# SMOKA / Tomo-e Gozen データ公開システムの開発

中島康,小澤武揚,小野里宏樹,森由貴\*,市川伸一

(2022年7月20日受付;2022年8月26日受理)

# **Development of SMOKA / Tomo-e Gozen Data Archive System**

Yasushi NAKAJIMA, Takeaki OZAWA, Hiroki ONOZATO, Yuki MORI\*, Shin-ichi ICHIKAWA

#### Abstract

We have developed and started the operation of a data archive system for the stacked data, made by stacking the "astronomical movie data" along the time axis, of the Tomo-e Gozen optical wide field camera for the Kiso Schmidt telescope. We have developed this system as a derivative of SMOKA (Subaru-Mitaka-Okayama-Kiso Archive System), a data archive system we have developed and are currently operating. In developing this system, we have taken measures to cope with some technical challenges that we recognized in SMOKA. In this paper, we report the development and the configuration of this system and discuss the results of the measures and the issues for the future.

#### 概要

我々は、木曽観測所シュミット望遠鏡に搭載の広視野可視光撮像装置 Tomo-e Gozen で取得・作成された動 画データの時間軸方向を足し合わせて作成されたスタック済みデータを公開するデータアーカイブシステムを 開発し、その運用を開始した.本システムは、我々が開発し運用中のデータアーカイブシステム SMOKA の派 生システムとして構築し、その開発にあたって、SMOKA で認識していた課題に対する技術的な挑戦を取り入 れた.本論文では、本システムの開発および構成について報告し、技術的挑戦の結果について論ずるとともに、 今後に向けた課題についても論ずる.

\* 東京大学(The University of Tokyo)



この記事はクリエイティブ・コモンズ[表示 4.0 国際]ライセンスの下に提供されています。 https://creativecommons.org/licenses/by/4.0/deed.ja

# 1 はじめに

天文学において観測データは根元的な研究基盤であ る.天体現象は定常的ではなく,観測装置や条件も一 般的には観測毎に異なるため,個々の観測データは唯 一無二であると言える.そのため,観測データを長期 間保管し世界中の研究者に提供する天文データアーカ イブシステムは天文学において極めて重要な役割を果 たしている.その使命のもと,SMOKAは20年以上に わたって運用を続け,日本の光学赤外線観測データを 国内外の研究者に提供してきた(論文1-7).

Tomo-e Gozen (本論文では以降 Tomo-e と略する) は東京大学木曽観測所105 cm シュミット望遠鏡に搭 載された可視光モザイク CMOS カメラであり、超広視 野の動画データを取得できる画期的な観測装置である [8]. 焦点面に84個のCMOS検出器(有効画素数2000× 1128×84=1億9千万画素)を配置して約20平方度の視 野を有し、毎秒2コマの動画データ取得が可能である。 その特色を生かし、天の広い領域をカバーする大規模 サーベイ観測および重力波対応天体などの突発現象の 追観測、時間軸天文学のための小中規模のキャンペー ン観測などが2019年10月1日より行われている.大規 模サーベイ観測として全天サーベイ (All-Sky Survey), 高頻度サーベイ (High-Cadence Survey),および超新星 サーベイ (Supernova Survey) の三つの観測プロジェ クトが実施されており、この三つのプロジェクトの観 測が全体の観測フレーム数の大部分(3章で述べる木 曽観測所から転送されてきた2019-10-01から2021-05-14までのデータでは95.7%)を占める(以降,これら を3大サーベイプロジェクトとよぶ. なお、プロジェ クト名やどのフレームがどのプロジェクトに属する かについて本論文ではFITS ヘッダに記載されている PROJECTの値を採ることにする).

Tomo-eは一晩約30TBの大容量の観測データ(生 データ)を生成するため、木曽観測所ではその全てを 長期保存することは不可能と考え、生データは観測後 5日で消去することとし、一次処理(ダーク補正、フ ラット補正、位置較正)およびスタック処理を行った データ(3章参照、本論文ではスタック済みデータとよ ぶ)を長期保存することとした.

SMOKAではこれまで生データ<sup>1)</sup>の公開を行ってき たが、Tomo-eについてはスタック済みデータをまず公 開することとした. Tomo-eのスタック済みデータが 含蓄する豊富で貴重な天文学的情報は後世に残して活 用すべきであり、国立天文台のSMOKAではそれらを 永続的に公開できる可能性が高いと考えたからである. その公開のルールとして, Tomo-e Gozen コラボレー ション(Tomo-e Gozenを利用する多機関の研究者から 成る共同体)によって以下が定められている.

- ・観測期間を二つに分け、それぞれに異なるルールを 適用する。
- ・第1段階のルールとして、観測日2019-10-01から
  2020-03-31までの3大サーベイプロジェクトのス タック済みデータについては観測日の2年後(730日 後と定義する)に公開を開始する(このルールによ る新規データ公開は2022年3月末で終了した).
- ・第2段階のルールとして,観測日2020-04-01以降の3 大サーベイプロジェクトおよび観測日2019-10-01以 降のいくつかの小中規模のキャンペーン観測等(木 曽観測所側で適宜追加する)のスタック済みデータ については観測日の3年後(1095日後と定義する)に 公開を開始する.
- ・ブロックリストは設定しない.ブロックリストとは、
  その科学的価値から Tomo-e Gozen コラボレーションが専有するデータのリストのことである.

Tomo-eのスタック済みデータのFITS ファイルには 以下の特徴がある.

- 3次元データ(3次元目は時間)である生データと は異なり、2次元データである。
- (2) データ配列のデータ型は32ビット浮動小数点数型 である.
- (3) 画素数2000 × 1128であり、1ファイルの大きさは 9,040,320バイトである.ただし画素数の異なる例 外がわずかにある.
- (4) 位置較正が施され、WCSパラメータとしてSIP
  (Simple Imaging Polynomial) [9]で表現されている。
  ただし位置較正が施されていない例外がわずかにある。
- (5) FITS ヘッダが整備されており、キーワードが欠損 していたり、値が定義外のものであったりという ことがなく、現時点では修正の必要がない.
- (6) データ圧縮があまり効かない (gzipやbzip2で圧縮 して80%ほどになる程度).
- (7) スタック処理済みであるがデータが大量である(1 月あたり平均150万フレーム/13.5 TB).

Tomo-eのスタック済みデータは一次処理済みかつ

一部の観測装置については生データに加えて処理済みデータも公開している。

位置較正済みであるため、利用者にとっては活用面 で好条件を備えている一方、生データの公開を主目 的として運用中のSMOKAに組み入れるには少なから ぬ工夫と改造が必要である。そのため、全天モニタ 画像公開システム[10]や写真乾板アーカイブシステム [11]のようにSMOKAの派生システムとしてTomo-eの スタック済みデータ公開システム(以降、本システ ム)を別途構築することとした。そしてこの機会に SMOKAの弱点・課題として認識されている事項(論 文7)のうちのいくつかに対して技術的な挑戦を行い、 かつ、これまでの経験や資産を生かして短期間での開 発に挑むこととした。

前者の技術的な挑戦とは以下である.

