

Tomo-e Gozen data platform on mdx

Satoshi TAKITA (IoA/UT)

about mdx



9 大学、2 研究機関が共同運営する
データ科学・データ駆動科学・データ活用応用
にフォーカスした高性能仮想化環境
(学術版の AWS のようなもの)

実体は東京大学 柏 II キャンパス

利用料金が発生

民間に比べると 1/10 程度

「JHPCN 共同研究課題」を通した研究費補助

「データ活用社会創成プラットフォーム」は
用途に応じてオンデマンドで**短時間に構築・拡張・融合**できる
データ収集・集積・解析機能を提供するプラットフォーム。

データ活用社会創成プラットフォーム 3 本柱

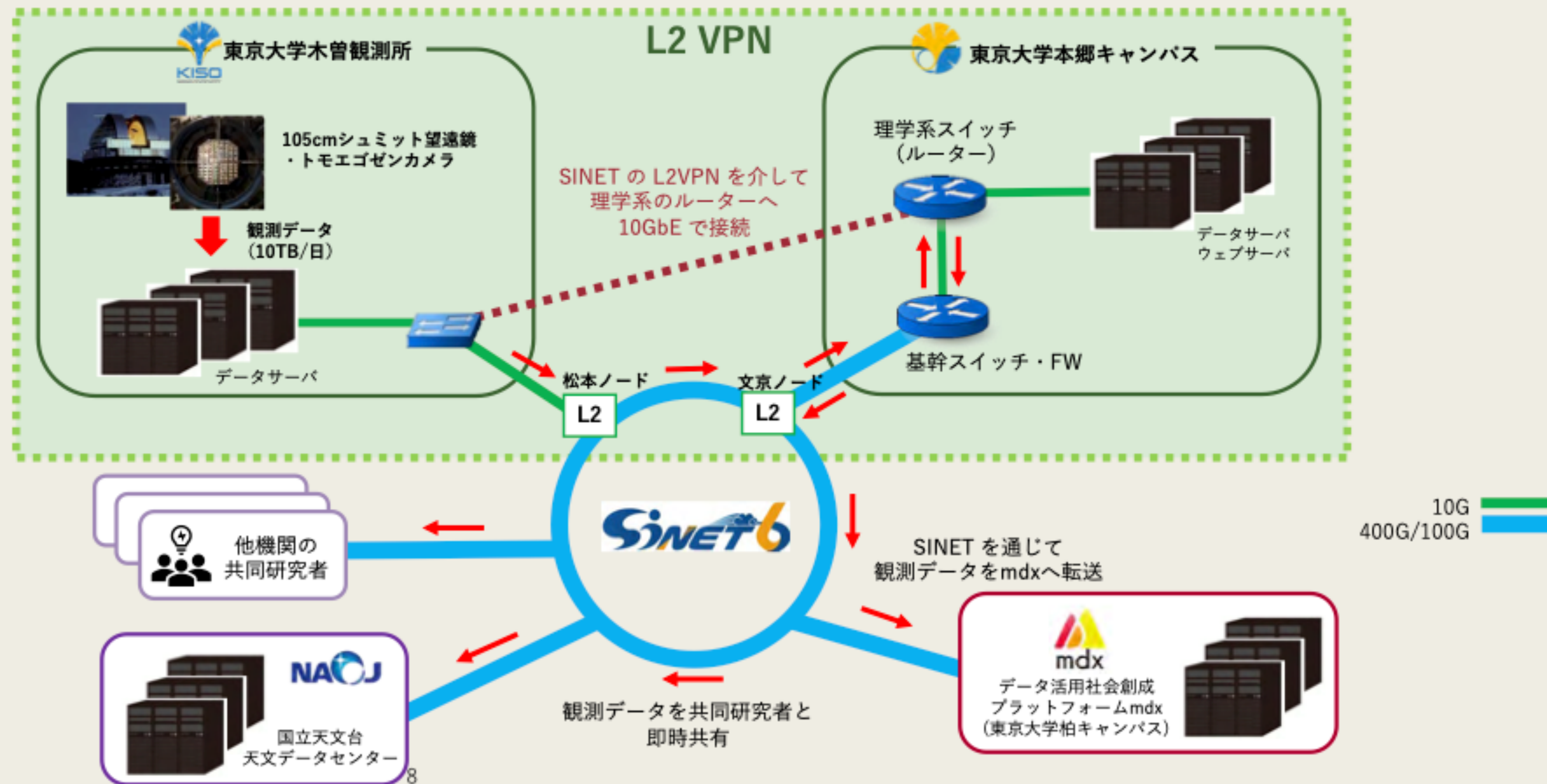
- 1 SINETを活かしたリアルタイム収集・集積・解析環境の動的な構築**
遠隔地のセンサーやストレージ、データプラットフォームの計算資源、ストレージをつないで、リアルタイムに入力から出力を得られるアプリケーションごとの収集・集積・解析環境（仮想データプラットフォーム：仮想DP）を、使いたいときに即時に構築する
SINETモバイル基盤によりセンサー等のデータを安定してセキュアにつなぐ
- 2 高性能計算環境によるデータ科学と計算科学の融合**
データ科学、計算科学の手法を融合し、さらに国内最高の計算環境を用いて他に無い高精度の予測を行えるようにする
- 3 異種データ・異種知識の融合活用の推進と利用者支援**
様々な分野のデータ保持者、解析者、利用者が産学にまたがって連携するコミュニティを形成し、新たな価値創造につなげる。
データ活用を目指す利用者へのコンサルティングや開発支援を実施する。

