

jh251015

多次元高精細地表情報（MHESD）の地球表層科学的基盤構築

早川裕弐（北海道大学）

概要

本研究は、多次元高精細地表情報（MHESD）について、ユーザ自身で計測したデータの保管・共有を効率化する情報基盤を構築するとともに、データアーカイブを可視化することで、多様な学術的利用の可能性を開拓することを目的とする。今年度は、mdx のストレージ領域を用いて Nextcloud によるユーザインタフェースを開発するとともに、archivematica を導入してメタデータ管理の枠組みを構築した。さらに、大規模言語モデル（LLM）を用いて自然言語から ISO 19115-1:2014 準拠の地理情報メタデータを抽出する試験を実施し、対話型のメタデータ整理ダイアログのプロトタイプを開発中である。これにより、Nextcloud から投入された MHESD の意味づけをユーザベースで行い、archivematica に接続するワークフローの設計を進めている。今後は GakuNin RDM 等の外部基盤との連携、および国内外の空間情報基盤との相互運用に向けたプロトタイプの構築・公開へと展開する。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

mdxI

(2) 課題分野

データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

早川裕弐（北海道大学）：研究総括、主導

齋藤 仁（名古屋大学）：データ可視化手法開発

小倉拓郎（兵庫教育大学）：データ出力手法開発

2. 研究の目的と意義

本研究は、地表面における空間情報の一種であ

る多次元高精細地表情報（Multidimensional High-definition Earth Surface Data, MHESD）について、ユーザ自身で計測したデータの保管・共有を効率化する情報基盤を構築するとともに、データアーカイブを可視化することで、多様な学術的利用の可能性を開拓することを目的とする。

MHESD は、地球環境科学分野における地形や森林などの自然景観構成要素を対象とし、無人航空機（UAS）から得られる空撮画像や、モバイル端末等で得られるレーザ点群などを含む。これらは地形や植生の高精細かつ高頻度な解析を可能とする一方で、データ容量が大きくなりがちで、その効率的なハンドリングが地形学、生態学等の分野で広く求められている。そこで、仮想化クラスタとオブジェクトストレージを備え、学術フェデレーション認証や広帯域 SINET との連携を有する mdx を活用することで、MHESD の効果的な利用を推進する。

MHESD の円滑な共有・可視化のワークフローが整備された場合、MHESD を取り扱わない他の

大規模空間情報基盤（G 空間情報センター，CSIS 研究用空間データ基盤 JoRAS 等）との連携をはかることで，日本における MHESD の一研究拠点として本プロジェクトが中核をなすことができる。さらに，代表者のアジアオセアニア地域における研究者ネットワーク，および欧米における同様な研究拠点（OpenTopography 等）との連携を通じて，その国際的な発展にも結び付けられる可能性がある。加えて，本研究を取り掛かりとして，地球表層に関わる自然・人文を含むあらゆる分野を通貫し，超学際的な MHESD の利活用を実験的に発展させることが期待される。

今年度は，研究計画全体のうちステップ I（データ保管・共有システム構築）の主要な構成要素を実装することを目標とした。

3. 当拠点の公募型共同研究として実施した意義

小型無人航空機や携帯端末などといった低空・地上のプラットフォームにより，写真測量，レーザ測量，マルチ／ハイパースペクトルセンサ等の多様なセンサによる MHESD が，研究者自身の手で容易に取得できるようになり，地球科学をはじめとしたさまざまな学術分野において，その活用が近年急速に広まっている。mdx はそうした大規模で多様なデータセットを適切に管理し，また即時に解析・共有することができるプラットフォームであり，地形学，森林科学，砂防，気象学，考古学，博物館学，歴史学，教育学などといったさまざまな分野において，MHESD の学際的な利活用に資するものである。

本研究では，必ずしも情報学に精通していない分野の研究者や学生，および実務担当者にとっても，こうした MHESD の円滑な利活用を普及させることが重要な目的となり，またそのための計算機資源が必要となる。従来，MHESD は各分野のそれぞれの研究者が独自に取得し，個人所有のデータストレージや外部サービスで保管・共有をすることが一般的であった。しかしながら，データ

