAが扱うすばる望遠鏡の観測データについては、SMOKAから取得した観測データを利用した成果論文は観測当事者による論文数の10%程度となっており、欧米に比べればまだ低いものの、日本国内においても一定の役割を担っていると言える (6.2)。アーカイブデータを有効に活用することで、天文研究においてより多くの選択肢を持てることによって可能性が広がるという意識は強くなってきており、その中のコミュニティーを構成する人々の役割もより重要になってきていると言える (例えば, [57][58])。

但し、アーカイブデータ利用による成果の多くは処理済みの画像やカタログによるところが大きく、Sloan Digital Sky Survey [59] のようなサーベイデータのデータアーカイブが大きな役割を担っている。一方、生データから科学的成果にまでつなげるためには、データ解析に関する様々な環境が整っていることが必要であり、今後も状況を改善していく必要がある。処理済みデータに比べて成果獲得までには時間がかかってしまうため、限られた人的、物的リソースの制約の中で、どのように整備を進めていくかは、研究者コミュニティの考え方の整理も含めて、今後の大きな課題であると言える。

二つめは、科学成果に対するエビデンスとしてのデータの公開や再利用促進に対する動きが活発になったことである。特に、公的資金を投じて建設・運用されている観測所のデータについては、その要求は近年強くなる一方で、例えば、外部資金提供による観測装置開発や文部科学省による共同利用・共同研究拠点事業²²⁾ に対しても、データの公開に関する具体的計画を提出することが要求されるようになったり、国際的科学論文誌の中には、すでにデータ利用可能性 (data availability) に関する情報の記述を著者に要求するものも出てきている。国内外を問わず、この動きはオープンデータ・オープンサイエンスと呼ばれ、天文学に限らず、すべての自然および社会科学分野の学問研究に対して強化されてきており、そのための支援体制やデータインフラストラクチャの整備が当面の重要な課題である。特に、いわゆる「STAP細胞問題」などに代

表される「データ改ざんによる論文詐称」などの「科学における不正行為」が世界的に見ても多く発生しているとされる日本において [7], 不正を起こさせない、また、第三者による検証を容易にしておくことで不正行為を抑制する効果を生み出すことは重要な取り組みである [60]。また、画期的な科学的成果が発表された場合にも、その結果や解釈を独立に再現し検証をするためにも必要不可欠な仕組みとしてのデータアーカイブの存在はその意義を以前に比べて大きく増してきていると言える。

一方で、三つめの大きな変化として、近年の観測データ量の巨大化や複雑化も観測データアーカイブを考えるうえで大きな要素となってきている。IT技術の格段の進歩により、データ生産レートは増大の一途をたどっている。観測装置の広視野化やCMOSセンサーをはじめとする高速読み出しを可能とする検出器の実用化といった要素が大きく影響しているが、限られた予算に対応するためコストパフォーマンスの最大化を目指す動きが先鋭化してきている点も見逃せない。一方で、観測装置から観測データを取得・保存することはできても、アーカイブされた観測データを高速に世界中にインターネットを介して配信することは決して簡単な状況ではない。個々のアーカイブ利用者が、高仕様の計算機間の通信速度 (数10 Gbps) に匹敵するほどの通信速度でインターネット回線を占有することは現状では難しいと思われる。従って、今後は観測データ本体の移動はなるべく起こさないような観測データの使い方を実現していく必要があり、そのためのデータサーバ側でのデータ処理能力および機能の大きな進展が期待される。実際、そのための様々な取り組みが開始されている²³⁾。また、観測データの巨大化は中小望遠鏡に装着される観測装置の観測データに関しても確実に浸透してきている (7.2節)。このため、観測装置の計画・設計段階からデータ取得や解析とともに観測データアーカイブ構築についても考慮しておくことが重要になってきている。これは、観測装置の開発プロジェクトの経費におけるデータ解析やアーカイブに関する割合が増加せざるを得ない可能性があり、費用が大きすぎて、他の大きなプロジェクトのついでにできるような規模の計算機システムおよび運用体制では対応できないからである。観測データが生産され始めてからデータ解析や観測データアーカイブを考え

20) HSC データを用いた論文数統計サイト, <https://archive.stsci.edu/hst/bibliography/pubstat.html>