<https://mdx.jp>

mdx and kiso-SINET



ネットワーク構成



project mdx

背景

- 木曽観測所の計算機群の問題
 - + 物理的、電力 (熱) 的制約
 - "実効的に" 利用可能なストレージは 1 PB 程度
 - + 管理コスト (観測所の人員)

目的

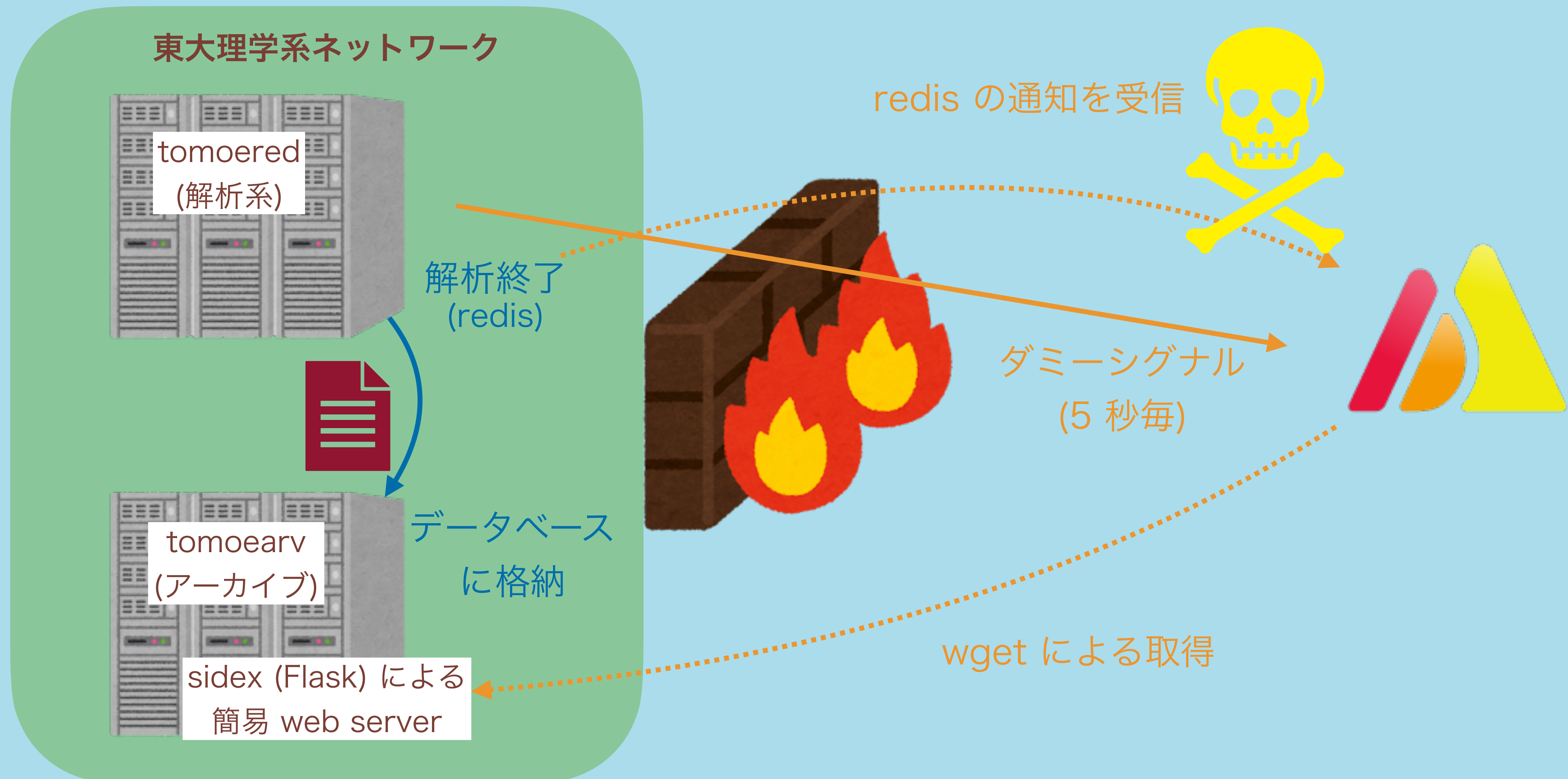
- データ解析機能の一部を木曽観測所から mdx へオフロード
 1. データアーカイブ
 2. ライトカーブデータベース
- mdx を起点としたデータ発信

