

データアーカイブの mdx への移行計画

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)

東京大学 柏 II キャンパスに一号機 (mdx-I)
(大阪大学でも mdx-II が稼働開始)

民間に比べると 1/10 程度の利用料金

「JHPCN 共同研究課題」を通じた研究費補助



「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ーを形成し、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>

changes on mdx

mdx の利用促進のため、2025 年度より利用料金が改定

- CPU は半額 (GPU は変更無し)
- HDD は無料 (SSD は変更無し)

機械学習関係は利用率が高いが
それ以外はあまり利用されていない

=> これまでのストレージ問題が解消

+ 本格的にデータアーカイブを mdx に移植させる

+ CPU の増加により、データ解析基盤としての機能も

データ公開は、データ活用を謡う mdx からの評価も非常に高い



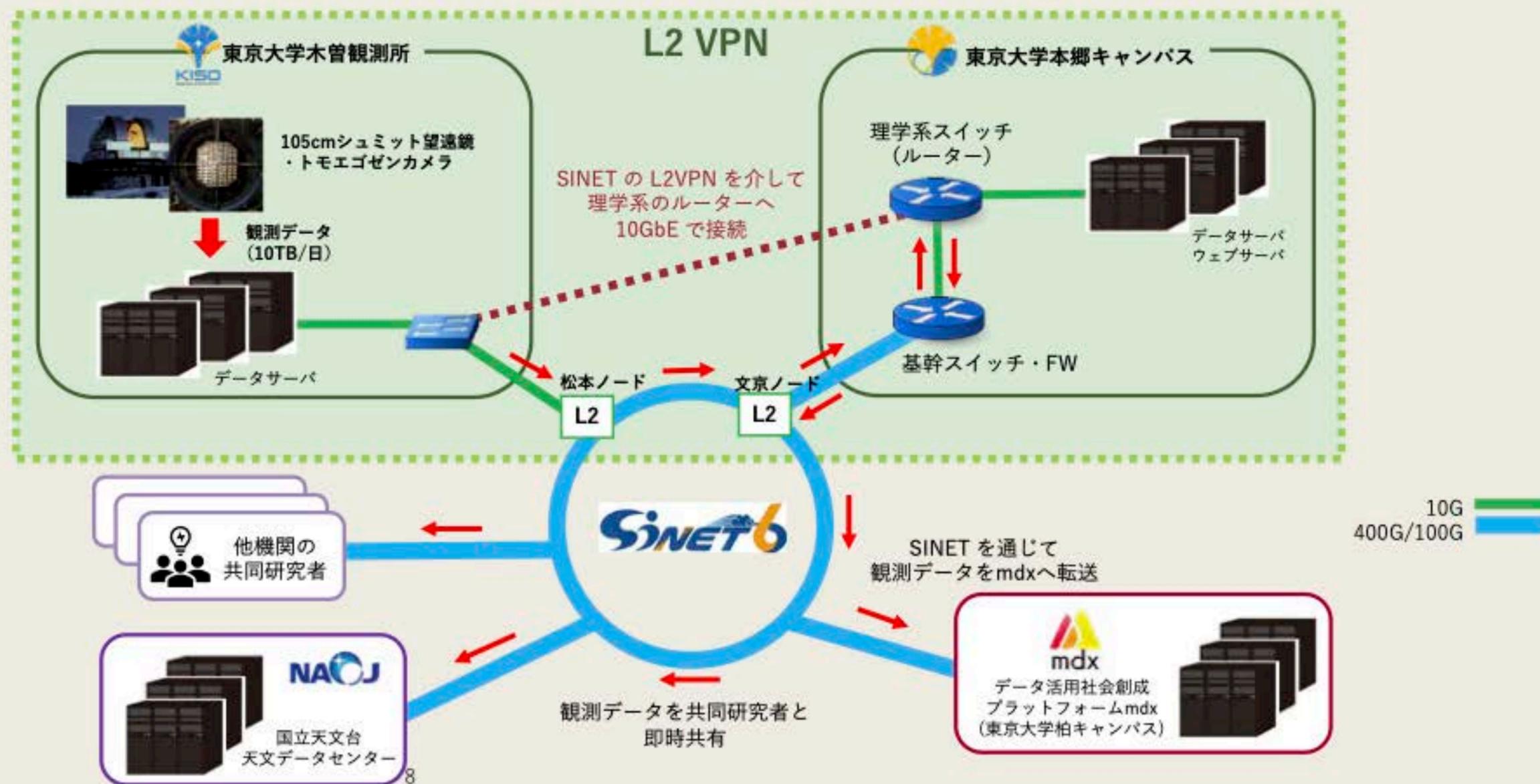
mdx利用事例 CASE 5

激動する宇宙を捉えたデータを
木曾からmdxへ、そして世界へ

mdx and kiso-SINET



ネットワーク構成



current data archive

現状のデータアーカイブ

- 観測システムと密に連携
- スタック画像のみ
- 二種類のアーカイブ

+ 内部向け (全データへのアクセス)

+ 一般向け (180 日経過したサーベイデータ)

(サーベイ以外のデータは SMOKA から公開)

=> 新機能追加にハードル

動画データの取得問題

他人のデータを見ないという紳士協定

TOMO E GOZEN

Tomo-e Gozen Archive

Available Dates

2025-05-19 (0)	2025-05-18 (418)	2025-05-17 (11)	2025-05-16 (11)
2025-05-15 (422)	2025-05-14 (673)	2025-05-13 (1065)	2025-05-12 (1075)
2025-05-11 (12)	2025-05-10 (182)	2025-05-09 (11)	2025-05-08 (654)