がより高精細で大規模になるに従い，そうした個別の対応では非効率さが際立ってきている。そこで，統合的かつ学術利用に特化したシステムである mdx を活用することで，こうしたデータ共有や解析手法の開発を，より効率的かつ学際的に進めることが可能となる。とくに mdx は，仮想化クラスタと S3 互換オブジェクトストレージ，学術フェデレーション（学認）認証，広帯域 SINET 連携を備え，大容量 MHESD の収容と配信，および学術コミュニティ内での安全な共有に適した基盤を有しており，本研究の遂行に不可欠な情報基盤である。

4. 前年度までに得られた研究成果の概要

本研究は，2022 年度（jh221008）および 2023 年度（jh231005）の研究課題を基盤的な準備段階として位置づけており，これらの先行課題を通じて，MHESD の統合的活用に向けた前提となる多様なデータの取得・整理と，アーカイブ・解析手法の検討を進めてきた。

まず，地球科学分野におけるデータ取得として，無人航空機（UAS）を用いた可視光（RGB）空撮画像，マルチスペクトル画像，および地上ベースまたは UAV ベースのレーザ測量（LIDAR）による 3 次元点群データを，国内外の複数の調査地で取得した。具体的には，北海道厚真の地震性斜面崩壊地，関東近郊の海食崖，トルコ・カイセリ周辺の地形・歴史遺跡地区，バングラデシュのティスタ川流域などにおいて，地形変化や土砂移動，地表構造の変遷などを対象とした MHESD を蓄積した。これらのデータは数十 GB～数 TB の容量に及び，個別のストレージで保管されてきたが，異なる時期・センサ・解像度のデータを統合的に管理する必要性が顕在化していた。

一方，歴史考古学・博物館学分野への展開として，博物館における収蔵品（地形景観模型，考古学資料等）の写真測量による 3D モデル化を進め，

地球科学分野で取得された MHESD と博物館・人文系資料を共通のデータ形式（点群・メッシュ・テクスチャ）で取り扱う可能性を補足的に検証した。これは、MHESD が自然科学のみならず人文科学的研究にも横断的に応用できることを示す試みであり、超学際的な利活用の可能性を示唆するものであった。

また、データ解析手法の検討として、UAS や LIDAR データからの SfM 多視点ステレオ写真測量および点群フィルタリングを通じた数値表層モデル（DSM）・数値地形モデル（DTM）・オルソ画像の生成ワークフローを整理した。あわせて、ScanX, Pix4Dcloud, OpenTopography 等の商用・オープンサービスを参照しながら、類似のオンライン解析機能を mdx 上で実装する場合のコスト・実現可能性を比較検討した。その結果、高度な商用解析ツールの完全な再実装は現実的ではなく、フリー／オープンソースのソフトウェアアルゴリズム（たとえば CloudCompare 等）を基盤としたオンライン基礎解析ツールセットの整備が現実的かつ有用であるとの結論に至った。

これら先行課題でのデータ取得・整理を通じて、現地調査による 1 次データ（点群、写真、分光画像）から、SfM 処理を経たオルソモザイク・DSM・DTM、さらには段彩図・陰影起伏図・等高線図・地形量（傾斜・曲率・粗度等）の解析プロダクトに至るまで、数段階のデータプロダクトを体系的に蓄積した。これらは合計で十数 TB 規模に達しており、現状ではローカルストレージや個別の外部サービスに分散して保管されている。本基盤の構築は、これらの分散資産を mdx 上に集約し、再利用可能な共通資源として整理することにも寄与するものである。

これらの先行課題における検討と知見をふまえ、2025 年度（本課題、jh251015）からは「多次元高精細地表情報（MHESD）の地球表層科学的基盤構築」と課題名を改め、データ収集・整理段階から、より具体的で恒常的な情報基盤の構築フェー

