

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、木曾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 $2000 \times 1128 \times 84 = 1$ 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の値を採ることとする)。

Tomo-eは一晩約30 TB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¹⁾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。

- (1) 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

1) 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ェアを再利用することやSMOKA用に用意されているハードウェアや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

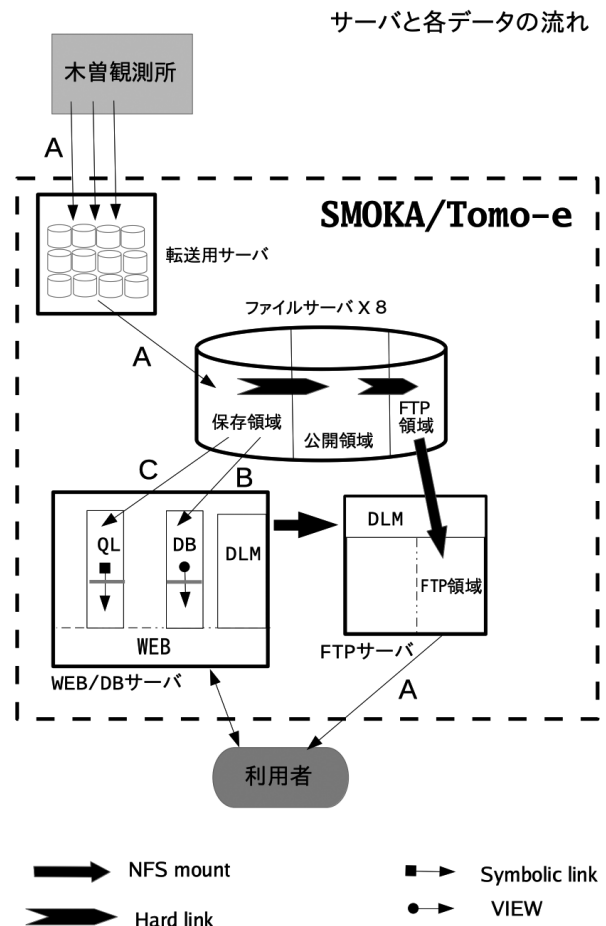


図1：SMOKA/Tomo-eシステムの全体像。本システムで用いられているサーバと各データ（以下のABC）の流れを示した。A：スタック済みデータ（4章参照）。B：スタック済みデータから抽出されたメタデータなどの情報。データベースサーバ（DB）に保存される（5章参照）。C：スタック済みデータから作成されたクイックルックデータ。WEBサーバの非公開の保存領域（QL）に保存される（7章参照）。各サーバの仕様は表1にまとめられている。

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.

	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

ここで保存される. 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.

TMQyyyymmddNNNNNNNNDD.fits

q: カメラモジュール番号 (1-4)

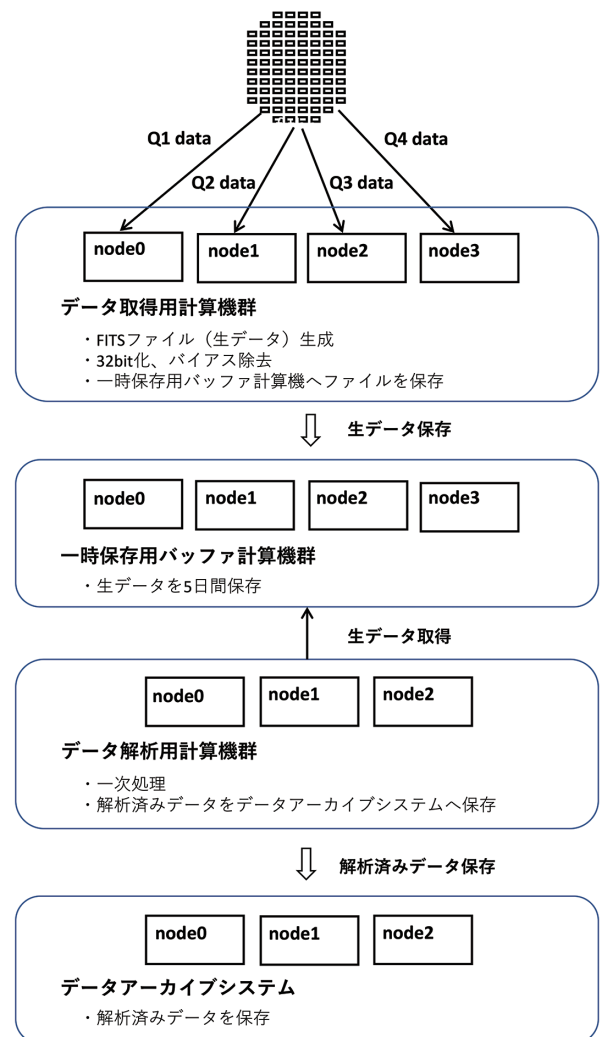


図2：Tomo-e Gozen のデータ生成から保存の流れ. 図中の計算機システムは全て木曽観測所内に設置されているものだが, そのうちデータアーカイブシステムは本館, それ以外はシュミット望遠鏡ドーム建物内に設置されている.