2025-05-07 (551)	2025-05-06 (100)	2025-05-05 (12)	2025-05-04 (1204)
2025-05-03 (39)	2025-05-02 (640)	2025-05-01 (584)	2025-04-30 (861)
2025-04-29 (1108)	2025-04-28 (12)	2025-04-27 (881)	2025-04-26 (1374)
2025-04-25 (940)	2025-04-24 (1066)	2025-04-23 (16)	2025-04-22 (12)
2025-04-21 (899)	2025-04-20 (684)	2025-04-19 (1183)	2025-04-18 (418)
2025-04-17 (1113)	2025-04-16 (771)	2025-04-15 (325)	2025-04-14 (72)
2025-04-13 (11)	2025-04-12 (451)	2025-04-11 (1398)	2025-04-10 (11)
2025-04-09 (1501)	2025-04-08 (1101)	2025-04-07 (1606)	2025-04-06 (0)
2025-04-05 (0)	2025-04-04 (1444)	2025-04-03 (1379)	2025-04-02 (11)
2025-04-01 (11)	2025-03-31 (11)	2025-03-30 (1304)	2025-03-29 (349)
2025-03-28 (78)	2025-03-27 (11)	2025-03-26 (1105)	2025-03-25 (1255)
2025-03-24 (26)	2025-03-23 (1851)	2025-03-22 (1618)	2025-03-21 (772)
2025-03-20 (1669)	2025-03-19 (789)	2025-03-18 (338)	2025-03-17 (1839)
2025-03-16 (0)	2025-03-15 (0)	2025-03-14 (0)	2025-03-13 (0)
2025-03-12 (0)	2025-03-11 (23)	2025-03-10 (166)	2025-03-09 (1973)
2025-03-08 (11)	2025-03-07 (1625)	2025-03-06 (370)	2025-03-05 (0)