1. alternative data archive

データアーカイブの作成

- スタック済み画像のリアルタイム転送
 - + Tomo-e のデータ解析で利用されている redis 通知
 - + sidex (Flask ベースの簡易 http server) 越しの wget
- データベース
 - + Apache Arrow 形式 (列指向データベース)
 - + pyarrow/duckdb
- web UI
 - + 現状のアーカイブを流用・強化

data transfer



database

観測ログ

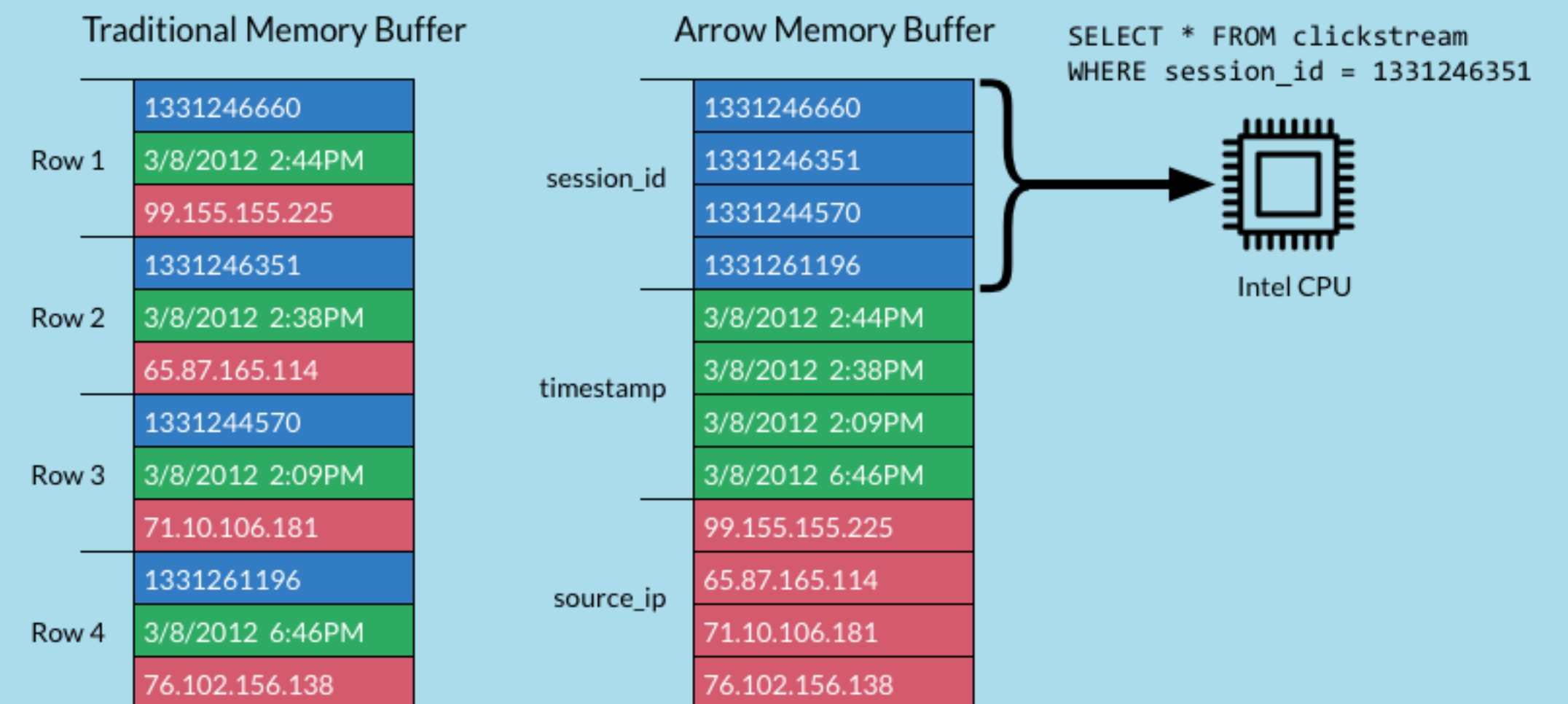
- 観測システムとは独立
- 毎晩の観測終了後に一度だけ追記
- 検索性能が重要

=> OLAP 的なデータベースが有効

Apache Arrow

- 列指向データベース
- parquet によるファイルベース
- pyarrow や duckdb 等のライブラリ
- (PG-Strom で PostgreSQL に組み込み可)

	session_id	timestamp	source_ip
Row 1	1331246660	3/8/2012 2:44PM	99.155.155.225
Row 2	1331246351	3/8/2012 2:38PM	65.87.165.114
Row 3	1331244570	3/8/2012 2:09PM	71.10.106.181
Row 4	1331261196	3/8/2012 6:46PM	76.102.156.138