21) ESO のデータを用いた論文数に関する統計サイト, <http://www.eso.org/sci/php/libraries/telbibstats/archive.php>

22) 文部科学省による共同利用・共同研究拠点事業, https://www.mext.go.jp/a_menu/kyoten/

23) <https://www.sciserver.org/>

るようなことをしてしまうと、計画初期から考慮していた場合と比べより大きなコストがかかることも明らかであり [61], 観測データの巨大化によって観測データアーカイブ構築等のデータ基盤の足下が揺らいでしまうと、早期の科学的成果の獲得にも大きな影響が出てしまうであろう。プロジェクト経費の中での観測データアーカイブなどデータインフラストラクチャに関する経費の割合を慎重に考慮しないといけない場合が出てきていると言える。

世界的に見れば、SKA²⁴⁾ や Rubin 観測所 [62]²⁵⁾ をはじめとする次世代観測所は膨大な量の観測データを生産する計画となっており、その観測データを用いた科学的成果を得るためには相応の規模の計算機システムが必ず必要になってくる。観測データの効率的利用のためには観測データアーカイブは必須の機能ではあるが、その実現のためには相応の費用を覚悟せねばならず、また、最新の IT 技術を導入するための技術・人材の確保も必要であり、また、どのような形で効率的かつ科学的正当性も保ちながら観測データ提供を行えるようにするかは、コミュニティ規模で慎重に考慮すべき時期に入ってきていると言える。

最後に、観測データそのものもそうであるが、観測所を構築・運用するためのプロジェクトの巨大化の影響も無視できない。多くの人員と長期間の建設によって運用にこぎ着けた観測プロジェクトに関わる科学者たちには特にそうであるが、自分たちだけが必要なデータを利用できるようにしようとすることはある程度当然の帰結として起こりやすい。ある程度の優先利用期間を経過した後は観測データが公開され、より多くの研究者が様々な研究を進められるようになることが重要であるが、この公開時期が必要以上に遅延しないようにするための取り組みも必要である。そのためには、巨大プロジェクトの中においてもデータ解析パイプラインや観測データアーカイブを早期に構築・運用し、データの優先使用期間内により多くの科学的成果が出せるような仕組みを機能させることが必要不可欠である。観測データに関する出口戦略を明確にもって臨まない場合、この点の消化不良から最後には「観測データの囲い込み」とみられても仕方がない状況になることも考えられ、そのような状況は是非とも避けたいところである。

以上のように観測データアーカイブを取り巻く環境は今後も変化を続けていくことと思うが、日本が中心的に運用する観測所や望遠鏡による観測データは日本の財産であると同時に、世界の天文学にとっても重要な財産となることを意識する必要がある。これらの観測データを観測データアーカイブシステムを介することで世界に向けて公開することで、世界の天文学コ

ミュニティーに対して日本から重要な寄与ができるのであるという意識をもつことは、日本の天文学研究者にとって必要なことではないかと考える。

7.2 観測装置の進化

天文学の発展にともなって新たな観測装置が次々と登場している [63]。例えばすばる望遠鏡では、CHARIS や VAMPIRES など [64][65] 新規観測装置が登場し、また、既存の観測装置も新しい機能を加え、例えば、HIDES が光ファイバー導入方式を採用した HIDES-F になり [40]、FOCAS が面分光機能を採用した FOCAS-IFU になった [66]。MOIRCS は相関二重サンプリングの全てのサンプリングデータを生データとして保存するために MEF を採用した [42][43]。

これまでになかった発想に基づく観測装置も多く登場し、ある装置はデータ構造が複雑になり、ある装置はデータ取得頻度が高速化している。SMOKA が現在公開している観測データの FITS ファイルの構造を表 6.7 にまとめたが、今後ますます複雑な構造のものが増えていくであろう。

高速データ取得の観測装置の中で、木曽観測所の Tomo-e Gozen [67] やせいめい望遠鏡の TriCCS²⁶⁾ などの CMOS カメラの観測データが今後 SMOKA で公開される予定となっている。前者は 84 個の 2k × 1k CMOS 検出器から 2 fps でデータが生成され、後者は 3 バンドの 2k × 1k CMOS 検出器から最大 98 fps でデータが生成される予定である。いずれも膨大な数と量の観測データが生成され、それらのデータ公開にはデータの転送や保存、管理方法（グループ化）と検索方法など多くの技術的課題はもちろん、経費的な課題も乗り越える必要がある。