yyyymmdd: 日付 (UTC)

NNNNNNNN: 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²⁾)

一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に、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。

aTMQyyyymmddNNNNNNNNDD.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; 単体ドライブ一つが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400Mbps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3,760Mbps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章で最後に述べた例外 (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.ia.s.u-tokyo.ac.jp/images/tomoe_distribution_of_sensors.png) に記載されている。

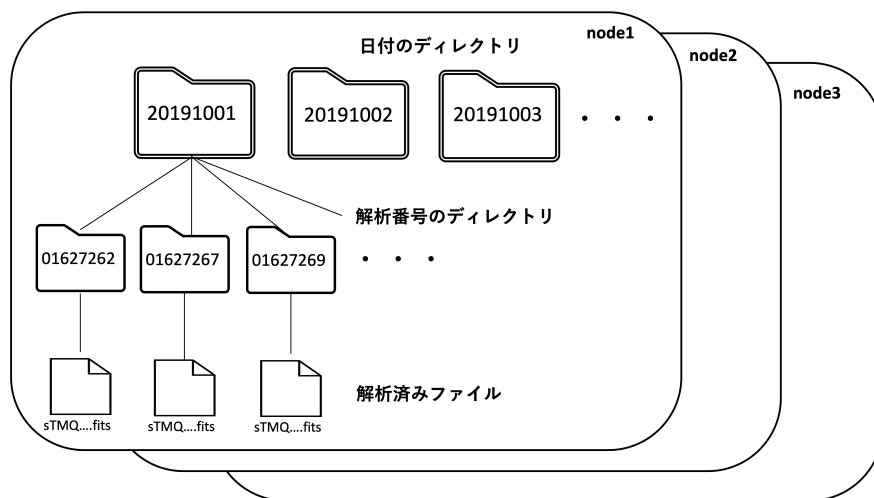


図3: データアーカイブシステムのファイル・ディレクトリ構造。木曽観測所に設置されている Tomo-e Gozen データアーカイブシステムは3台並列の構成となっており、観測日毎のディレクトリが3台に分割して置かれている。ここから SMOKA/Tomo-e システムにスタック済みデータが送られている。

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徴として、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。

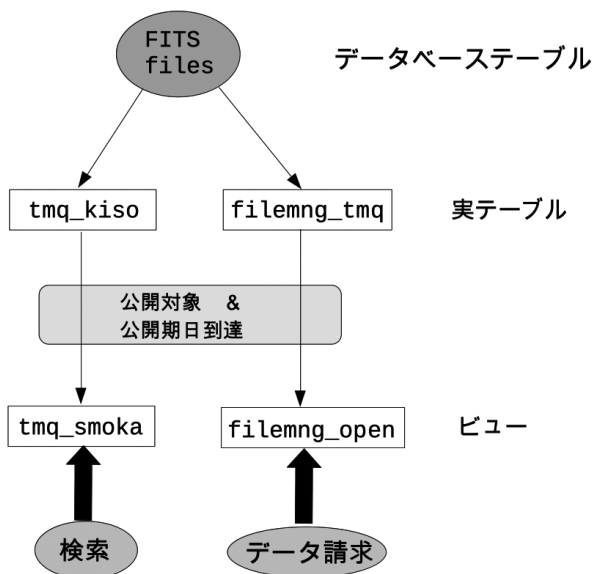


図4: データベーステーブルの相互関係。スタック済みデータの管理情報およびFITSヘッダから抽出したメタデータはそれぞれ実テーブルである。ファイル管理テーブル `filemng_tmq` およびヘッダキーワードテーブル `tmq_kiso` に登録される。公開対象かつ公開日に達したエントリは公開データ用ビューである `tmq_smoka` および `filemng_open` として参照可能になり、データ検索とデータ請求に利用される。

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的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思われ、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 ☐ NED Resolve
Coordinates	RA RA(J2000) DEC DEC(J2000)
Observation Date	YYYY-MM-DD
Output Format	
☒ TABLE (max 5,000 rows, HTML) ☐ ASCII (max 20,000 rows, text)	
Search	Reset to defaults Help...
Output Options	
Order by:	Rows per page:
EXP_ID reverse	100