<https://arrow.apache.org>

interface

Tomo-e Gozen Archive

観測年ごとに表示

2024

Available Dates

2024-05-09 (1416)	2024-05-08 (48)	2024-05-07 (8)	2024-05-06 (8)
2024-05-05 (1235)	2024-05-04 (1468)	2024-05-03 (1468)	2024-05-02 (228)
2024-05-01 (8)	2024-04-30 (8)	2024-04-29 (8)	2024-04-28 (1445)
2024-04-27 (166)	2024-04-26 (8)	2024-04-25 (1428)	2024-04-24 (8)
2024-04-23 (8)	2024-04-22 (8)	2024-04-20 (8)	
2024-04-18 (624)	2024-04-17 (1358)	2024-04-16 (8)	
2024-04-14 (1274)	2024-04-13 (848)	2024-04-12 (1453)	
2024-04-10 (819)	2024-04-09 (1144)	2024-04-08 (8)	
2024-04-06 (539)	2024-04-05 (8)	2024-04-04 (1347)	
2024-04-02 (679)	2024-04-01 (1602)	2024-03-29 (1525)	
2024-03-27 (1641)	2024-03-26 (1049)	2024-03-25 (8)	
2024-03-23 (41)	2024-03-22 (498)	2024-03-21 (1482)	
2024-03-19 (8)	2024-03-18 (242)	2024-03-15 (892)	
2024-03-05 (8)	2024-03-04 (501)	2024-03-03 (534)	
2024-03-01 (766)	2024-02-29 (8)	2024-02-28 (1513)	
2024-02-26 (1131)	2024-02-25 (8)	2024-02-24 (571)	
2024-02-22 (8)	2024-02-21 (8)	2024-02-20 (99)	
2024-02-18 (8)	2024-02-17 (8)	2024-02-16 (54)	
2024-02-14 (447)	2024-02-13 (1291)	2024-02-12 (1347)	

DataTables (JavaScript) を用いたテーブル描画
カラムに対する検索
"FPS, nFits" カラムの追加

Tomo-e Gozen Archive / 2024-05-09

Show 10 entries

Search:

Date	Exp ID	nFits	Object	Observer	Project	Format	Gain	FPS	(RA, Dec)
2024-05-09 08:45	1201968	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x360	HIGH	2.0	—
2024-05-09 08:51	1201969	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x360	HIGH	1.0	—
2024-05-09 09:00	1201970	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x360	LOW	2.0	—
2024-05-09 09:06	1201971	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x360	LOW	1.0	—
2024-05-09 09:15	1201972	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	400x24x40000	HIGH	57.5	—
2024-05-09 09:30	1201973	84	DARK	Tomo-e Gozen Scheduler	Kiso Observatory	400x240x4500	HIGH	18.9	—
2024-05-09 09:37	1201974	84	FLAT	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x180	HIGH	1.0	—
2024-05-09 09:43	1201975	84	FLAT	Tomo-e Gozen Scheduler	Kiso Observatory	2000x1128x180	LOW	1.0	—
2024-05-09 10:16	1201988	3	J0459+2600_dith1	Kiso Observatory	Twilight Survey	2000x1128x18	LOW	2.0	(74.767, +26.007)
2024-05-09 10:17	1201989	3	J0459+2600_dith2	Kiso Observatory	Twilight Survey	2000x1128x18	LOW	2.0	(74.767, +26.407)