ズへと展開している。とくに、研究者がユーザ自身で計測した MHESD を安全に保管・共有・公開し、他分野・他機関との学際的・国際的連携を支える持続的な情報基盤を、mdx を基盤として段階的に整備することが本課題の主眼となっている。

5. 今年度の研究成果の詳細

本年度は研究計画のステップ I（データ保管・共有システム構築）に集中的に取り組み、mdx を基盤とする MHESD のための統合的な情報基盤の主要な構成要素を実装するとともに、それらを連携させるためのワークフロー設計を進めた。具体的な実装と検討の内容は以下の通りである。

(1) Nextcloud によるユーザインタフェースの構築

mdx の仮想化クラスタ上に Nextcloud インスタンスをデプロイし、ユーザがウェブブラウザを通じて MHESD をアップロード・閲覧・共有・ダウンロードできるインタフェースを開発した。バックエンドのストレージとしては mdx の S3 互換オブジェクトストレージを利用し、Nextcloud の外部ストレージ機能を介して接続する構成とした。これにより、UAS による RGB 空撮画像（GeoTIFF 形式のオルソモザイク）、マルチスペクトル画像、DSM/DTM 等のラスタデータ（GeoTIFF）、LAS/LAZ 形式の 3D 点群データ、PLY/OBJ 形式のメッシュデータ、および各種 2 次プロダクト（等高線、傾斜・曲率図、断面図等）を、フォルダ階層に基づいて段階的に保管できる構成を実現している。

アクセス制御については、個人ストレージ・プロジェクト共有・公開リンクといった粒度の異なる共有モードを設けることで、研究フェーズに応じた柔軟な運用を可能にした。とくに、未公開の研究データは限定的な共有とする一方で、論文化を経た公開段階のデータについては広く外部に公開できる設計とし、研究データのライフサイクル全体に対応する基盤としての位置づけを明確にし

ている。

(2) archivematica によるメタデータ管理基盤の整備

デジタルプリザベーション基盤として広く利用されている archivematica を mdx 上に導入し、MHESD に対する標準準拠のメタデータ記述・長期保存ワークフローの枠組みを構築した。ここで、PREMIS (Preservation Metadata Implementation Strategies) に基づく保存メタデータ、Dublin Core を用いた記述メタデータ、および地理空間情報の記述に必要な ISO 19115-1:2014 準拠のメタデータを取り扱うことを想定している。Nextcloud から投入されたデータに対しては、AIPs (Archival Information Packages) 化を経て長期保存し、必要に応じて DIPs (Dissemination Information Packages) として公開・配信する経路を整備しつつある。

さらに、archivematica の出力を AtoM (Access to Memory) と連携させることで、メタデータカタログとして外部から検索可能にする構成についても並行して検討している。これにより、本基盤に蓄積された MHESD が、他の研究データ基盤からも横断的に発見・参照可能となることをめざしている。

(3) AIによる自然言語からのメタデータ抽出試験

archivematica に投入するメタデータ XML を効率的に整備するため、大規模言語モデル (LLM) を活用し、研究者が記述する自然言語の調査記録、現地調査メモ、および報告文から、ISO 19115-1:2014 準拠の地理情報メタデータ要素を抽出する試験を実施した。具体的には、識別情報 (タイトル、要約、作成日、作成者)、範囲 (地理的範囲、時間的範囲、標高範囲)、品質情報 (データ精度、解像度、処理履歴)、参照系 (測地系、座標系、投影法)、流通情報 (フォーマット、配布制限、ライセンス) といった要素について、LLM による抽出結果と人手による正解との一致度を評価した。

試験の結果、一定の精度で自動抽出が可能であ

ることを確認した一方で、専門用語 (地形学・森林科学独特の語彙) や、「〇〇の上流側」といった曖昧な空間記述については補完が必要であることがわかった。これを踏まえ、対話型でユーザに不足項目を問い返しながらメタデータを段階的に整理・補完するダイアログ型インタフェースの開発を進めている。これにより、Nextcloud からアップロードされた MHESD の意味づけをユーザ自身が行いやすくし、archivematica へとシームレスに接続するワークフローを実現することをめざす。