- (A) SMOKAでは人手によって行っている期日に達し たデータの公開作業(論文7)を自動化する.
- (B) SMOKAで行っている、公開作業の際の未公開領域から公開領域へのデータのコピー(論文7)と、データ請求の際の利用者がアクセスできる領域へのデータのコピー(論文7)を行わず、作業時間とCPU負荷の軽減をはかるとともに磁気ディスク領域の節約をはかる。
- (C) SMOKAではピンポイント検索にHEALPixを利用 しているが(論文5),本システムではHEALPixを 利用せずに実現し,HEALPixインデックスの算出 にかかる手間と時間を省く.
- (D) データ請求時に Header Replacer による FITS ヘッ ダの修正(論文7)を行わないなど、できるだけ単 純なシステムとして構築し、他所への移植などが 容易に行えるようにする。

上記特徴(4)と(5)を備え、公開期日の十分前に スタック済みデータが転送済みで下準備や確認が行え るという有利な条件の下での挑戦ではあるが、今後の 観測データアーカイブ公開システムを考えるうえでの 貴重な経験になると考えている。

後者の経験や資産については、SMOKAのために開 発したソフトウエアを再利用することやSMOKA用に 用意されているハードウエアやOSなどを利用するこ とを意味する.

本論文はそのような状況の下での本システムの開発 について論じるものであり、上記の第1段階のデータ 公開ルールに基づく開発について述べ、最後にそれ以 降の展望や課題について論じる。

なお,上記特徴の例外などにあたる若干数のフレー ムについては,本システムでの第1段階のデータ公開 の対象外とした.

本論文では1フレームとは一つのFITSファイルを指

すものとする.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指すが、本論文では混乱を避け るためにその意味では用いないことにする.

# 2 システムの概要

本システムを構成するサーバおよびその中でのデー タの流れを図1に示す.本システムを担うサーバは, 転送用サーバ,ファイルサーバ,WEBサーバ,デー タベースサーバ,FTPサーバである.これらサーバの 仕様の概要を表1に示した.転送用サーバは,スタッ ク済みデータを木曽観測所からインターネット経由で 本システムに転送するために使用した.ファイルサー バはRAID構成された磁気ディスクを有し,木曽観測 所から転送されたスタック済みデータは最終的にこ



図1:SMOKA/Tomo-eシステムの全体像.本システムで用 いられているサーバと各データ(以下のABC)の流れを 示した.A:スタック済みデータ(4章参照).B:スタッ ク済みデータから抽出されたメタデータなどの情報.デー タベースサーバ(DB)に保存される(5章参照).C:ス タック済みデータから作成されたクイックルックデータ. WEBサーバの非公開の保存領域(QL)に保存される(7 章参照).各サーバの仕様は表1にまとめられている.

	WEB/データベース	FTP ファイル		転送用		
台数	1	1	8	1		
CPU	Intel Xeon CPU E5-2667 v4 3.20 GHz			Intel Xeon CPU E5-2640 v4 2.40 GHz		
CPU×コア数	2 CPU × 8 コア			2 CPU × 10 コア		
メモリ	32 GB		192 GB	128 GB		
OS	RedHat 7.9			CentOS 7.9		

表1:各サーバの主な仕様、ファイルサーバ8台のうち第1段階の公開で使用されるのは2台である.

こで保存される. なお, ファイルサーバは本システ ム専用ではなくSMOKAなど他のシステムと共用であ る. WEBサーバとデータベースサーバは同じ計算機 上で稼働しており,利用者による検索とデータ請求の 機能を提供する. FTPサーバは利用者が請求したデー タをダウンロードするために使用される. 本システム で扱うデータは主として,スタック済みデータ(FITS ファイル),メタデータ(データベース),データ請求 情報(テキストファイル),およびクイックルック画像 (PNGファイル)である.

本システムの各部の詳細は以下の各章で論ずる.

# 3 データの生成と転送

Tomo-eは84個のCMOS検出器を同時に駆動してそ れぞれ毎秒2コマで観測するため,膨大な観測データ が次々に生成される.観測データは即時に解析パイプ ライン[12]で一次処理が行われ,木曽観測所のデータ アーカイブシステムに保存されていく.これらの処理 は木曽観測所に設置された計算機群(データ取得用計 算機,一時保存用バッファ計算機,データ解析用計算 機)によって行われる.データ生成から保存までの流 れを図2に示す.

84個のCMOS検出器からのデータ取得と画像生成 (FITS形式の生データ)・保存を遅延なく実行するた め、データ取得用計算機と一時保存用バッファ計算機 はそれぞれ4台並列で処理を実施している.カメラか らは16ビット整数型で画像データを読みだしているが、 データ取得用計算機でFITSファイルを生成する際に バイアス信号除去を行ったのちに、画像データを32 ビット浮動小数点数型として記録している.生成した FITSファイルは4台の一時保存用バッファ計算機内の 日付ごとのディレクトリ下に分散して保存される.こ のFITSファイルは毎秒2コマで撮影した複数の静止画 から構成される3次元FITS(動画データ)ファイルで あり、Tomo-eではこの時点のFITSファイルを「生デー タ」と定義している. 生データのFITSファイルは1回の動画撮影(3大 サーベイプロジェクトの大部分では露出時間0.5秒× 12コマ,あるいは18コマ)で一つ生成される.ファイ ル名には下記の命名規則により一意に決まる文字列が 割り当てられる.

# TMQqyyyymmdd NNNNNNDD.fits

q:カメラモジュール番号 (1-4)



図2: Tomo-e Gozenのデータ生成から保存の流れ. 図中の 計算機システムは全て木曽観測所内に設置されているもの だが,そのうちデータアーカイブシステムは本館,それ以 外はシュミット望遠鏡ドーム建物内に設置されている. yyyymmdd: 日付(UTC) NNNNNNN: Exposure ID(通し番号) q+DD(+は文字列の連結):検出器番号(酒向ら[8]の 図13を参照<sup>2)</sup>)

ー時保存用バッファ計算機に保存した生データに は、データ解析用計算機でダーク補正、フラット補正、 位置較正等の一次処理が順次施される.さらに3次元 データのスタック処理を行ってデータサイズを約1/10 に削減した後、データアーカイブシステム内の日付ご とのディレクトリ下に保存される.一次処理、スタッ ク処理および保存処理も遅延なく行う必要があるため、 データ解析用計算機とデータアーカイブシステムはそ れぞれ3台並列の構成とした.

解析済みデータのファイル名には、下記のように先 頭にデータの種別を表すプレフィックスが付加され る.しかし、解析のやり直しなど一つの生データに対 して解析処理を複数回行う場合には、ファイル名の重 複が発生する.その場合にもデータを上書きすること がないよう、解析済みデータは解析パイプラインの実 行回数に対応した通し番号(リダクションID)の名前 のディレクトリ下に格納されている.本システムにス タック済みデータを転送しているデータアーカイブシ ステムのファイル・ディレクトリ構造を図3に示す.

