

令和6年度 成果報告書

基本情報(公開)

事業名 プログラム名		戦略的イノベーション創造プログラム(SIP) 「統合型ヘルスケアシステムの構築」
研究開発課題名		E-1 大容量医療データの高速処理・高効率管理・高次解析基盤技術の開発
研究開発 担当者* 1	機関名	国立大学法人東京大学
	所属	生産技術研究所
	役職	教授
	氏名	合田 和生
実施期間*2		令和6年4月1日～令和7年3月31日

*1 委託研究開発契約書に定義

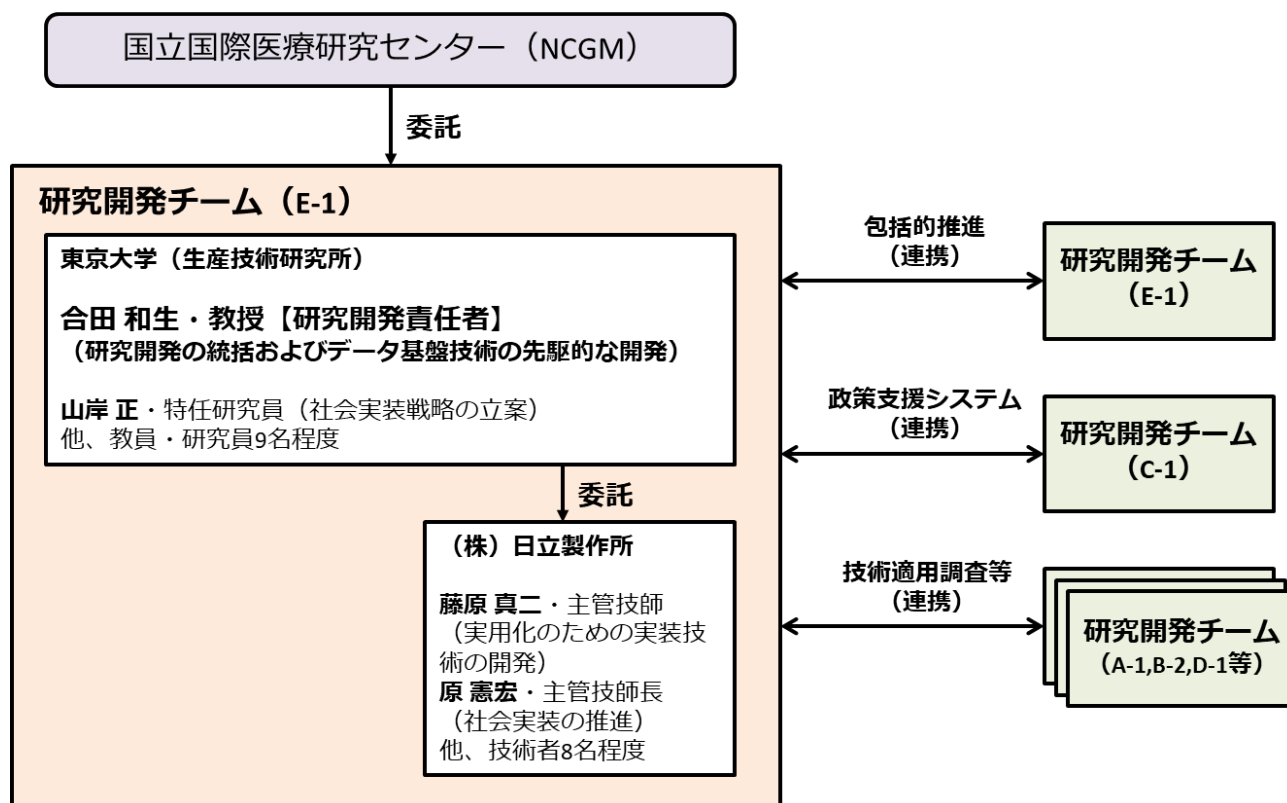
*2 年度の契約に基づき、本委託研究開発を行った期間又は中止までの期間

1. 研究開発テーマ概要

1.1 研究開発内容

本研究開発テーマでは、我が国のヘルスケアシステムで生成される国家規模の膨大な医療データを集約し、機動的にデジタルサービスを生成してヘルスケアシステムに提供可能とするデータ基盤技術確立することを目指す。SIP が構想する「医療デジタルツイン」の実現に向けた基礎を確保するものであり、具体的には、(1)一連の解析プロセスを飛躍的に高速化する技術、(2)複雑度の高い解析処理を飛躍的に高速化もしくは高精度化する技術、(3)システム管理者の意思決定を支援することによりその作業を飛躍的に高効率化する技術確立することを目指す。また、当該技術をもとにソフトウェアを試作し、実験室環境に於いて当該技術の有効性を検証するほか、他の研究開発チームと連携の下でヘルスケアフィールドに於ける実証試験を行い応用の観点からの有効性を検証する。更に産業展開を想定した実装技術確立する。