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、 $38 < \text{CRPIX1} < 1979$, $17 < \text{CRPIX2} < 1117$ 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 $0 \leq X \leq 2001$ かつ $0 \leq Y \leq 1129$ 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³⁾。

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

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 `fconesearch` `tmqsmokaCRVAL` 関数を利用した。

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。

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 to Retrieve	FRAME_ID	RA_FRAME	DEC_FRAME	DATE_OBS	TIME_OBS	Pinpoint_X	Pinpoint_Y	EXPTIME	EXP_ID
1	thumbnail	☐	TMQ1201910020015182815	0:41:37.6	+41:09:07.3	2019-10-02 weather	18:29:28.5	1634.3	213.8	6.00	151828
2	thumbnail	☐	TMQ4201910020015183315	0:43:33.5	+41:12:32.0	2019-10-02 weather	18:31:10.5	535.5	377.1	6.00	151833
3	thumbnail	☐	TMQ1201910020015188415	0:41:40.5	+41:08:58.5	2019-10-02 weather	19:39:51.2	1606.9	206.3	6.00	151884
4	thumbnail	☐	TMQ4201910020015188915	0:43:36.1	+41:12:20.6	2019-10-02 weather	19:41:33.6	511.6	367.2	6.00	151889
5	thumbnail	☐	TMQ4201910040015233915	0:43:31.3	+41:13:19.7	2019-10-04 weather	15:28:43.9	556.2	417.5	6.00	152339

図6: ピンポイント検索の結果画面例。

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

[Click here 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	SHOT	RA_SHOT	DEC_SHOT	DATE_OBS	TIME_OBS	EXPTIME
1	thumbnail	☐	TMQX2019100200151827XX	00:43:01.2	+37:08:17	2019-10-02 weather	18:29:09.7	6.00
2	thumbnail	☐	TMQX2019100200151828XX	00:43:01.1	+37:32:16	2019-10-02 weather	18:29:28.5	6.00
3	thumbnail	☐	TMQX2019100200151829XX	00:40:14.6	+37:32:15	2019-10-02 weather	18:29:47.6	6.00
4	thumbnail	☐	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 1

More Information Date:2020-1-5
 ExpID:219336-221447 (2091 exposures)

Table (HTML)
 Ascii (Text)

2020/1 Stacked Data [Weather and Skymonitor](#)

1	O	LINK
2	O	LINK
3	O	LINK
4	O	LINK
5	O	LINK
6	O	LINK
7	-	LINK
8	O	LINK
9	O	LINK
10	O	LINK
11	O	LINK

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。DS9⁴⁾およびXPA⁵⁾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

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

5) <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。

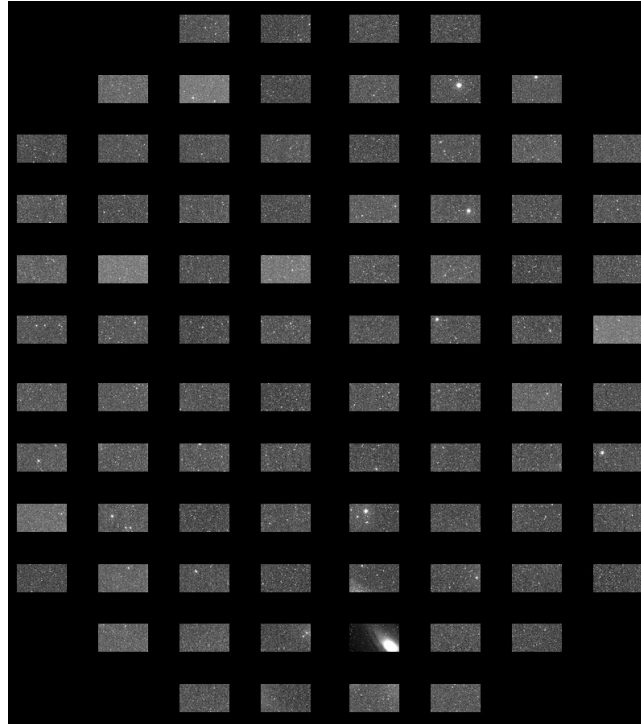


図10：SHOT 画像の例。図9の Thumbnail を含む SHOT である。

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ェアの ImageMagick⁶⁾ 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

6) <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核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