#### aTMQqyyyymmddNNNNNNNDD.fits

a: データの種別を表すプレフィックス 例) d: ダーク較正用画像, f: フラット較正用画像, s: 一次処理済み2次元スタック画像, x: 一次処理なし2 次元スタック画像

データアーカイブシステムに保存された解析済み データのうち、観測日が2019-10-01から2021-05-14ま でのデータ(解析処理が複数回行われファイル名が 重複するものも含む)が木曽観測所から3多重で本シ ステムの転送用サーバへ転送された. 転送するデー タを8TB(転送用サーバに装着された単体ドライブ の容量;単体ドライブーつが1ファイルシステムとし てマウントされている) 以内に収まるように日付の ディレクトリ単位で分割して予めリスト化し、単体ド ライブ1本分ずつrsyncコマンドを用いて転送した.合 計269TBの転送に要した期間は2021年5月26日から9 月10日までの108日間であった(手動でコマンドの実 行を行っていたため、断続的に転送が途絶えた期間 があった).なお、木曽観測所からSMOKAへのデー タ転送試験は多重度を上げても400 Mbpsで頭打ちに なっており、木曽観測所から本郷へのデータ転送では 3,760 Mbps 出ていることから木曽観測所側ではないど こかにボトルネックがあるものと思われる. このボト ルネックがどこにあるのかは未解明である.

転送されたプレフィックスがsであるスタック済み データは29,163,330フレームであり,この中で、3大 サーベイプロジェクトのいずれかの観測のデータで あり、かつ画像サイズがNAXIS1 = 2000, NAXIS2 = 1128であるものは27,918,959フレームである.そのう ち、2019-10-01から2020-03-31までの間のフレーム数 は9,153,769であるが、1章で最後で述べた例外(詳細 は9.2節で論ずる;計13フレーム)を除いた9,153,756

2) 全てのカメラモジュールについての図はTomo-e GozenのWEBページ (https://tomoe.mtk.ioa.s.u-tokyo.ac.jp/images/tomoe\_distribusion\_of\_sensors.png) に記載されている.



図3:データアーカイブシステムのファイル・ディレクトリ構造.木曽観測所に設置されている Tomo-e Gozen データアーカ イブシステムは3台並列の構成となっており,観測日毎のディレクトリが3台に分割して置かれている.ここから SMOKA/ Tomo-e システムにスタック済みデータが送られている.

フレームが第1段階公開データである.なお,第1段 階公開データは3大サーベイプロジェクトのうちの 「Supernova Survey」のみであった.

# 4 データの配置

#### 4.1 保存領域へのデータ配置

本システムの転送用サーバには活性交換可能な合計 23本の単体ドライブを各々別のファイルシステムとし てマウントすることが可能である. 個々の単体ドラ イブを着脱することによって磁気ディスク領域の不 足を補いつつ、サーバ外で保管するなど高い自由度 の下で用いることができる. しかしながら, RAIDを 構成していないためにデータの安全度は不十分であ り, また, 本システムのWEBサーバやFTPサーバか ら通常はアクセスすることができない設定になってい る. そこで転送されたスタック済みデータを順次本シ ステムのファイルサーバの保存領域にコピーした. そ れぞれのスタック済みデータにとっては本システム 内でのコピーはこれが最初で最後となる.3章の最後 で述べた転送済の2019-10-01から2021-05-14までのス タック済みデータ(8TBの単体ドライブ39本を費やし た) は八つのファイルシステム (図11の/da0[1-8]y/ TomoeNP0[1-8]/. 今回の転送分ではそれぞれ5-42TB 使用) に分割して配置された. 八つものファイルシ ステムを用いたのは、ファイルサーバはSMOKA など 他のシステムと共用であり、今回の転送分だけでなく 次回の転送分をも含めた空き容量を確保するためであ る. 木曽観測所のデータアーカイブサーバ上のデータ 構造に合わせて3分割された同じ観測日のスタック済 みデータは3多重で単体ドライブ3本に分かれてネット ワーク転送されているため、保存領域へのコピー作業 の際には同じ観測日のデータをまとめる手順を要した. 第1段階での公開対象となったスタック済みデータは 八つのファイルシステムのうちの二つ (/da0[1-2]y/ TomoeNP0[1-2]/) に置かれている.

#### 4.2 公開領域へのデータ配置

公開日に達したスタック済みデータは、八つのファ イルシステムそれぞれの中の公開領域に自動的に配置 される.この配置はファイル単位でのハードリンク (図11参照)によって実現され、毎朝9時(UT0時)す ぎにcronによって実行される.この処理は公開作業 の自動化の要の一つである.ハードリンクによる配置 であるので、スタック済みデータはファイルサーバの 保存領域から移動することはない. この過程は利用者 による検索やデータ請求などにとっては必須なもので はないが、プログラムの誤動作や作業中の誤操作など によって未公開のデータが利用者に見えてしまうなど を防ぐべく、スタック済みデータの公開・未公開の区 別を明確にするために行っている. この過程を実現す るハードリンク以外の手段として、シンボリックリン ク,移動 (mv),コピー (cp) が考えられる.しかし 1章で技術的目標として述べた通り、作業時間と記憶 装置容量を多く費やすコピー (cp) は行わないことと し、8章で論ずるようにFTP領域による制限によって シンボリックリンクも除かれる. また, ハードリンク の場合と同じ同一ファイルシステム内という条件の下 では移動(mv)は高速であり、かつ、新たな記憶装置 容量を要しないが、元ファイルの消去を伴うことによ る危険性を考えて避けることとした. ハードリンクに よるデータ配置の処理は高速であり、公開を開始する スタック済みデータは1日あたり平均で約5万フレーム であったがその配置作業の所要時間は2ないし3分間で ある.

# 5 データベース

SMOKAには多くのデータベーステーブルが用意さ れており、それらを駆使することによってデータの管 理や検索などの機能が実現されている(論文2,論文7). 本システムでも SMOKAと基本的に同様のデータベー ス構造を採用し、ファイル管理、ヘッダキーワード登 録、データ検索の各機能を支援するテーブル(実テー ブルと仮想テーブル (ビュー)) を用意した. SMOKA ではこれら機能に加えてデータ修正を支援するテーブ ルも存在するが、Tomo-eではデータ修正の必要がない ため、そのようなテーブルを用意しなかった. さらに SMOKAとは異なる本システムの特色として、実テー ブルに公開予定のスタック済みデータの情報を(公開 予定のFITSファイルが保存領域に配置され次第随時) その公開に先立って入力し、ビューで公開対象かつ公 開日を過ぎたスタック済みデータの情報のみを実テー ブルから参照するようにビューの定義を設定すること で、自動的にスタック済みデータをその公開日に検索 および請求可能な状態にする.これも,公開作業の自 動化の要の処理の一つである. ビューと実テーブルの 役割と関係の概要を図4に示した. また、SMOKAで は利用していなかったデータ検索用のSQLユーザ定 義関数を開発し利用している.以下に、各テーブルと データ検索用関数について述べる.