1.2 研究開発実施体制



1.3 研究推進のスケジュール

項目	令和5年度	令和6年度	令和7年度	令和8年度	令和9年度
(サブテーマ1) 大規模医療データを対象とする高速解析技術の開発	GPU/DPU活用 方式検討	PMEM活用方式検討・ ソフトウェア設計	ソフトウェア試作・ ベンチマーク試験 (規模限定)	ネットワーク化 複数サーバ方式検討	ソフトウェア高度化・ ベンチマーク試験 (大規模)
(サブテーマ2) 大規模医療データを対象とする高次解析技術の開発	近似問合せ 方式検討	類似検索方式検討・ ソフトウェア設計	ソフトウェア試作・ ベンチマーク試験 (規模限定)	近似処理と 類似検索の融合 方式検討	ソフトウェア高度化・ ベンチマーク試験 (大規模)
(サブテーマ3) 大規模医療データを対象とする高効率管理技術の開発	データ編成 自動化方式検討	システム構成 自動化方式検討・ ソフトウェア設計	ソフトウェア試作・ ベンチマーク試験 (規模限定)	クラウド環境への 拡張方式検討	ソフトウェア高度化・ ベンチマーク試験 (大規模)
(サブテーマ4) プロジェクトの総合的推進					

2. 本年度の成果・進捗の概要

本年度は、サブテーマ(1)大規模医療データを対象とする高速解析技術の開発については、前年度に実施した GPU ならびに DPU 等の異種プロセッサ技術を活用する超高並列実行手法の検討に加えて、広帯域フラッシュメモリ等を活用したデータ配置手法について検討を行い、潜在的な有効性を確認した他、ソフトウェアの構成法を検討し設計を纏めた。また、大規模医療データを模した合成データを用いて、データベース編成を行う試験的な並列実装を試作し、各種の実装技法の効果を実験的に確認した。サブテーマ(2)大規模医療データを対象とする高次解析技術の開発については、類似検索に適合した問合せ最適化手法について検討を行い、潜在的な有効性を確認した他、前年度の研究開発成果に基づき、ソフトウェアの構成法を検討し設計を纏めた。また、索引にシノプシス(概要情報)を埋込む近似問合せ実行方式を試作実装し予備試験を実施し、その効果を実験的に確認した。サブテーマ(3)大規模医療データを対象とする高効率管理技術の開発については、前年度に行ったデータベースのアクセスログ等に加えて、ストレージシステムへの入出力ログ等を解析することにより効率的なシステム構成を推薦するシステム構成自動化手法を検討し、潜在的な有効性を確認した他、ソフトウェアの構成法を検討し設計を纏めた。また、模擬データを用いてデータベース管理技能の被験者試験を実施し、性能チューニングの余地と重要性を確認した。サブテーマ(4)プロジェクトの統括と総合的推進に於いては、ステアリング会議を 1 回、技術検討会を 13 回、個別の打ち合わせ(技術、対外発表等)を26回行った。また、事業・制度・社会的受容性の視点から、研究開発成果の社会展開を効果的に推進するために、実証活動に向けて健康医療分野の事業者等との対話イベントを実施した。更に

【経理様式 1 別添】【R6】

人材の視点から、優秀な若手の研究員・技術者を積極的に配置した。

3. 成果物の公表

3.1 論文など(原著論文、学位論文、プロシーディングス、総説、解説、速報など)

論文数(総数)	(内国際誌)	(内国内誌)
2	2	0

1. Hiroyuki Yamada, Masaru Kitsuregawa, Kazuo Goda. LakeHarbor: Making Structures First-Class Citizens in Data Lakes. IEEE International Conference on Data Engineering (ICDE2024), 5583-5592, DOI:10.1109/ICDE60146.2024.00446. (※1)
2. Hirotaka Yoshioka, Yuto Hayamizu, Kazuo Goda, Masaru Kitsuregawa. Synchronization Options and Their Performance Benefits for Spatial Index on Persistent Memory. Information Integration and Web Intelligence (iiWAS2024). (※1)