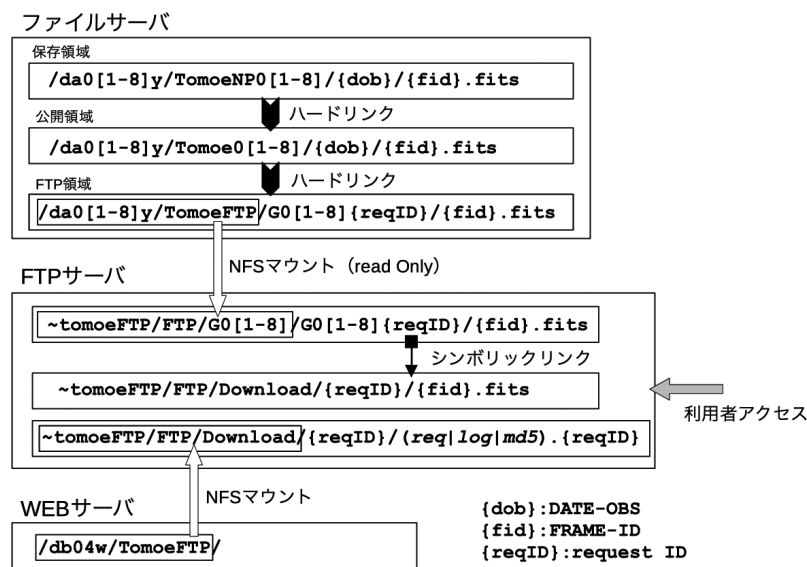


図11: スタック済みデータの配置。スタック済みデータはファイルサーバの八つのファイルシステム（/da0[1-8]y）上の保存領域のどれかに置かれ、公開日に達するとそれぞれのファイルシステムの公開領域にハードリンクされる。さらに、データ請求を受けたスタック済みデータはそれぞれのファイルシステムのFTP領域にハードリンクされる。FTPサーバにはファイルサーバのFTP領域がNFSマウントされる（~tomoeFTP/FTP/G0[1-8]）。利用者がアクセスしやすいよう、それらの下のファイルは請求毎に作成されるディレクトリ（~tomoeFTP/FTP/Download/{reqID}）下にまとめてシ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

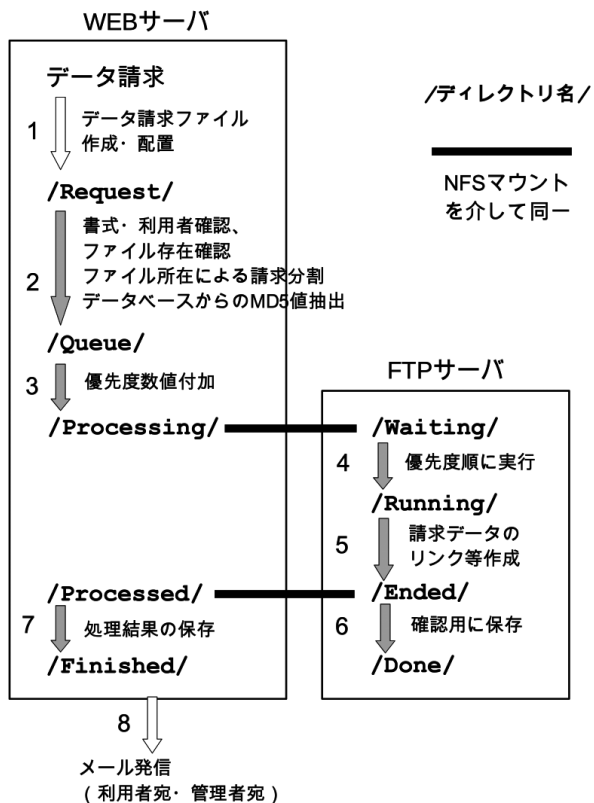


図12: Data Link Manager (DLM) によるデータ請求処理の流れ。利用者がデータ請求を行うとデータ請求ファイルが作成されて「Request」ディレクトリに置かれる（処理1）。その後数字の順に処理が行われ、データ請求ファイルや処理の途上で作成された関連ファイルが次々と移動される。最後（処理8）に処理の結果が利用者や管理者に電子メールで通知されてデータ請求処理は終了する。なお、WEBサーバ上の「Processing」ディレクトリと「Processed」ディレクトリはそれぞれFTPサーバの「Waiting」ディレクトリと「Ended」ディレクトリにNFSマウントされており実質上同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

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ェ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⁷⁾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ある（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

7) <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1.8 TB）を保存することになった。これは、データ処理／解析ソフトウェ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、300 m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は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、木曾観測所、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ます。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, 古莊玲子, 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, 古莊玲子, 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, 古莊玲子, 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, 古莊玲子, 西村高德, 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, 古莊玲子, 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: 東京大学木曾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: 木曾観測所報告, 木曾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).