TOMO E GOZEN

Tomo-e Gozen Public Archive

Available Dates

2024-11-19 (420)	2024-11-18 (2038)	2024-11-17 (204)	2024-11-15 (132)
2024-11-13 (874)	2024-11-12 (1323)	2024-11-11 (794)	2024-11-09 (1556)
2024-11-08 (1201)	2024-11-07 (1601)	2024-11-06 (1344)	2024-11-05 (115)
2024-11-04 (318)	2024-11-03 (614)	2024-11-02 (751)	2024-10-31 (590)
2024-10-30 (888)	2024-10-26 (197)	2024-10-25 (144)	2024-10-21 (424)
2024-10-20 (714)	2024-10-19 (140)	2024-10-18 (48)	2024-10-17 (341)
2024-10-16 (200)	2024-10-15 (137)	2024-10-14 (812)	2024-10-13 (1271)
2024-10-12 (521)	2024-10-11 (749)	2024-10-10 (669)	2024-10-09 (880)
2024-10-06 (471)	2024-10-02 (527)	2024-10-01 (993)	2024-09-30 (165)
2024-09-27 (605)	2024-09-26 (436)	2024-09-25 (750)	2024-09-24 (64)
2024-09-23 (523)	2024-09-22 (965)	2024-09-21 (145)	2024-09-20 (331)
2024-09-17 (116)	2024-09-13 (211)	2024-09-10 (600)	2024-09-09 (789)
2024-09-07 (360)	2024-09-06 (119)	2024-09-05 (138)	2024-09-04 (9)
2024-09-01 (144)	2024-08-30 (173)	2024-08-28 (90)	2024-08-26 (409)
2024-08-23 (190)	2024-08-21 (488)	2024-08-11 (972)	2024-08-10 (1076)
2024-08-09 (822)	2024-08-06 (314)	2024-08-02 (764)	2024-08-01 (832)
2024-07-31 (1412)	2024-07-29 (1243)	2024-07-28 (567)	2024-07-23 (1270)
2024-07-22 (707)	2024-07-21 (443)	2024-07-20 (130)	2024-07-19 (127)
2024-07-18 (108)	2024-07-09 (823)	2024-07-08 (1039)	2024-07-07 (1114)

new data archive

新たなデータアーカイブに求める機能

- 検索性能の強化

+ SQL -> NoSQL, 列指向型データベース

Apache Arrow + Arrow IPC format

polars (DuckDB)

- 動画データへのアクセス

+ "cube" データの mdx への転送

- 統一的なアーカイブ

+ ユーザ認証とアクセスコントロール



DuckDB

-  [Select by calender list](#)
-  [Search by coordinate](#)
-  [Data technical note](#)
-  [Return to project top](#)

blocklist?

blocklist って機能してますか？
現状の mdx archive では未対応

2021 年に定義されて以降、一度も手を付けられていない
= ブロックされているデータはない

Coordinate Block Registration

[Block List](#)

Login to register a new request

Filter:

Request ID	Effective Period	Description	Applicant Name	Email	Coordinates
00000004	2021-09-09 – 2022-09-09	Dummy Entry for Function Test	Ryou Ohsawa		(83.634070, 22.0155)
00000005	2021-09-09 – 2022-09-09	Dummy Entry for Function Test	Ryou Ohsawa		(83.634070, 22.0155)
00000006	2021-09-27 – 2022-09-27	Dummy Entry for Function Test	Sako Shigeyuki		(83.634070, 22.0155)
00000007	2021-09-27 – 2022-09-27	Dummy Entry for Function Test	Sako Shigeyuki		(83.634070, 22.0155)
00000008	2021-09-27 – 2022-09-27	Dummy Entry for Function Test	Tomoki Morokuma		(83.634070, 22.0155)
00000009	2021-11-07 – 2022-11-07	Coordinate Block Registration f	Shigeyuki Sako		(32.331000, 316.0000)
00000010	2021-11-07 – 2022-11-07	Coordinate Block Registration f	Ryou Ohsawa		(359.000000, 33.0000)
00000011	2010-01-01 – 2011-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000012	2011-01-01 – 2012-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000013	2012-01-01 – 2012-12-31	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000014	2013-01-01 – 2014-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000015	2014-01-01 – 2015-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000016	2015-01-01 – 2016-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000017	2016-01-01 – 2016-12-31	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000018	2017-01-01 – 2018-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000019	2018-01-01 – 2019-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000020	2019-01-01 – 2020-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000021	2020-01-01 – 2020-12-31	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000022	2021-01-01 – 2022-01-01	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000023	2021-11-08 – 2022-11-08	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000024	2021-11-08 – 2022-11-08	Block Request Insertion Test	Ryou Ohsawa		(10.000000, -10.0000)
00000025	2021-11-22 – 2022-11-22	Block Registration Test	Ryou Ohsawa		(60.000000, 38.0000)
00000026	2021-11-23 – 2022-11-23	Block Registration Test	Ryou Ohsawa		(70.000000, 32.0000)