Showing 1 to 10 of 1,416 entries

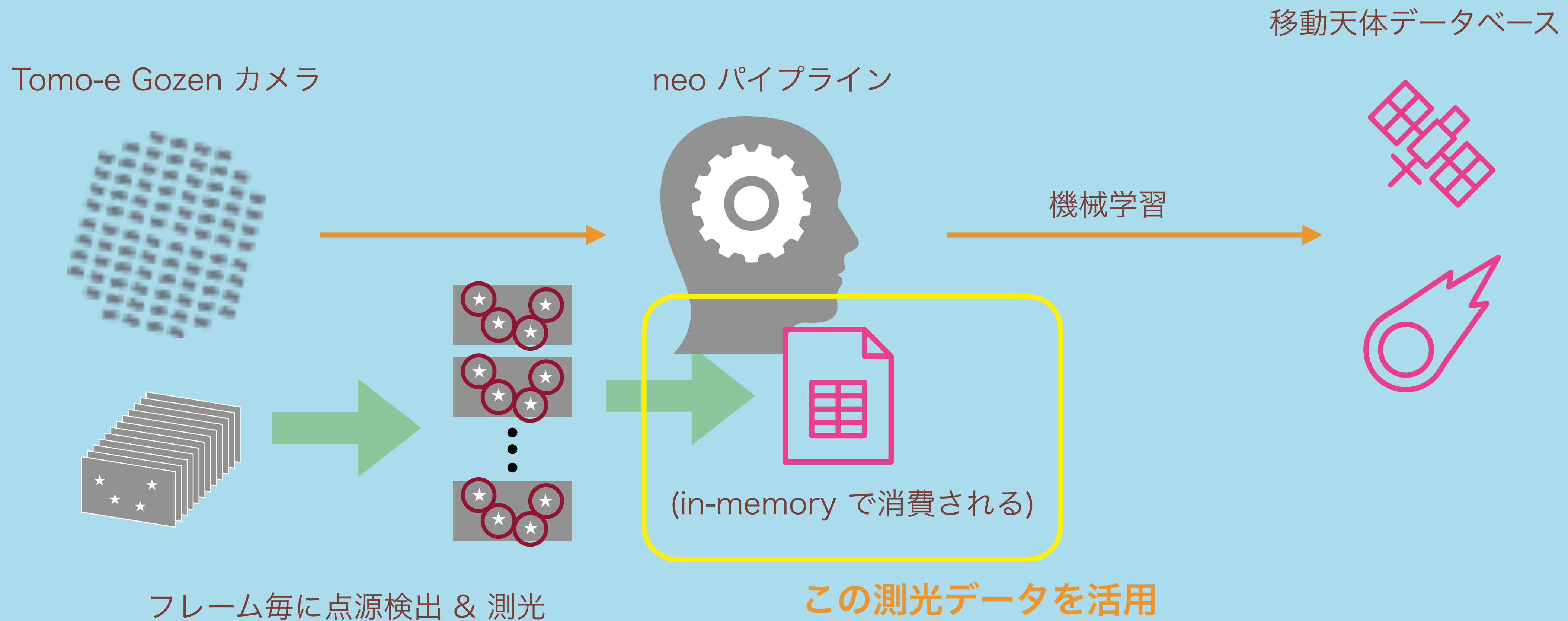
Previous 1 2 3 4 5 ... 142 Next

2. light-curve database

Tomo-e が誇る「秒」スケールのライトカーブ

- neo / 高速移動天体検出パイプラインによるフレーム単位での測光データ
 - + 6 or 9 秒のライトカーブ
- Gaia データとのマッチング
 - + 数時間や日を跨いだデータの結合
 - + flux calibration
- データ発信
 - + skyatlas との連携
 - + デブリデータ