図4:データベーステーブルの相互関係.スタック済み データの管理情報およびFITS ヘッダから抽出したメタ データはそれぞれ実テーブルである、ファイル管理テーブ ル filemng\_tmqおよびヘッダキーワードテーブル tmq\_kiso に登録される.公開対象でかつ公開日に達したエントリは 公開データ用ビューである tmq\_smoka および filemng\_open として参照可能になり、データ検索とデータ請求に利用さ れる.

#### 5.1 ファイル管理テーブル

ファイル管理テーブルfilemng\_tmqは公開済みお よび公開予定分も含むスタック済みデータの保存場所 や公開ステータスを管理する実テーブルである.本 テーブルは、フレームID,観測日、データの保存場所、 データのMD5、データ公開日、データ公開フラグ(公 開対象かの判定)の6個の列から構成されている.フ レームIDを主キーとしてデータの重複を防ぎ、6個の 列のうちデータ公開日とデータ公開フラグにインデッ クスを定義することで高速化をはかっている.

後述するデータ請求時に使用されるビューfilemng\_ openは本テーブルfilemng\_tmqを参照している.

#### 5.2 ヘッダキーワードテーブル

ヘッダキーワードテーブルtmq\_kisoは公開済み および公開予定分も含むスタック済みデータのFITS ファイルのヘッダのキーワードの値を記録する実テー ブルである.ヘッダのキーワードはしだいに増えてお り,現在本テーブルに含まれるキーワードは第1段階 ルールでの公開データに含まれる145のキーワードで ある.本テーブルは154の列から構成されており,内 145列がFITSキーワードの値を保管するためのもの. 9列がHISTORY行を保管するためのものである.フレームIDを主キーとしてデータの重複を防ぎ,データ検索で使用される列にインデックスを定義することで高速化をはかっている.

後述するデータ検索時に使用されるビューtmq\_smoka は本テーブルtmq\_kisoを参照している.

#### 5.3 公開データ用ビュー

公開データ用ビューとしてデータ検索に用いられる tmq\_smokaとデータ請求に用いられるfilemng\_open を作成した.両ビューとも公開対象かつ公開期日に達 したデータのみを実テーブルから参照することでデー タ公開の自動化を実現した.

tmq\_smokaは本システムのウェブページ上でデータ 検索が要求された際に問い合わせを受けるビューであ る.本ビューは検索テーブルtmq\_kisoからデータ検 索に必要な列を参照する.また本ビューでは、6章で 述べるデータ検索のため、赤道座標系の値を持つra列 とdec列から直交座標系の値を持つx列、y列、z列が ユーザ定義関数feq2x,feq2y,feq2zそれぞれを使っ て定義されている.

通常列に対する演算の結果はインデックスを持たないことからx列, y列, z列に対する問い合わせは順次スキャンとなり検索性能が大きく低下してしまう.本ビューでは参照元テーブルtmq\_kisoに複合・式インデックス[13]としてbtree (feq2x,feq2y,feq2z)を定義したことでビューのx列, y列, z列へのインデックス検索を有効にし高速な検索を実現した.

式インデックス以外の高速化方法としてマテリア ライズドビュー[13]並びに式結果の実テーブル化が 考えられるが,前者は第1段階で公開予定の9,153,756 フレームの情報を入力した時点でマテリアライズド ビューの更新に10分以上時間がかかり今後データが増 加していった際に運用に支障をきたすため採用しな かった.後者はra列とdec列に更新の必要が生じた際 に実テーブルのx列,y列,z列を同時に更新するトリ ガ関数を作成すれば検索元テーブルと実テーブルとの 間の齟齬を防ぎつつ運用することが可能であるが,一 方のテーブルがもう一方のテーブルに影響を与えると いうトリガ関数の仕組みから多用すると保守性の低下 を招き,また自動的に動作し運用者からその存在を忘 れられがちであるため採用しなかった.

filemng\_open は本システムのウェブページ上で データ請求が行われた際に問い合わせを受けるビュー である.ファイル管理テーブルfilemng\_tmqのフ レームID, 観測日, データの保存場所, データのMD5 の4個の列を参照している.

#### 5.4 データ検索用関数

本システムのデータベースでは、データ検索の際 に必要な演算の一部をPostgreSQLのユーザー定義関 数を使って実現した.ユーザー定義関数を処理内容 ごとに定義しておくことでデータ検索時に生成する SELECT文が平易となり、検索プログラムの保守性を 高めることができる.

作成した関数は5.3で述べた関数feq2x, feq2y, feq2zのほか,検索中心から任意角度内にある天体を 検索する関数fconesearchtmqsmokaCRVALと,同関数 の実装に必要な各種関数群である.使用した言語は SQLとPL/pgSQLである.

関数fconesearchtmqsmokaCRVALはいわゆるCone Search [14]を実装するものであり、検索中心と天体の 位置ベクトルのなす角を求め、求めた角度が任意角度 以下となる天体を抽出する.その処理内容をシンプル なSQL文で書くと以下である.

## SELECT \* FROM tmq\_smoka WHERE

ACOS(x1x2+y1y2+z1z2) < R; (Rは任意角度) しかし上記SQL文ではWHERE句内でx列, y列, z列 に対する演算を行っているため, 実テーブルtmq\_kiso に設定したインデックスbtree (feq2x,feq2y,feq2z) が使用できず, 順次スキャンとなり検索性能が悪い. 検索性能の向上のため,本関数では検索中心を中心と したxyz空間での1辺2Rの立方体の内側の天体をイン デックス検索で絞り込んだのち,上記SQL文を実行し ている.本関数を実装したことで6章で述べるラフ検 索で生成するSQL文を平易にできた.

# 6 データ検索

本システムのデータ検索機能としてピンポイント検索, ラフ(Rough)検索,カレンダー検索の3種類を用意した.なお,本システムでは1回の動画撮影から作成されるスタック済みデータの全フレームの集合をショットとよぶ.ピンポイント検索ではフレームが,ラフ検索とカレンダー検索ではショットが検索の直接の対象である.

SMOKAで使用しているプログラム (jsp および java)を再利用することで短い期間で開発を行うこと ができた.ただし,SMOKAのSQL文作成プログラム については採用しなかった.本システムは単一の観測 装置の観測データを扱うため、複数の観測装置の観測 データを対象としたSMOKAのSQL文作成プログラ ムは本システムでは不要に複雑であり、そのプログラ ムを改造するよりも各検索機能ごとにSQL文を含む プログラムを新規作成する方が負担が軽いと判断した. 新規作成したプログラムにおいては可読性も高まった と思われる.新規観測装置の観測データを取り込む際 にSQL文作成プログラムの改造の手間がかかること がSMOKAの欠点の一つであった(論文7)が,本シス テムではその問題を回避することができた.

以下,各検索機能の詳細を述べる.

#### 6.1 ピンポイント検索

図5はピンポイント検索の入力画面である.入力必 須項目は赤経・赤緯の天球座標であり,ピンポイン ト検索ではその座標点(以降,検索座標点)を視野 に含むフレームを検索する.Tomo-eのスタック済み データは位置較正済みであり,各フレームのヘッダの WCS情報を利用して1秒角程度の精度での検索ができ るようにした.オプションとして,観測日(特定の日 または期間)で検索結果を絞り込めるようにもした.