また、SMOKA は 1 章などで述べたように主に観測生データを扱うシステムであるが、このような高速データ取得における膨大な生データをどう扱うべきなのか悩ましい課題である。観測装置のデータ取得系から出力されたデータ（＝生データ）を足し合せたものと、データ取得系内部で積分されたうえで出力されたデータ（積分しないデータを別途出力する場合もある。これも生データである）をどう考えるのか。保存する情報量と保存するための経費や手間との兼ね合いを検討していかなければならないであろう。

また、すばる望遠鏡の PFS [68] では、各 3 バンドの 4 分光器モジュールから生成される 12 個の FITS ファイルに加えて、ファイバ配置を測定する Metrology

24) <https://www.skatelescope.org/>

25) <https://www.lsst.org/scientists/keynumbers>

26) <http://www.kusastro.kyoto-u.ac.jp/~kazuya/p-triccs>

表6：観測データのFITS ファイルの構造（2次元でFITS 拡張を持たないものは除く）.

装置	構造	詳細
COM	4次元	撮像：空間方向2次元+時間軸+検出器の番号（1のみ） 分光：空間方向1次元+波長方向1次元+時間軸+検出器の番号（低分散では1のみ）
CRS	2次元+BTE ^a	
FCS	2次元（+ATE ^b ）	ATE は一部のデータ
FMS	2次元+ATE（+IE ^c ）	IE は一部
HDS	2次元+ATE	ATE はスペクトルのフォーマット（回折次数，波長，CCD 上での位置）
MIR	3次元+ATE	空間方向2次元+時間軸，ATE は動作パターン
MCS	2次元（+2N IE）	IE は2016年5月20日以降のデータ Nは相関二重サンプリングの回数（執筆時現在で最大35）
MMZ	3次元	空間方向の2次元+スタッキング1次元のイメージデータ

^a BTE: Binary Table Extension^b ATE: Ascii Table Extension^c IE: Image Extension

表7：複数フレームが組になっている観測データ.

装置	組となるフレーム
モザイクカメラ	
FCS	2枚のフレーム
HDS	2枚のフレーム
HSC	104枚のサイエンス用フレーム+8枚のフォーカス用フレーム
IRD	2枚のフレーム
SUP	10枚のフレーム
MCS	2枚のフレーム
KWF	8枚または4枚のフレーム
HID	3枚のフレーム（2007年11月まではシングルCCD）
HWP	2枚のフレーム
多色同時観測装置	
MIR	近赤外線と中間赤外線の同時撮像フレーム
SWS	Blue（0.9–1.4 μm ）と Red（1.4–2.5 μm ）の2波長帯での撮像または分光フレーム
MCT	g'_2 , r'_2 , z_{s2} の3色同時（個別取得も可）撮像フレーム
HNR	可視光線と近赤外線の2波長帯の撮像，分光，偏光フレーム
MTA	g' , R_C , I_C の3色同時撮像フレーム
MTO	g' , R_C , I_C の3色同時撮像フレーム
NIC	J , H , K_S の3色同時撮像フレーム
撮像・分光同時観測装置	
COM	中間赤外線での同時撮像・分光フレーム

Camera のデータ，Auto Guider のデータなど異質なデータが組になって生成される予定である．これまでの観測装置でも分光データとスリットビューデータのように関連付けて扱うべきデータの組があったが，より高度で複雑な関連付けが必要になるであろう．観測データの公開にあたってはデータ間の関連付けが重要になるであろう．

すばる望遠鏡には，AO（補償光学）と組み合わせた

観測装置が多く台頭している [64]．研究成果の検証のためには AO の制御データをある程度保存すべきであるが，大量のデータをどう抽出・選択してどう残すべきか，現時点では全く未定である．

これらの新たな観測装置の観測データを公開するためには，まず，複雑なデータ構造やデータ間の関連付けに対応できる柔軟なアーカイブシステムの構築が必要である．また，膨大なデータ量に対応できる容量の