(4) メタデータプロファイル設計の検討

中間評価でのコメントを踏まえ、「現実解」となるメタデータスキーマの整備に着手した。具体的には、ISO 19115-1 の共通コア (識別・範囲・品質・参照系・流通) を基盤として、これに MHESD 固有の拡張項目を加えたプロファイルを文書化することとした。MHESD 固有の拡張としては、センサ仕様 (機材名、搭載センサ種別、分光特性等)、飛行ログ (飛行高度、速度、撮影間隔、天候等)、点群処理履歴 (フィルタリング手法、分類アルゴリズム、地表抽出手法等)、SfM 処理パラメータ (マッチング手法、バンドル調整条件、再投影誤差等) などが含まれる。

これらの拡張項目について、Nextcloud のカスタムプロパティ機能および archivematica のメタデータ取り込みフィールドとの対応関係を整理する作業を開始した。これは最終的に、ユーザが Nextcloud 上で記述した情報がそのまま archivematica 側でも再利用される自動連携の基礎となることが期待される。

(5) 外部基盤との連携可能性の検討

本基盤を孤立したシステムとせず、国内外の既存の研究データ基盤と相互運用可能な存在として位置づけるための連携可能性について検討した。国内では、国立情報学研究所 (NII) の GakuNin RDM との連携を想定し、研究データのライフサイクル全体 (取得・整理・保管・公開) における役割分担をとり、必要なメタデータ互換性の維持

に関する情報を NII と共有しながらシステム開発を進めている。具体的には、GakuNin RDM にログインしたユーザが、連携している本基盤 (Nextcloud+archivematica) を背後に置きながら違和感なくデータ管理を行える構成を構想している。

さらに、G 空間情報センター (公的・準公的な空間情報の集積・配信ハブ) や、東大 CSIS 研究用空間データ基盤 (JoRAS) との連携シナリオの検証も開始した。国外では、米国 NSF が支援する地形・地球物理データの公開基盤 OpenTopography を参照モデルとし、カタログ同期 (タイトル, 要約, 主題語, 空間範囲, ライセンス等の項目) と軽量リンク (DIP 参照 URL) の試作を計画している。これらの相互運用は、MHESD の発見可能性を高めるとともに、本基盤を国内外のデータエコシステムにおけるハブとして位置づけるうえで重要である。

(6) NII 研究データエコシステム構築事業との連携

NII による「AI 等の活用を推進する研究データエコシステム構築事業」のユースケース創出課題 (2024~2025 年度) と連携し、本 JHPCN 課題と相互補完的にシステム構築を加速している。NII の事業はより広範な研究分野におけるデータエコシステムの構築を視野に入れており、本 JHPCN 課題で開発している MHESD 基盤は、地球表層科学領域における先進的なユースケースを提供する位置づけにある。具体的には、NII 事業からは GakuNin RDM との連携技術や標準化の知見を得る一方で、本 JHPCN 課題からは大容量・多次元データの実運用に基づくフィードバックを提供している。両者の相補的な関係は、研究データ基盤を分野固有の要件と汎用的な仕組みの両面から発展させるうえで効果的である。

(7) 計算機利用上の工夫

mdx の仮想化クラスタを用いて Nextcloud および archivematica を並行運用するにあたっては、両者がアクセスする共通のオブジェクトストレージ

層を mdx の S3 互換ストレージに据え、各サービスのデータベースとアプリケーションサーバ層を分離する構成を採用した。これにより、大容量 3D データ (数十 GB~数 TB 単位) に対しても安定した収容と配信が可能となるとともに、サービスごとの独立したスケールアウトが可能となる。

アクセス制御については、将来的には学術フェデレーション (学認) 認証と連携することを想定し、ユーザ管理の独立性を保ちつつ統一的な入退管理が可能となる構成を志向している。広帯域 SINET を介した他研究機関とのデータ共有や、遠隔地からの大容量データアップロードについても、mdx の計算機資源と通信基盤を活かすことで実用的な性能が見込まれる。