SMOKAのピンポイント検索(論文5)ではHEALPix を利用しているが、本システムの場合フレーム数が膨 大であるためHEALPixの計算に膨大な時間がかかる と見積もり、HEALPixを利用しない手法を開発した. なお、HEALPixの計算時間はHEALPixの階層と視野 の大きさの兼ね合いや演算の多重度などに左右される が、MITSuME [15](視野一辺30 arcmin弱)の5万フレー ム(おおむね1か月あたりのフレーム数)の場合、14階 層で視野のHEALPixインデックス化(10多重)におよ そ3時間、そのデータベース化(2多重)におよそ2時間 を要している.

Tomo-eのピンポイント検索は2段階の工程からなる. 最初の工程では, FITS ヘッダのCRVAL1, CRVAL2の

# Tomo-e Gozen Pinpoint Search

Click here to know how to search.

Search Conditions						
Object Name (for name resolve) Resolver						
Object Name		SIMBAD	ONED Resolve			
Coordinates	<b>RA</b> RA(J2000) <b>DEC</b> DEC(J2000)					
Observation Date	YYYY-MM	1-DD				
Output Format  OASCII (max 20,000 rows, text)						
Search	Reset to defaults	Help.				
Output Options						
EXP_ID	der by:		Rows per page:			

図5:ピンポイント検索の入力画面. Search Conditionsの枠 内で検索条件を指定し, Output Optionsの枠内で検索結果 の表示方法を選択する.

値が検索座標から半径46分角の範囲にあるフレーム に絞り込む検索を行う. CRVAL1, CRVAL2の参照点 CRPIX1, CRPIX2の値はフレームによって異なるうえ, その分布は広範囲にわたり(DATE OBSの範囲2019-10-01..2020-02-09のデータにおいて、38 < CRPIX1 < 1979, 17 < CRPIX2 < 1117である), Tomo-eの各フレー ムの視野対角は45.6分角であるため絞り込み検索の半 径を46分角とした.ただし、ここで円領域での検索 を行うと三角関数の多用により検索速度が落ちるため, 検索座標±半径の矩形領域での絞り込み検索を行った. 同様の絞り込み手法をSMOKAで採用しており(論文 5), そのプログラムをSMOKAから再利用すること で開発期間を短縮することができた.2番目の工程で は、上記で絞り込んだフレームそれぞれについて、検 索座標に対応するピクセル座標を WCS 情報 (SIP 項を 含む)に基づいて計算し、その値が0≤X≤2001かつ 0≤Y≤1129の場合には、検索座標がそのフレームに 含まれると判断する. この手法は検索座標点を含むフ レームの検索(点の検索)をする場合に有効であるが、 HEALPixとは違い、広がりをもつ領域を含むフレーム の検索(面の検索)はできない<sup>3)</sup>

図6はピンポイント検索の結果画面の例である.検 索条件に適合した各フレームのフレームID, 主要 なメタデータ,サムネイル画像へのリンク,さらに 当該フレームでの検索座標に対応するピクセル座標 (Pinpoint\_X, Pinpoint\_Y)を表示するようにした.メタ

3) SMOKAでは当初,重複領域検索(論文5)のために面の検索が可 能なHEALPixの利用を開始し,HEALPixは単純なアルゴリズムによ り点の検索も可能であることがわかったため,HEALPixをピンポイン ト検索に適用した.SMOKAは現在では重複領域検索のサービスを停 止しているため,HEALPixはピンポイント検索のみに利用されている. データのうち, RA\_FRAME および DEC\_FRAME は当 該フレームの中央のピクセル (X, Y = 1000, 564) での 赤経・赤緯である (WCS 情報から算出した). EXP\_ID (ショットのID) にはその EXP\_ID 全てのフレームの一 覧をリンクし, 当該フレームを含むショットの (全て または一部の) ファイル請求およびサムネイルの閲覧 ができるようにした.

#### 6.2 ラフ検索

ピンポイント検索と同様の入力画面とした. ラフ検 索では検索座標をその視野範囲(フレーム間の隙間も 含む)に「おおむね」含むショットを検索する. ピン ポイント検索と同様に, 観測日(特定の日または期間) で検索結果を絞り込めるようにした.

Tomo-eのFITS ヘッダの (RA, DEC) は望遠鏡のポ インティングに基づく各ショットの視野範囲中心の概 算値である.ポインティング精度は観測時期によって 異なり、2021年9月21日以前はrmsで10分角程度であり、 それ以降は20秒角程度である[16]. ラフ検索では検索 座標から半径6度以内に(RA. DEC)をもつショットを 検索する.ショットの全フレームを内包する最小半径 は約4.5度であるが、上記ポインティングの精度を考慮 に入れるとともに、Tomo-eのサーベイ観測ではフレー ム間の隙間を埋めるディザリングを行っており、同じ ディザリングのセットのショットも検索結果に含める ために検索の半径を6度とした.ここでも、検索座標 +/-半径の矩形領域での検索を行って絞り込みを行っ たのちに、円領域での検索を行った.円領域の検索に は5章に記載のfconesearchtmqsmokaCRVAL 関数を利 用した.

#### Results 00h42m44.32, 41d16m07.5, J2000

Click here to know how to look search results.

279 frames were found.

To retrieve data, mark checkboxes at the "Check to Retrieve" column of rows which correspond to the frames which you'd like to retrieve. Then push "Datarequest" button located above/below the table.

To view other page of the query results, select the range of numbers from the list box located at the bottom of the table, then push "Go" button next to it.

Mark all Unmark all Datarequest (for 1-100)

No.	Thumbnail	Check	FRAME ID	<b>RA FRAME</b>	DEC FRAME	DATE OBS	TIME OBS	<b>Pinpoint X</b>	<b>Pinpoint</b> Y	EXPTIME	EXP ID
		to Retrieve									
1	<u>thumbnail</u>		TMQ1201910020015182815	0:41:37.6	+41:09:07.3	2019-10-02 weather	18:29:28.5	1634.3	213.8	6.00	<u>151828</u>
2	<u>thumbnail</u>		TMQ4201910020015183315	0:43:33.5	+41:12:32.0	2019-10-02 weather	18:31:10.5	535.5	377.1	6.00	<u>151833</u>
3	<u>thumbnail</u>		TMQ1201910020015188415	0:41:40.5	+41:08:58.5	2019-10-02 weather	19:39:51.2	1606.9	206.3	6.00	<u>151884</u>
4	<u>thumbnail</u>		TMQ4201910020015188915	0:43:36.1	+41:12:20.6	2019-10-02 weather	19:41:33.6	511.6	367.2	6.00	<u>151889</u>
5	<u>thumbnail</u>		TMQ4201910040015233915	0:43:31.3	+41:13:19.7	2019-10-04 weather	15:28:43.9	556.2	417.5	6.00	<u>152339</u>

図6:ピンポイント検索の結果画面例.