future plans

利便性向上のための機能追加 (予定・願望)

- JWT (json web token) によるコマンドラインアクセス
- advanced search

+ 複数条件による検索 (SQL は叩かない)

- API 周りの整備

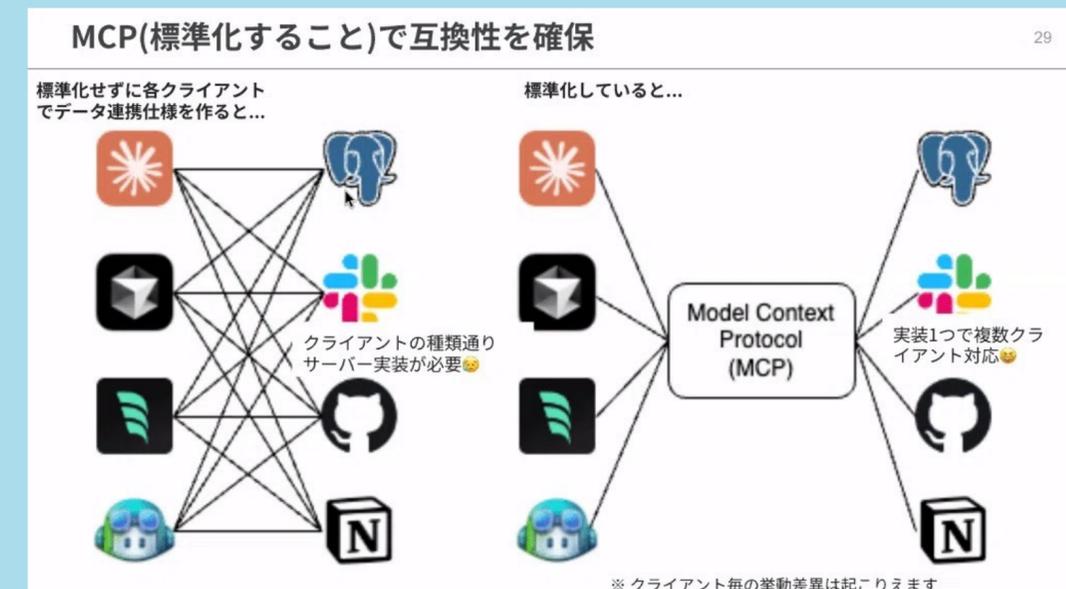
(既存の flask ベースからの脱却)

=> 今後の機械学習を見据えて FastAPI に移植?

- API ドキュメントの自動生成

- MCP (Model Context Protocol) との連携

+ 生成 AI 向けの共通プロトコル (なんかすごいらしい)



<https://note.com/fujimohige/n/n93f7d448eac7>

summary

mdx 上に新たなデータプラットフォームを構築、試験運用中

- 統一的なインターフェース
- + 動画データの取得
- + ユーザ認証によるアクセスコントロール
- NoSQL (Apache Arrow) による検索性能の向上
- どこかでちゃんと書き直したい

過去のデータの転送

- => ある程度たまったところで一般公開へ
- + ブロックリストの利用？