(8) 全体システム構成と動作確認

以上の(1)~(7)の構成要素を総合すると、本基盤は、フロントエンド (ユーザ向けインタフェース) としての Nextcloud, メタデータ管理・長期保存基盤としての archivematica, 自然言語からのメタデータ抽出を担う AI/LLM コンポーネント, 国内外との連携を担う GakuNin RDM 等の外部基盤連携層, およびそれらをホストする mdx の仮想化クラスタ・S3 互換オブジェクトストレージという、複数の層から構成されることになる。各層の独立性を保ちつつ、API レベルでの自動連携を志向する設計を採用しており、将来的な機能追加や置換にも柔軟に対応できる構成としている。

動作確認の段階としては、本年度はサンプルとなる MHESD (小規模な UAS オルソモザイクおよび LIDAR 点群データ) を用いて、Nextcloud へのアップロード, archivematica でのメタデータ生成, AI による自然言語からのメタデータ抽出補助, AIPs 生成までの一連の流れを単体で確認するに至った。一方で、これらを連結した完全自動化の End-to-End ワークフローについては設計段階に留まっており、今後のステップ II において統合動作試験と利用者評価を実施することを計画している。

6. 進捗状況の自己評価と今後の展望

本節では、まず申請時に設定した今年度の計画に対する進捗を構成要素別に自己評価し、次に中間評価における審査委員会からのご指摘への対応方針を整理し、最後に 2026 年度以降のステップ II・III の計画について述べる。

[自己評価]

今年度（2025 年度）に設定した計画の中核は、研究計画ステップ I（データ保管・共有システム構築）の実装であった。これに対し、主要構成要素ごとの進捗を以下のように自己評価する。Nextcloud インタフェースの構築については、mdx の S3 互換ストレージとの接続を含めて基本的なデプロイ・動作確認を完了している。archivematica の導入とメタデータ管理枠組みの構築については、基本動作確認とワークフロー設計を完了しており、AIPs/DIPs の試作を残す段階である。AI を用いたメタデータ抽出試験は基礎試験を完了し、対話型インタフェースの開発に着手している。メタデータプロファイル設計（ISO 19115-1 ベース+MHESD 拡張）については検討中で、統合ワークフロー（Nextcloud → AI → archivematica → GakuNin RDM）の自動化は設計レベルで進行中である。代表的データセットによる試験運用と利用者評価については未着手である。

これらを総合して、今年度の計画全体の達成度はおおそ目標の半分以下と自己評価する。中核的な構成要素の単体実装は計画通りに進んだ一方で、それらを統合した一気通貫の自動化ワークフローと、代表的データセットを用いた試験運用・利用者評価については、2026 年度以降に実施検討するステップ II で具体化することが妥当と判断している。

[中間評価コメントへの対応]

中間評価において、審査委員より以下の点についてご指摘をいただいた。すなわち、(a) 当初計

画に示されたステップ I～III との対応関係、および統合ワークフローの実装状況をより明確にすること、(b) 代表的データセットによる試験運用と利用者評価の具体化、(c) G 空間情報センター等との具体的連携シナリオの提示、(d) 国際ネットワークを活かした共同利用体制の明示、(e) プロトタイプ公開や論文文化など外部発信の加速、の 5 点である。

(a) については、本最終報告書において、今年度の各取り組み（5 章の(1)～(7)）を、研究計画ステップ I に対応する個別の構成要素として明示的に位置づけ整理した。さらに今後、各構成要素の関係を 1 枚のシステム図として文書化し、ウェブ上でも公開できると良いと考えている。あわせて、統合ワークフローの自動化スクリプトを、2026 年度の早期に整備する計画である。

(b) については、2026 年度以降のステップ II（試験運用と機能追加）の主要タスクとして組み込み、地形学分野（北海道厚真の地すべり、海食崖モニタリング等）および森林科学分野（北海道大学研究林等）の代表的データセットを対象とした試験運用と利用者評価を計画する。利用者評価には、代表者が指導する大学院生・若手研究者、および国内外の共同研究者を被験者として含める予定である。