大きなシステムである必要があり、経費面に加えて消費電力（必要な冷房能力にも対応する）も考慮し、磁気ディスクだけでなく磁気テープライブラリの活用もはかるべきであろう。

いずれにせよ、今後の観測装置の動向を十分に把握したうえで観測データアーカイブシステムの検討と設計がますます重要になっていくであろう。

7.3 これからの観測データアーカイブへの提言

これまで述べてきたように、SMOKAの20年間には様々な課題が立ち塞がり、その多くを乗り越えてきた。SMOKAが今後も長く継続できるのかについては、現時点では不確定な要素が多いが、それにこだわることなく、これからの観測データアーカイブについて論ずる。

観測データアーカイブの実現にとって最も重要なことは、観測データを生み出す源（観測所や観測装置グループ）の「観測データを公開する」という強い意思と使命感である。7.1節で述べたように、観測データの公開は時代の流れであり社会の要請でもあるが、観測データ生成源に相応の意思と決意がなければその意義を十分に実現することは難しいであろう。もちろん多くの研究者が天文学研究における観測データの重要性を理解し、観測データを大切に作る姿勢をもつことも欠かせない要素である。

具体的に重視すべきことは本論文で述べてきたSMOKAの方針と活動で端的に示されていよう。すなわち、SMOKAは1章で述べた意義目的を実現すべく、「できるだけ多くの観測データをできるだけ早く」公開することを第一の活動目標とし、さらに、観測データの天文学的意義と可用性を高めるべく、

- ・ FITSヘッダ中の情報を修正し、また、欠落情報を補って観測データの自己記述性を高める、
- ・ 環境データなどの付随情報や各種資料を用意して、観測データの解析などを支援する、
- ・ クイックルック画像（サムネイルやショットイメージなど）や観測データの品質評価など様々な検索方法を提供して必要な観測データに迅速に至れるようにする、

ことの実現を目指し、そのために

- ・ 観測装置の開発段階からデータフォーマットやFITSヘッダの策定に参画する、

- ・ 各観測所との連携の下、観測データ自体やデータ解析などに関する様々な情報を収集する、

- ・ 望遠鏡や観測装置、天文学に熟達した目で観測データを取り扱い、評価判断する、

活動をも行ってきたのである。

これからの観測データアーカイブにあってもこれらを重視することをSMOKAからの提言としたい。そしてこの提言はとりもなおさず、SMOKAの20年にわたる開発と運用を通じて得られた成果だと言えるであろう。

しかしながら、人員や予算や組織などの体制的問題、様々な技術的問題、あるいは未知の問題による制約が想像され、これまで行ってきたSMOKAの活動全てがこれからの観測データアーカイブでも実行できるとは限らない。適切な取捨選択を行いつつ、天文学や観測装置の発展に対応していく観測データアーカイブシステムを築いていくことが求められよう。

8 まとめ

本論文では20年にわたるSMOKAの活動を論文6以降の時期を中心にまとめ、その間に直面した技術的問題（4章）、運用面での問題（5章）についてそれぞれへの対応とその評価、なおも残された検討課題などについて論じた。7章では観測データに関する最近の動向や観測装置の発達を論じ、これからの観測データアーカイブシステムに向けた提言をまとめた。

様々な困難を乗り越え、試行錯誤や紆余曲折を経て20年にわたる運用を続けてきているSMOKAであるが、今後も長く継続していくためには幾多の技術的・運用的な課題を乗り越えていかねばならず、それは安易な道ではない。しかしながら、昨今のデータ重視の流れに加え、研究偽装の多発や一面的・即席的な成果ばかりを求める風潮に鑑みると観測生データを基本としたデータアーカイブは、天文学の健全で多様の永続的な発展のために必要不可欠なものであり、観測所や関係者や利用者などとともにその努力を続けていかねばなるまい。

本論文はそのための検討材料、あるいは、道標としての役割を担うことを願っている。

謝辞:

SMOKAの20年間にわたる開発と運用に携った論文1, 2, 3, 4, 5, 6の著者の方々, および, 吉田鉄生, 樋口祐一, 大越克也, 小林正和, 小池美知太郎, 空華智子, 田口弘子の各氏に感謝する. また, 種々の支援をいただいた古澤久徳氏の助力に深く感謝する. 長年にわたりSMOKAの運用に欠かせぬ連携をいただいた, ハワイ観測所, ハワイ観測所岡山分室(旧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, 兵庫県立大学西はりま天文台, 京都大学岡山天文台, の皆様に感謝する. 本論文作成の際に貴重な助言をいただいた小澤武揚氏に感謝する. 匿名の査読者には多くの貴重なご指摘ご意見をいただいた. ここで厚く感謝する.