# Results 00h42m44.32, 41d16m07.5, J2000

<u>Click here</u> to know how to look search results.

#### 897 exposures were found.

To retrieve data, mark checkboxes at the "Check to Retrieve" column of rows which correspond to the shots which you'd like to retrieve. Then push "Datarequest" button located above/below the table.

To view other page of the query results, select the range of numbers from the list box located at the bottom of the table, then push "Go" button next to it.

Mark all Unmark all Datarequest (for 1-100)

No.	Thumbnail	Check to Retrieve	<u>SHOT</u>	RA_SHOT	DEC_SHOT	DATE_OBS	TIME_OBS	EXPTIME
1	<u>thumbnail</u>		TMQX2019100200151827XX	00:43:01.2	+37:08:17	2019-10-02 weather	18:29:09.7	6.00
2	<u>thumbnail</u>		TMQX2019100200151828XX	00:43:01.1	+37:32:16	2019-10-02 weather	18:29:28.5	6.00
3	<u>thumbnail</u>		TMQX2019100200151829XX	00:40:14.6	+37:32:15	2019-10-02 weather	18:29:47.6	6.00
4	<u>thumbnail</u>		TMQX2019100200151830XX	00:40:14.5	+37:08:14	2019-10-02 weather	18:30:06.5	6.00

図7:ラフ検索の結果画面例.

図7はラフ検索の結果画面の例である.ピンポイント 検索の結果画面とは違い,検索条件に適合したショッ トを各行に表示し,赤経・赤緯は各ショットのFITS ヘッダの(RA, DEC)を表示し,ショット名にはその ショットに含まれるフレーム一覧をリンクした.検索 結果一覧からショット単位でデータを選んでデータ請 求ができるほか,ショット名からのリンク先でフレー ム単位でのデータ請求もできるようにした.

#### 6.3 カレンダー検索

SMOKAのカレンダー検索機能(論文4)と同様の機能をTomo-e用に構築した.図8はカレンダー検索の検索画面である.検索結果は,選択したDATE-OBSに観測されたショットについてラフ検索と同じ項目を表示するようにした.

2020	v 1 v	More ExpID	Information Date:2020-1-5 :219336-221447 (2091 exposures)
Table	(HTML)		
Ascii	(Text)		
2020/1	Stacked Data	Weather and Skymonitor	
1	<u>0</u>	<u>LINK</u>	
2	<u>0</u>	<u>LINK</u>	
3	<u>0</u>	<u>LINK</u>	
4	<u>0</u>	<u>LINK</u>	
5	<u>0</u>	<u>LINK</u>	
6	<u>0</u>	<u>LINK</u>	
7	-	<u>LINK</u>	
8	<u>0</u>	LINK	
9	Ò	LINK	
10	<u>0</u>	<u>LINK</u>	
11	<u>0</u>	<u>LINK</u>	

図8:カレンダー検索の検索画面.

# 7 クイックルック

利用者がスタック済みデータの画像の概要を迅速に 把握できるようにするため、本システムではクイック ルックとしてThumbnail画像とSHOT画像を用意して いる.前者はフレーム単位で縮小した画像であり、後 者は1ショット分のThumbnail画像をまとめて表示し た画像である.Thumbnail画像とSHOT画像はすべて のスタック済みデータについて公開日を待たずに作 成してWEBサーバの非公開の保存領域に置いておき、 公開日に達したものについてWEBサーバの公開領域 (検索結果画面からリンク可能な領域)へ自動的にシ ンボリックリンクを貼ることで配置している.この配 置を行う処理がcronにより毎朝9時過ぎに実行される.

#### (1) Thumbnail 画像

スタック済みデータ(画素数2000×1128)の各フレームを0.06250倍(1/16倍)に縮小してThumbnail 画像 (画素数124×70)を作成した. Thumbnail 画像の例 を図9に示す. DS9<sup>4)</sup> および XPA<sup>5)</sup>の機能を利用し て表示階調の自動調整やファイルの自動作成を行い, Thumbnail 画像のPNGファイルを作成している. この 方法では10万フレームあたり約5時間の作成時間がか かる.

#### (2) SHOT 画像

1ショット分(84枚. ただし欠損がある場合あり) のThumbnail 画像(作成済みであることが前提)を

<sup>4)</sup> https://sites.google.com/cfa.harvard.edu/saoimageds9

<sup>5)</sup> http://ds9.si.edu/doc/ref/xpa.html



図9: Thumbnail 画像の例. M31が視野に含まれている.



図10:SHOT 画像の例. 図9の Thumbnail を含む SHOT である.

Tomo-eのカメラのチップ配列を模した配置に並べて SHOT 画像を作成した.SHOT 画像の例を図10に示 す.SHOT 画像は画素数800×900のPNG ファイルで あり、その作成にはアプリケーションソフトウエアの ImageMagick<sup>6)</sup>を用いた.WEB上の表示サイズの都合 から Thumbnail 画像を1/2に縮小して用いている.作成 所要時間は、2000ショット(約16万フレームに相当) あたり約200分である.

# 8 データ請求処理

データ請求処理とは、利用者が求めるフレームに対応したFITSファイルを利用者がアクセスできる場所(本システムではFTPサーバ)に配置する処理のことである.

本システムではSMOKAと同様にデータ請求には利 用者アカウントの取得を必須としている. SMOKAと 共通の利用者アカウントとし,利用者登録もSMOKA の利用者登録機能を用いることとした.

データ請求処理の概要は以下の通りである.利用者 によってWEB画面を用いてデータ請求が行われると, WEBサーバによりデータ請求内容(利用者アカウント, 請求日時,請求フレーム)を記録したテキストファイ ル(データ請求ファイル)が所定のディレクトリ下に 作成され,それが引き金となって処理が始まり,請求 されたFITSファイルがFTPサーバ上に配置され,そ れが利用者にメールで通知されて処理が終了する.こ のデータ請求処理もSMOKAの機能を踏襲したもので あり,SMOKAで用いられているプログラムを再利用 することで開発を省力化することができた.ただし, 本システムではデータのコピーを行わない,という方 針の下,データの配置の方法がSMOKAのそれとは若 干異なっている.なお,本論文執筆時においてはデー

<sup>6)</sup> https://imagemagick.org/

タ請求は一つずつ処理する(同時処理数=1)としている.また、1人の利用者の連続したデータ請求が他者のデータ請求を長時間阻害せぬように処理の優先度制御を行っている.