neo data



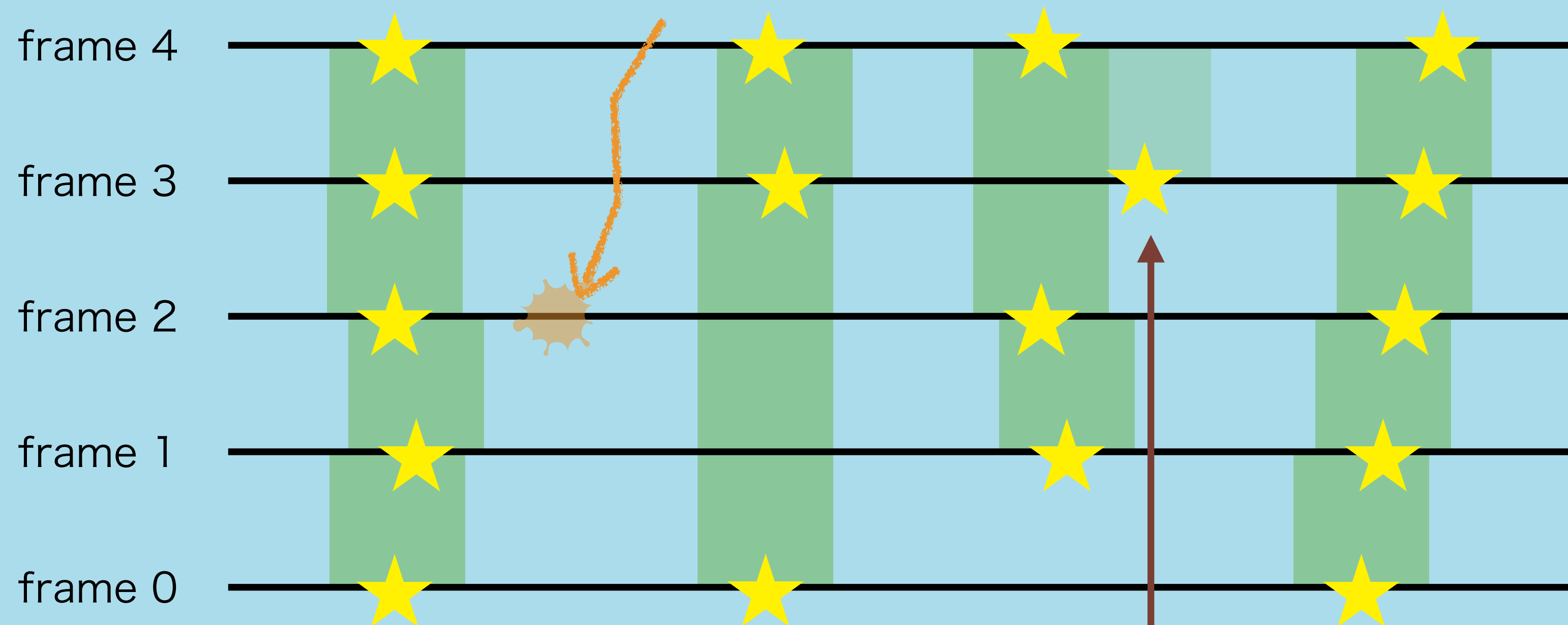
試験として 1 exposure (84 fits) 分のデータを出力

data grouping / matching

画像上の x, y 座標 $\rightarrow$ ± 3 px でマッチング

直近の検出位置を基準

$\Rightarrow$ Gaia DR3 と 1 arcsec でマッチング



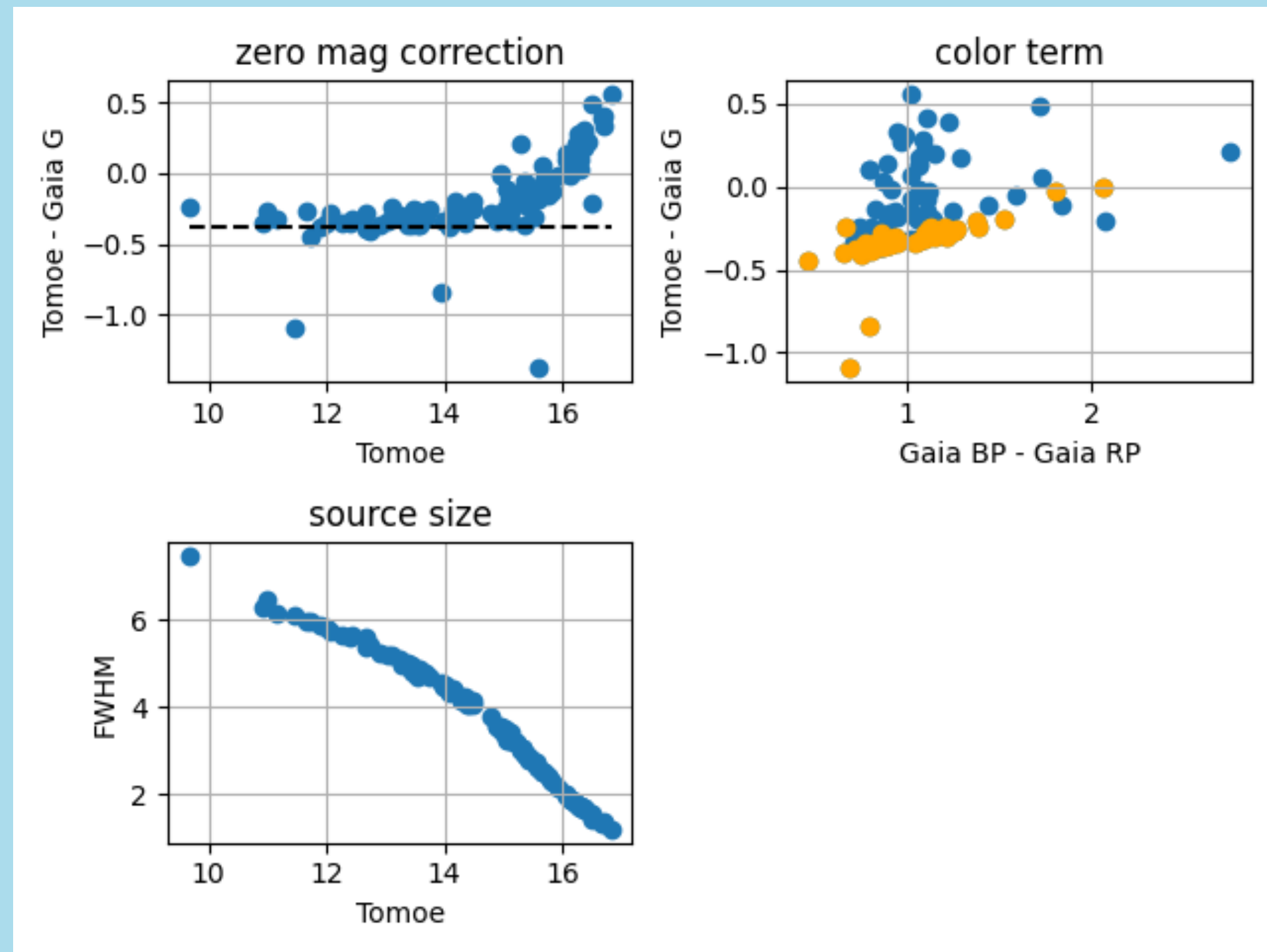
宇宙線イベントや
高速移動天体は
単発のイベント

seeing dance により
外れ値がでることもある

移動量の少ない小惑星やデブリは引っかかる

flux calibration

Gaia との比較により zeromag, colour-term を導出