(c) については、G 空間情報センターおよび JoRAS との間で、カタログ同期（タイトル・要約・主題語・空間範囲・ライセンス）と軽量リンク（DIP 参照 URL）の試作を行い、国内ユーザがこれら既存基盤を経由して本基盤上の MHESD を参照できる経路の確立をめざす。

(d) については、代表者がアジアオセアニア地域における地形学・GIS 分野の研究者ネットワーク（バングラデシュ、ニュージーランド、中国等）を有していることを踏まえ、これら地域の研究者・大学院生による本基盤の試験利用を、国際共同研究プロジェクト等を通じて段階的に進めることを計画する。また、OpenTopography（米国）等

の海外データ基盤プロジェクトとの相互運用検証も視野に入れている。

(e) については、本基盤の設計・実装に関する論文化（地球科学分野の国際学術誌，および情報科学分野の関連誌への投稿），国内外の学会発表（JpGU-AGU，日本地形学連合，日本地理学会等）を計画している。プロトタイプの一部については，GitHub 等でのオープンソース公開も検討している。

[今後の展望]

2026 年度以降は，本研究の中核を以下のように展開する。ステップ II（情報基盤の試験運用と機能追加の計画）では，地球表層科学（地形学，森林科学等）の研究者・学生・実務担当者を対象に，ステップ I で構築した情報基盤の試験運用を実施する。具体的には，km・m・cm と空間スケールの異なる景観構成要素（地形・植生）の 3D データを対象とし，シンプルで扱いやすく，かつ大容量でも高速処理が可能な，各分野で好まれる解像度・精度を模索しながらシステムを最適化する。これにより，情報技術に詳しくない当該分野の学生や若手研究者にあっても，自ら取得した高精細 3D データを容易に管理・活用できるプラットフォームへと洗練させる。

あわせて，保管，共有，分析，可視化といった機能の一気通貫したプラットフォームを目標に追加機能を実装してゆく。さらに，僻地の現地調査における通信網（衛星通信，4G/5G 携帯回線等）を活用し，現場で取得した MHESD を即座に基盤上にアップロード・可視化・他者と共有するモバイルワークフローの試作も検討する。具体的な利活用シナリオとして，過去の計測データに基づく活動的な地すべりの 3D タイムラインと現地調査による新規 3D データをリアルタイムに照合し，変化箇所を即座に抽出・可視化することや，下層植生を含む森林構造を高頻度に計測してビデオの早回しのように 3D で可視化したものを半自動的にウェブ公開すること，分野横断的な利活用とし

て地すべり移動体上の樹木の形状変化と土塊変動・植生フェノロジーを連関的に可視化することなどを想定している。

さらに先のステップ III（情報基盤の一般公開と展開）では，少数の限定利用から，より多くの利用者に構築した情報基盤を開放し，本格的運用を開始する。メタデータ設計について，ISO 19115-1 基盤+MHESD 固有プロファイルの策定を完了し，Nextcloud/archivematica との対応関係を GakuNin RDM の上で閲覧可能とする。代表的データセットによる試験運用と利用者評価を経て，G 空間情報センター・JoRAS・OpenTopography 等の国内外の空間情報基盤との連携を本格化させる。国際連携としては，アジアオセアニア地域における研究者間での利用や，欧米における同様な研究拠点との協働を本格化し，本情報基盤の国際的・学際的な展開をめざす。

さらに，集約したデータそのものや解析済みのデータセットについて，3D プリントといった実物への出力や，仮想現実（VR）・メタバース等の仮想空間にもシームレスに接続できる出力形式の整備も並行して進める。これにより，地球表層科学領域における先進的な研究データ基盤として，国内外で確固とした位置づけを確立することをめざす。

最終的に本研究の成果は，論文化（国際学術誌への投稿），学会発表（国内外），およびオープンソースとしてのプロトタイプ公開を通じて広く外部に発信し，地球表層科学のみならず人文学・教育を含む超学際的な MHESD 利活用の促進に資するものとする。

※7.研究業績はウェブ入力です