データ請求処理でのデータ配置の手法を制限する最 大の要因はFTPサーバである. FTPサーバではセキュ リティ保持のため、利用者は求めるデータが置かれた 領域(FTPサーバのFTP領域、とよぶ)以外を見るこ とができないように設定される. そのためFTPサーバ で利用者に提供すべきデータを配置するには、一般的 なLinuxの機能の範囲では(1) FTP領域下にコピーす る、(2) FTP領域下にディレクトリごとNFSマウント する,のいずれかしかない.そして「データのコピー を行わない」という方針の下では(2)を採ることにな る. なお, SMOKAはHeader Replacer (HR) 処理のた めにコピーを行うわけであるが、その負荷を軽減分散 させるために結局は手法(2)を採っている(論文7). データのコピーは時間や計算機資源を大きく費やす処 理であり、SMOKAの開発はその軽減の努力とともに あると言えよう. さらに、NFSマウントの際には元領 域でのシンボリックリンクをそのままでは伴うことが できないことも制約条件となる、そのため、ファイル サーバ上でのFTP領域(NFSの元領域)に対しても4 章で述べた保存領域/公開領域と同様にハードリンク を行うこととした. このようにして本システム全体の データ配置は図11で示した構成となった.

SMOKAではデータ請求処理の中枢を担当するプロ グラム群をDCM (Data Copy Manager, 論文1, 論文7) と呼んでいるが,本システムでは,コピー処理をハー ドリンクやシンボリックリンク等の処理に置き換えた ため,DLM (Data Link Manager)と呼んでいる.その 処理の流れを図12に示した.WEBサーバとFTPサー バの両方にわたった多くのディレクトリを用いてデー タ請求処理を多段階化したのは,SMOKAのDCMと 同様(あるいはそれ以上)に処理の途上でトラブルが 生じた場合の原因の切り分けを容易にするとともに, 途中の段階からのトラブルの再現と処理の再開を容易 にするためである.

# 9 本システム開発の総括と今後の課題

#### 9.1 開発の総括

本開発では、Tomo-eのスタック済みデータの第1段 階の公開システムを構築して公開した.公開期日に達 したデータを順次公開していく、という基本機能は実 現でき、大きな問題もなく運用が継続できている.ま た、1章で述べた技術的目標:(A)データの公開作業の 自動化、(B)公開作業とデータ請求対応作業でデータ のコピーを行わない、(C) HEALPix を介さないピンポ イント検索、(D) できるだけ単純なシステム構造も実



図11:スタック済みデータの配置.スタック済みデータはファイルサーバの八つのファイルシステム(/da0[1-8]y)上の保存領域のどれかに置かれ、公開日に達するとそれぞれのファイルシステムの公開領域にハードリンクされる.さらに、データ請求を受けたスタック済みデータはそれぞれのファイルシステムのFTP領域にハードリンクされる.FTPサーバにはファイルサーバのFTP領域がNFSマウントされる(~tomoeFTP/FTP/G0[1-8]).利用者がアクセスしやすいよう、それらの下のファイルは請求毎に作成されるディレクトリ(~tomoeFTP/FTP/Download/{reqID})下にまとめてシンボリックリンクされる.また、WEBサーバ上のデータ請求関連ファイル(データ請求ファイル、ログ、MD5値一覧)もNFSマウントを経てFTPサーバ上で利用者がアクセスできるようにしている.



図12: Data Link Manager (DLM) によるデータ請求処理の 流れ.利用者がデータ請求を行うとデータ請求ファイルが 作成されて「Request」ディレクトリに置かれる(処理1). その後数字の順に処理が行われ、データ請求ファイルや処 理の途上で作成された関連ファイルが次々と移動される. 最後(処理8)に処理の結果が利用者や管理者に電子メー ルで通知されてデータ請求処理は終了する. なお、WEB サーバ上の「Processing」ディレクトリと「Processed」ディ レクトリはそれぞれFTPサーバの「Waiting」ディレクト リと「Ended」ディレクトリにNFSマウントされており実 質上同一のものである.

現することができた. さらに. 木曽観測所からのデー タ転送終了(2021年9月10日)後, MD5値の照合など の確認を経てデータの保存領域への配置を2021年9月 24日に終え、その間にデータ公開の方針(第1段階ルー ル)が定められて本システムの開発に着手し、動作確 認作業を経たうえで2021年11月1日に公開を開始する ことができた。1か月間ほどの短期間開発が実現でき た要因は、1章で述べたTomo-eのスタック済みデータ の特徴(特に(4)と(5)),単一の観測装置の観測デー タを扱っていること、および技術的目標(D)の「でき るだけ単純なシステム構造」の実現にあると考えてい る、このことは、今後の観測データアーカイブ開発・ 構築の際には大いに参考となるであろう、この経験を 生かすことによって、例えば、SMOKAの機能の一部 を省いたデータアーカイブシステムを他所で容易に構 築する、などが期待できるだろう.

本システムで選択した上記技術目標にはメリットだ

けでなくデメリットも存在する. それぞれのデメリットを以下に挙げる. これらも今後の観測データアーカ イブ開発・構築の際に参考となるであろう.

- (A) データの公開作業の自動化
  - ・FITS ヘッダに欠損などがないことが自動化の前提 であるが、その前提が成り立たないフレームが存 在すると自動処理の途上でエラーが発生して手戻 りを余儀なくされ、復旧作業に膨大な手間と時間 がかかる場合が生ずる.
- (B) データのコピーを行わない
  - ・SMOKAでは利用者が観測データをダウンロードする 際に、ファイルをtarで固めたり、圧縮したりする選択 が行えるが、コピーを行わない、という方針を堅守す ると、実質的なコピーであるtarはできず、複数のハー ドリンクがあるファイルの圧縮に必要なコピーも行えず、 これらの選択ができない.
  - ・保存領域から公開領域にコピーが行われる場合と 比べて、データの冗長性が低くなり万が一のハー ドウエア不具合に対して脆弱となる.
- (C) HEALPix を利用しない
  - ・SMOKAでかつて行われていた重複領域検索(論 文5および6)を導入することができない.
  - NASA が配布している HEALPix パッケージ<sup>7)</sup>には、多くの便利なプログラムが含まれている。例えば、任意の円形領域に含まれる HEALPix インデックスを検索するプログラムが含まれている(論文5). HEALPix インデックスを算出したうえでHEALPix パッケージを利用する、ということを行わなければ様々な機能拡張の際に自力での開発が必要となってくる。
- (D) できるだけ単純なシステム(Header Replacer を使用しない、単一観測装置を対象とする)
  - ・ヘッダ修正が必要となった場合に迅速には対応できない、あるいは、Header Replacer 以外の手段を別途開発する必要が生じる(9.2の(4)および論文7).
  - ・SMOKAで扱っている多くの観測装置の観測デー タと連携しての検索が容易ではない.

#### 9.2 今後の課題

今後の課題としてまず生データの公開が挙げられ

<sup>7)</sup> https://healpix.sourceforge.io/documentation.php

る.1章で述べたように、Tomo-eの生データはこれま で取得5日後に捨てられてきたが、2022年3月観測分か ら、毎晩取得した観測生データの一部(darkとflatと最 初の100ショット、おおむね1.8TB)を保存することに なった.これは、データ処理/解析ソフトウエアの再 現性などの確認を可能とすることにより成果の信頼性 を確保するために行うものであるが、天文学研究への 直接利用も期待できるだろう.