SMOKAはその開発と運用の経費の大部分を国立天文台天文学データ解析計算センター, および, 天文データセンターによって賄われた.

参考文献

- [1] 馬場 肇, 他: 2002, 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**, 23–26.
- [2] 山本直孝, 他: 2003, 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.
- [3] 榎 基宏, 他: 2004, 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.
- [4] 出田 誠, 他: 2005, 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.
- [5] 山田善彦, 他: 2009, すばる望遠鏡公開データアーカイブシステムの開発5, 国立天文台報, **12**, 53–78.
- [6] 野田祥代, 他: 2012, すばる望遠鏡公開データアーカイブシステムの開発6, 国立天文台報, **14**, 35–61.
- [7] 松澤孝明: 2014, 諸外国における国家研究公正システム(1) 基本構造モデルと類型化の考え方, 情報管理, **56**, 697–711.
- [8] Barbon, R., et al.: 1999, The Asiago Supernova Catalogue - 10 years after, *A&AS*, **139**, 531–536.
- [9] Groff, T. D., et al.: 2015, The CHARIS IFS for high contrast imaging at Subaru, *Proc. SPIE*, **9605**, 96051C.
- [10] Groff, T. D., et al.: 2016, Laboratory testing and performance verification of the CHARIS integral field spectrograph, *Proc. SPIE*, **9908**, 99080O.
- [11] Brandt, T. D., et al.: 2017, Data reduction pipeline for the CHARIS integral-field spectrograph I: detector readout calibration and data cube extraction, *JATIS*, **3**, 048002.
- [12] Miyazaki, S., et al.: 2018, Hyper Suprime-Cam: System design and verification of image quality, *PASJ*, **70**, S1.
- [13] Komiyama, Y., et al.: 2018, Hyper Suprime-Cam: Camera dewar design, *PASJ*, **70**, S2.
- [14] Kawanomoto, S., et al.: 2018, Hyper Suprime-Cam: Filters, *PASJ*, **70**, 66.
- [15] Furusawa, H., et al.: 2018, The on-site quality-assurance system for Hyper Suprime-Cam: OSQAH, *PASJ*, **70**, S3.
- [16] Tamura, M., et al.: 2012, Infrared Doppler instrument for the Subaru Telescope (IRD), *Proc. SPIE*, **8446**, 84461T.
- [17] Kotani, T., et al.: 2014, Infrared Doppler instrument (IRD) for the Subaru telescope to search for Earth-like planets around nearby M-dwarfs, *Proc. SPIE*, **9147**, 914714.
- [18] Kotani, T., et al.: 2018, The infrared Doppler (IRD) instrument for the Subaru telescope: instrument description and commissioning results, *Proc. SPIE*, **10702**, 1070211.
- [19] Motohara, K., et al.: 2014, Development of a simultaneous two-color near-infrared multi-object spectrograph SWIMS for the TAO 6.5-m telescope, *Proc. SPIE*, **9147**, 91476K.
- [20] Motohara, K., et al.: 2016, NIR camera and spectrograph SWIMS for TAO 6.5m telescope: overview and development status, *Proc. SPIE*, **9908**, 99083U.
- [21] Konishi, M., et al.: 2018, Development status of the simultaneous two-color near-infrared multi-object spectrograph SWIMS for the TAO 6.5m telescope, *Proc. SPIE*, **10702**, 1070226.
- [22] Konishi, M., et al.: 2020, The University of Tokyo Atacama Observatory 6.5m telescope: On-sky performance of the near-infrared instrument SWIMS on the Subaru telescope, *Proc. SPIE*, **11447**, 114475H.
- [23] Miyata, T., et al.: 2010, Development of a new mid-infrared instrument for the TAO 6.5-m Telescope, *Proc. SPIE*, **7735**, 77353P.
- [24] Kamizuka, T., et al.: 2012, Development of MIMIZUKU: a mid-infrared multi-field imager for 6.5-m TAO telescope, *Proc. SPIE*, **8446**, 84466P.