現状の問題点

- Tomo-e データは sep (SourceExtractor) の magauto の値を出力
- > 固定 aperture の測光値が必要

超新星チームの知見を再利用

- 21 px diameter の測光値
- Gaia データから Tomo-e mag への変換

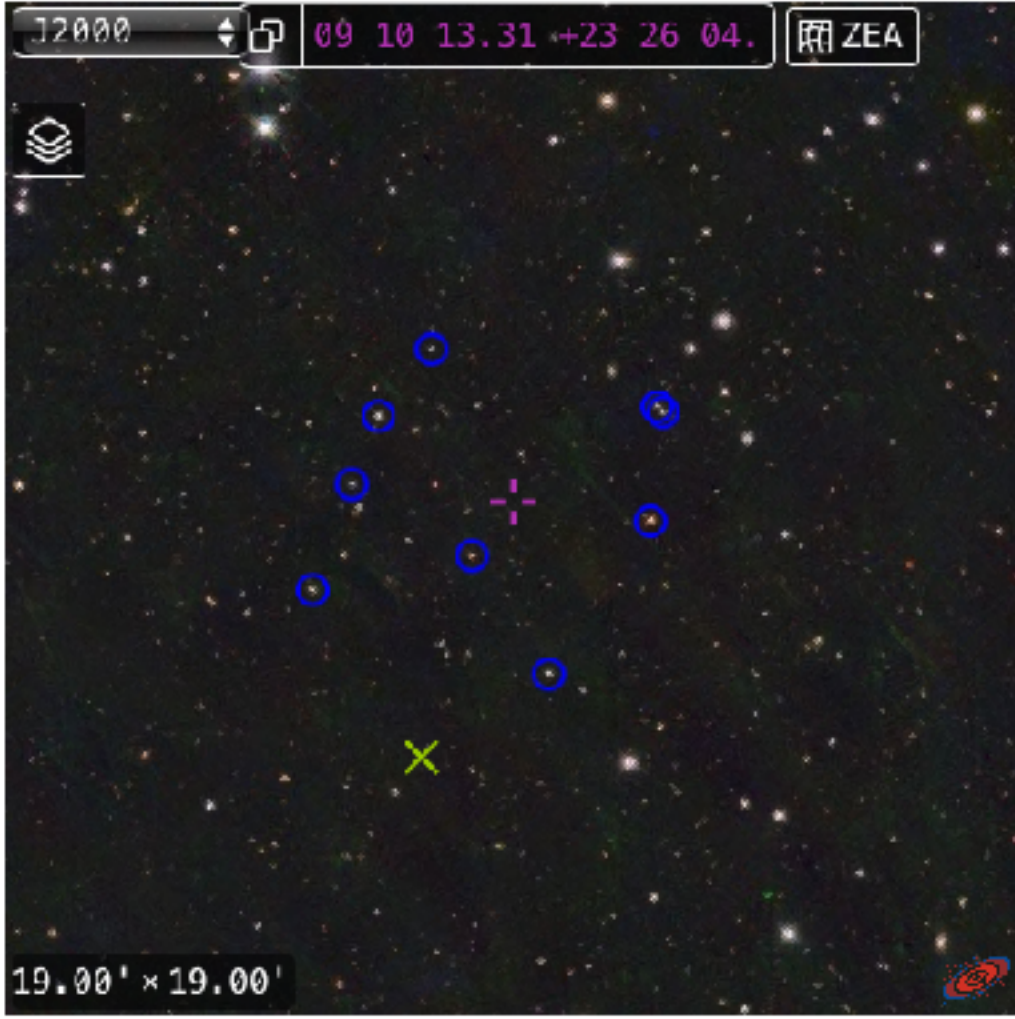


interface

Tomo-e Gozen (Sub-)Second-scale Lightcurve

12000

09 10 13.31 +23 26 04. 天 ZEA



19.00' × 19.00'

Target:
137.55545, 23.43462
Jump

Radius: 5 [arcmin] (max: 60)