(※1)情報学分野では、査読を伴う論文を編纂した会議録が出版される学会発表は、原著論文と同等に扱われていることから、本項目に分類している。

3.2 学会発表など((国内・国際)学会口頭発表・ポスター発表、招待講演で成果を公表したもの)

1. QRAM 上のデータベース格納に関する一検討, 御手洗陽紀, 合田和生, xSIG2024, 2024/8/7, 国内, 口頭.
2. 複数 SSD を有する計算機上における GPU 直接 IO を用いたハッシュ結合実行方式の実験的考察, 小沢健史, 三浦優也, 合田和生, IPSJ 第 179 回 DBS 研究発表会, 2024/9/11, 国内, 口頭.
3. Optimizing Nested Loops for Big Data Analytics, Kazuo Goda, Korea-Japan (Japan-Korea) Database Workshop (KJDB 2024), 2024/11/16, 国外, 口頭.
4. 大規模医療データ分析における列指向の非順序型データベースエンジンの性能評価, 藤本太郎, 飯田敏成, 木村耕治, 清水晃, 藤原真二, 第 44 回医療情報連合大会, 2024/11/22, 国内, 口頭.
5. GPU 直接 IO を用いた圧縮データに対するデータベース問合せ処理方式の検討, 小沢健史, 合田和生, DEIM2025, 2025/2/27, 国内, 口頭.
6. 結合を含むクエリに対するシノプシス埋込みによる近似問合せ処理の評価, 高田実佳, 合田和生, DEIM2025, 2025/2/27, 国内, 口頭.
7. 量子ランダムアクセスメモリにおける関係データベースの編成, 御手洗陽紀, 合田和生, DEIM2025, 2025/2/28, 国内, 口頭.
8. 大規模近似最近傍探索を対象とする入出力特性の測定と性能評価, 呉果行, 小沢健史, 合田和

【経理様式 1 別添】【R6】

生, DEIM2025, 2025/2/28, 国内, 口頭.

9. Performance Analysis of Different IO Methods between GPU Memory and Storage, Tarun SREEPADA, 小沢健史, ラゲ・ウダイキラン, 合田和生, DEIM2025, 2025/2/28, 国内, 口頭.
10. anagodb: 挑戦的なデータベースコア技術研究のための実践的コードベース, 早水悠登, 川道亮治, 小沢健史, 合田和生, DEIM2025, 2025/2/28, 国内, 口頭.
11. 多次元ソートのための整数値変換手法の提案, 高尾大樹, 黒川能毅, 西川記史, 茂木和彦, 藤原真二, DEIM2025, 2025/3/1, 国内, 口頭.

3.3 診療ガイドライン、省令、基準、日本薬局方、添付文書改訂、国の技術文書(通知)等への反映

なし

3.4 研修プログラム、カリキュラム、シラバス、教材、e-learning 等の公表

なし

3.5「国民との科学・技術対話」に対する取り組み

1. データストレージの歴史, 合田和生, Database Society of Japan (DBSJ), 2024/8/7, 国内.
2. データ管理ソフトウェア技術の動向や課題:クラウドや生成 AI, 合田和生, International Disk Drive Equipment and Materials Association (IDEMA JAPAN), 2024/11/29. 国内.
3. ビッグデータ価値協創プラットフォーム工学社会連携研究部門の活動, 合田和生, ビッグデータ価値協創プラットフォーム工学社会連携研究部門シンポジウム「価値協創が拓く データ活用・生成 AI の未来」, 2024/12/13, 国内.
4. 大容量医療データの高速処理・高効率管理・高次解析基盤技術の開発, 合田和生, SIP 第 3 期「統合型ヘルスケアシステムの構築」2024 年度公開シンポジウム, 2025/2/28, 国内.
5. 「人手不足」の社会課題に生成 AI×データ管理で挑む!, 西川記史, DEIM2025 ランチョン・セミナー, 2025/3/3, 国内.

3.6 その他

1. 東大生研と日立が協創で挑む、生成 AI 時代の社会課題解決【第 1 回】最新のデータ基盤技術を基に、フロントラインワーカーを支える, 日立 Digital Evolution Headline(https://deh.hitachi.co.jp/_ct/17749077), 2025/3/5.
2. 東大生研と日立が協創で挑む、生成 AI 時代の社会課題解決【第 2 回】価値創出のためのデータ基盤、人財に求められるもの, 日立 Digital Evolution Headline(https://deh.hitachi.co.jp/_ct/17749078), 2025/3/6.

以上

【経理様式 1 別添】【R6】