生データは, Tomo-eの一時保存用バッファ計算機か ら木曽観測所内に設置された SMOKA のファイルサー バに接続された磁気ディスクユニット(=単体ドライ ブ計22本)へ転送される.シュミット望遠鏡ドーム建 物内にある一時保存用バッファ計算機と、300m離れ た本館にある SMOKA ファイルサーバの間は専用の光 ファイバで接続され、10 Gbpsの速度で通信できる設 計となっている.データ転送は観測に影響しないよう, 観測終了後の時間に実施される。生データは一時保存 用バッファ計算機4台から4本の単体ドライブそれぞれ に並列で送られる.ファイルサーバは、どの単体ドラ イブに書き込みを行うかの選択や、満杯に近づいた単 体ドライブの切り替えを自動的に行うように管理を 行っており、生データは日々自動的に磁気ディスクユ ニットへ転送されている.満杯になった単体ドライブ はある程度まとまったところで三鷹に郵送などを利用 して送られている.

生データは成果の検証の手段を提供するという面で ぜひとも公開すべきであるが、本論文で述べてきたス タック済みデータとどのように結び付けて公開してい くのか今後の課題である.

またさらに、以下の課題が挙げられる.

(1) 例外データの組み入れ

1章の最後で述べたように本論文のシステムでは 例外の若干フレームを対象から除いている.例外 としたのは、画素数が2000×1128でないもの(部 分読み出し)336フレーム、位置較正がうまくいか なかった12フレーム、ファイル名重複1組2フレー ムのうち日付の古い方の1フレームの合計349フ レームである.これらを本システムにどう組み入 れるかは今後の課題である.

(2) 公開データが増えていくことによって懸念される問題

第2段階のデータ公開が控えており,今後扱う データの数も量も増えていく.その時に,検索速 度は実用の範囲にとどめられるのか,など安定運 用に対する懸念がある.動作を見守りつつ必要が 生ずれば対策を検討することになろう.

(3)次期計算機システムへの移行 今回用いた計算機システムは2023年2月末で契約 が終了する予定である.次期システムでは磁気 ディスク容量を補うべく、磁気テープライブラリ の積極的利用を行う予定であり、その活用を含め た次期システムへの発展的移行と開発を行わなけ ればならないだろう.

(4) FITS ヘッダの修正

今後FITS ヘッダの修正の必要が生じた場合に どうするかという課題がある.本システムに SMOKA 同様のHeader Replacer (論文7)を組み 入れるという選択肢があるが,コピーを行わない とする方針からは外れるほか,SMOKAのHeader Replacer には利点のみならず課題も存在するため, 慎重に検討する必要がある.別の選択肢として, 論文7で述べたFITS ヘッダ修正内容をファイルと して付け,利用者がそれぞれの目的に基づいて判 断し,修正を行わない,あるいは,必要に応じて 利用者が手元で修正プログラム (本システム上で プログラムの例を提示する)を走らせる,という 案を考えている.

(5)検索速度のさらなる改善をはかる 小澤ほか[17]ではデータアーカイブに用いられる データベースの検索を高速化するための方策と実 験結果が論じられている.その結果を採り入れ, 上記(2)の課題の克服も踏まえて検索のさらなる 高速化をはかっていきたい.

# 10 まとめ

木曽観測所のTomo-e Gozenのスタック済みデータ の公開システムを開発し、公開に供することができた. 本システムの開発にあたっては、Tomo-e Gozenのス タック済みデータに備わった好条件の下で、SMOKA が包含するいくつかの技術的課題の解決への挑戦を行 い、今後の観測データアーカイブシステムへの指針を 得ることができた.

前章で述べたようになおもいくつかの課題が残され ており、本システムの真価を評価できるのはその対 処の後になるであろうが、SMOKAを初めとする観測 データアーカイブの発展のための一段階として本論文 が役立てば幸いである。

## 謝辞:

本システムの立ち上げに際しては東京大学天文学教育 研究センターの酒向重行氏,大澤亮氏にご尽力いただ いた.両氏には本論文の執筆にあたっても多大なる助 言をいただいたことを感謝する.また,木曽観測所, および,Tomo-e Gozenチームの方々には本システムの 開発と運用にあたってひとかたならぬご協力とご支援 をいただいており,ここで厚く感謝する.匿名の査読 者には有益なご指摘ご意見をいただいた.ここで厚く 感謝する.本システムはその開発と運用の大部分を国 立天文台天文データセンターの経費によって賄われて いる.

## 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,6,23-26 (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, 奥村真一郎, 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] 山田善彦,小澤友彦,西澤 淳,古荘玲子,西村高 徳,榎 基宏,吉野 彰,古澤順子,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5,国立天文台報,12,53-78 (2009).
- [6] 野田祥代,古荘玲子,古澤順子,山田善彦,山内千里,小澤友彦,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6,国立天文 台報,14,35-61 (2012).

- [7] 中島康,樋口あや,格和純,小野里宏樹,野田 祥代,古澤順子,本間英智,高田唯史,市川伸 一:光学赤外線観測データアーカイブシステム
   SMOKA: 20年間の開発と運用,そして将来,国 立天文台報,22,1-44 (2022).
- [8] Sako, S., et al.: The Tomo-e Gozen wide field CMOS camera for the Kiso Schmidt telescope, *Proc. SPIE*, 10702, 107020J (2018).
- [9] Shupe, D. L., et al.: The SIP Convention for Representing Distortion in FITS Image Headers, Astronomical Data Analysis Software and Systems XIV, ASP Conference Series, 347, 491–495 (2005).
- [10] 小澤友彦,野田祥代,古澤順子,吉田鉄生,樋口 祐一,市川伸一: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 (2015).
- [11] 青木勉,征矢野隆夫,中嶋浩一,宮内良子,森由 貴,樽澤賢一,小林尚人,古澤順子,市川伸一, SMOKAグループ:東京大学木曽観測所写真乾 板のデジタル化,天文月報,114,523-533 (2021).
- [12] Ohsawa, R., et al.: Development of a real-time data processing system for a prototype of the Tomo-e Gozen wide field CMOS camera, *Proc. SPIE*, **9913**, 991339 (2016).
- [13] PostgreSQLグローバル開発グループ: PostgreSQL 12.4文書, https://www.postgresql.jp/document/12/html/ index.html, 日本PostgreSQL ユーザ会, 2020 (最終 閲覧日: 2022年7月20日).
- [14] Power, R. A.: Large Catalogue Query Performance in Relational Databases, *Publications of the Astronomical Society of Australia*, 24, 13–20 (2007).
- [15] Kotani, T., et al.: MITSuME.Multicolor Imaging Telescopes for Survey and Monstrous Explosions, *Il Nuovo Cimento C*, 28, 755–758 (2005).
- [16] 高橋英則:木曽観測所報告,木曽シュミット シンポジウム2021,2021-10-04,http://www.ioa. s.u-tokyo.ac.jp/kisohp/RESEARCH/symp2021/ kisosymp2021\_NTakahashi.pdf,(最終閲覧日:2022 年7月20日).
- [17] 小澤武揚,小野里宏樹,中島康:光学赤外線天文 学観測データアーカイブにおける検索高速化の 研究,国立天文台報,投稿済(2022).