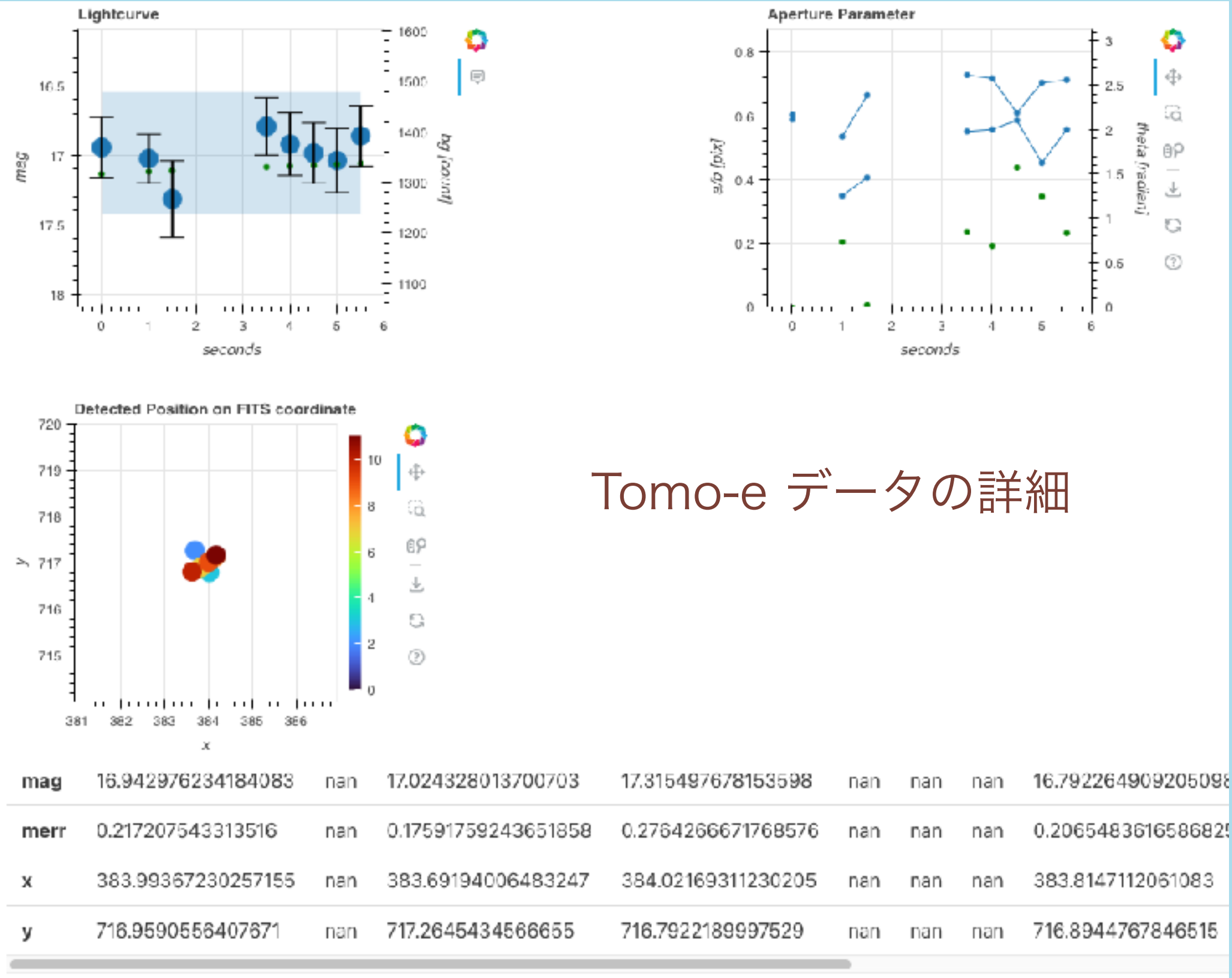
Date: 2024/04/04 to 2024/04/04

Search

Gaia objects
Show 10 entries

Gaia source ID	R.A.	Dec.	#Obs	Gaia GMag
687130931262755968	137.54348	23.38293	1	16.366
687132340012028496	137.62378	23.40925	1	16.204
687132992847054464	137.61042	23.44218	1	16.373
687143060250399232	137.50889	23.43058	1	14.973
687144228481502848	137.56951	23.41987	1	17.011
687144778237313920	137.60136	23.46345	1	15.716
687145053115220480	137.58326	23.48434	1	16.854
687145392416366592	137.50461	23.46465	1	16.613

青○: Gaia とマッチングした天体
緑x: Gaia にない天体



Tomo-e データの詳細

issues to do

データアーカイブ

- ブロックリストの対応
- ストレージ容量問題

利用料金の問題で mdx のみでは困難
=> HPCI 共有ストレージ利用研究課題

ライトカーブ

- データの準備
 - + neo パイプラインの改修・常駐化
- データベースの検索が遅い
 - + Tomo-e と Gaia の JOIN が問題
 - + 恐らく parquet (ファイルベース) で扱っているのが原因
 - => Tomo-e 側に Gaia の情報 (一部) を加えるべきか

summary

mdx 上に新たなデータプラットフォームを開発中

- JHPCN 共同研究課題や HPCI 共有ストレージの利用
- SINET による 100 Gbps ネットワーク

データアーカイブの機能強化・「秒」スケールライトカーブ

- skyatlas との連携
- 動画データの有効活用
- + 産学官連携

=> 分野を超えたデータ活用へ