- [25] Kamizuka, T., et al.: 2014, Revised specifications and current development status of MIMIZUKU: the mid-infrared instrument for the TAO 6.5-m telescope, *Proc. SPIE*, **9147**, 91473C.
- [26] Kamizuka, T., et al.: 2016, Development status of the mid-infrared two-field camera and spectrograph MIMIZUKU for the TAO 6.5-m Telescope, *Proc. SPIE*, **9908**, 99083W.
- [27] Kamizuka, T., et al.: 2018, Laboratory performance evaluation of the mid-infrared camera and spectrograph MIMIZUKU for the TAO 6.5-m telescope, *Proc. SPIE*, **10702**, 107022H.
- [28] Kamizuka, T., et al.: 2020, The University of Tokyo Atacama Observatory 6.5m telescope: On-sky performance evaluations of the mid-infrared instrument MIMIZUKU on the Subaru telescope, *Proc. SPIE*, **11447**, 114475X.
- [29] Narita, N., et al.: 2015, MuSCAT: a multicolor simultaneous camera for studying atmospheres of transiting exoplanets, *JATIS*, **1**, 045001.
- [30] Yoshida, S., et al.: 1995, CCD Cameras for the Kiso 105 cm Schmidt Telescope, in *IAU Colloq. 148: The Future Utilisation of Schmidt Telescopes*, eds. J. Chapman, R. Cannon, S. Harrison, and B. Hidayat, volume 84 of Astronomical Society of the Pacific Conference Series, 33–37.
- [31] Itoh, N., et al.: 2001, A very wide-field CCD camera for Kiso Schmidt telescope, *PNAOJ*, **6**, 41–48.
- [32] Sako, S., et al.: 2012, KWFC: four square degrees camera for the Kiso Schmidt Telescope, *Proc. SPIE*, **8446**, 84466L.
- [33] Sakimoto, K., et al.: 2012, An optical and near-infrared multipurpose instrument HONIR, *Proc. SPIE*, **8446**, 844673.
- [34] Akitaya, H., et al.: 2014, HONIR: an optical and near-infrared simultaneous imager, spectrograph, and polarimeter for the 1.5-m Kanata telescope, *Proc. SPIE*, **9147**, 91474O.
- [35] Kotani, T., et al.: 2005, MITSuME—Multicolor Imaging Telescopes for Survey and Monstrous Explosions, *NCimC*, **28**, 755–758.
- [36] 石黒正晃, 他: 2011, 西はりま天文台近赤外線カメラ NIC の性能評価, 西はりま天文台年報, **21**, 13–42.
- [37] 高橋 隼, 禪野孝広, 石黒正晃: 2013, 西はりま天文台近赤外線カメラ NIC 限界等級の再測定, 兵庫県立大学天文科学センター紀要, **1**, 17–22.
- [38] Matsubayashi, K., et al.: 2019, KOOLS-IFU: Kyoto Okayama Optical Low-dispersion Spectrograph with optical-fiber Integral Field Unit, *PASJ*, **71**, 102.
- [39] Horaguchi, T., et al.: 1999, An Astronomical Data Archive System with a Java-Based User Interface, *PASJ*, **51**, 693–701.
- [40] Kambe, E., et al.: 2013, A Fiber Link between the Okayama 188-cm Telescope and the High-Dispersion Spectrograph, HIDES, *PASJ*, **65**, 15.
- [41] 小澤友彦, 他: 2015, 全天モニタ画像公開システムの開発, 国立天文台報, **17**, 1–17.
- [42] Walawender, J., et al.: 2016, The nuMOIRCS project: detector upgrade overview and early commissioning results, *Proc. SPIE*, **9908**, 99082G.
- [43] Fabricius, M., et al.: 2016, Detector upgrade of Subaru's Multi-object Infrared Camera and Spectrograph (MOIRCS), *Proc. SPIE*, **9908**, 990828.
- [44] 青木 勉, 他, SMOKA グループ: 2021, 東京大学木曾観測所写真乾板のデジタル化, 天文月報, **114**, 523–533.
- [45] Guillochon, J., et al.: 2017, An Open Catalog for Supernova Data, *ApJ*, **835**, 64.
- [46] Bertin, E., Arnouts, S.: 1996, SExtractor: Software for source extraction, *A&AS*, **117**, 393–404.
- [47] Chambers, K. C., et al.: 2016, The Pan-STARRS1 Surveys, arXiv:1612.05560.
- [48] Astropy Collaboration, Robitaille, T. P., et al.: 2013, Astropy: A community Python package for astronomy, *A&A*, **558**, A33.
- [49] Astropy Collaboration, Price-Whelan, A. M., et al., and Astropy Contributors: 2018, The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package, *AJ*, **156**, 123.
- [50] Tamura, M.: 2009, Subaru Strategic Exploration of Exoplanets and Disks with HiCIAO/AO188 (SEEDS), *AIP Conf. Proc.*, **1158**, 11–16.
- [51] Aihara, H., et al.: 2018, The Hyper Suprime-Cam SSP Survey: Overview and survey design, *PASJ*, **70**, S4.
- [52] Tanaka, M., et al.: 2021, Hyper Suprime-Cam Legacy Archive, *PASJ*, **73**, 735–746.
- [53] 伊藤信成, 西浦慎悟, 山縣朋彦: 2021, SMOKA の教育への活用 教員養成系学部での事例紹介, 天文月報, **114**, 396–404.
- [54] Ishiguro, M., et al.: 2009, 2006 Fragmentation of Comet 73P/Schwassmann-Wachmann 3B observed with Subaru/Suprime-Cam, *Icarus*, **203**, 560–570.

- [55] Fuse, T., et al.: 2007, Observations of Fragments Split from Nucleus B of Comet 73P/Schwassmann-Wachmann 3 with Subaru Telescope, *PASJ*, **59**, 381–386.
- [56] Beagrie, N.: 2006, Digital curation for science, digital libraries, and individuals, *International Journal of Digital Curation*, **1**, 3–16.
- [57] Vaduvescu, O., Curelaru, L., Popescu, M.: 2020, Mega-Archive and the EURONEAR tools for data mining world astronomical images, *Astronomy and Computing*, **30**, 100356.
- [58] Shaw, R. A., et al.: 2018, Enabling new science with MAST community contributed data collections, *Proc. SPIE*, **10704**, 1070414.
- [59] Gunn, J. E., Knapp, G. R.: 1993, The Sloan Digital Sky Survey, *ASP Conf. Proc.*, **43**, Sky Surveys: Protostars to Protogalaxies, ed B.T. Sofier (San Francisco: ASP), 267–279.
- [60] 林 和弘: 2016, オープンサイエンス時代の研究公正, 情報の科学と技術, **66**, 98–102.
- [61] Berriman, G. B., et al.: 2015, Data and Metadata Management at the Keck Observatory Archive, *ASP Conf. Ser.*, **495**, 535–538.
- [62] Ivezić, Ž., et al.: 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, *ApJ*, **873**, 111.
- [63] Ground-based and Airborne Instrumentation for Astronomy VIII, eds. C. J. Evans, J.J. Bryant, and K. Motohara, *SPIE Proceedings*, **11447**, (2020).
- [64] Hattori, T., et al.: 2020, Instrumentation at the Subaru Telescope, in Ground-based and Airborne Instrumentation for Astronomy VIII, eds. C. J. Evans, J.J. Bryant, and K. Motohara, *Proc. SPIE*, **11447**, 1144702.
- [65] Norris, B., et al.: 2015, The VAMPIRES instrument: imaging the innermost regions of protoplanetary discs with polarimetric interferometry, *MNRAS*, **447**, 2894–2906.
- [66] Ozaki, S., et al.: 2020, Integral field unit for the existing imaging and spectroscopy instrument, FOCAS, *PASJ*, **72**, 97.
- [67] Sako, S., et al.: 2018, The Tomo-e Gozen wide field CMOS camera for the Kiso Schmidt telescope, *Proc. SPIE*, **10702**, 107020J.
- [68] Tamura, N., et al.: 2020, Prime Focus Spectrograph (PFS) for the Subaru telescope: a next-generation facility instrument of the Subaru telescope has started coming, *Proc. SPIE*, **11447**, 1144713.