

実証を通じた横断検索システム機能の高度化及び Web システムの開発

研究分担者 荻島 創一 東北大学東北メディカル・メガバンク機構医療情報 ICT 部門 教授

研究要旨

次世代バイオデータ基盤の構築に向けて、2019 年度に実施した調査研究で取りまとめた「コホート・バイオバンクの産業利活用促進策」の具体化のために、コホート・バイオバンクの産業利用促進のための調査研究を踏まえて 2020 研究開発したコホート横断検索システムの時系列検索機能を拡充し、高度化した。また、コホート研究により収集されるデータの項目のメタデータを標準化し、各コホート研究のデータ項目のメタデータを公開するコホート横断検索システムカタログの研究開発を行った。さらに、企業向け相談窓口の整備として Web システムの開発を行い、データの利活用を促進するため、データの説明書の研究開発を行った。

A. 研究目的

バイオ戦略では、「医療とヘルスケアが連携した末永く社会参加できる社会」の実現を目指し、「バイオとデジタルの融合のためのデータ基盤の整備」、なかでも「大規模統合コホート・バイオバンクの構築」の取組みが必要とされ、「健常人コホート等の実施主体が連携し、データを統合・強化する大規模健常人コホート・バイオバンクの構築」、「健常人コホート・バイオバンクについては、多様な分野において、科学的エビデンスに基づいたサービスを提供できる環境を整備」することとされた。

このような政策的位置付けの中で、東北大学東北メディカル・メガバンク機構は、2019 年度の官民研究開発投資拡大プログラム(PRISM) バイオ技術領域で実施した調査研究において、コホート・バイオバンクの産業利用促進策を取りまとめ、その具体化のために、2020 年度はコホート横断検索システムを研究開発した。

2021 年度は、2020 年度研究開発したコホート横断検索システムの時系列検索機能を拡充し、高度

化する。また、コホート研究により収集されるデータの項目のメタデータを標準化し、各コホート研究のデータ項目のメタデータを公開するコホート横断検索システムカタログの研究開発を行う。さらに、企業向け相談窓口の整備として Web システムの開発を行い、データの利活用を促進するため、データの説明書の研究開発を行う。

B. 研究方法

1. コホート横断検索システムの高度化の研究開発

コホート・バイオバンクの産業利用促進策として、コホート横断検索システムを高度化するため、株式会社とせ研究所による「コホート・バイオバンクの産業利用促進のための調査研究」を通じて、コホート横断検索システムの高度化に対する産業界のニーズを把握する。

把握したニーズに基づき、コホート横断検索システムの高度化を行う。

2. コホート横断検索システムカタログの研究開発
コホート研究により収集されるデータの項目のメタデータを標準化し、各コホート研究のデータ項目のメタデータを公開するコホート横断検索システムカタログの研究開発を行う。

連携する官民研究開発投資拡大プログラム (PRISM) バイオ技術領域「糖尿病個別化予防を加速するマイクロバイーム解析 AI の開発」、「認知症に関与するマイクロバイーム・バイオマーカー解析」から協力を得て、各コホート研究で収集したデータの項目のメタデータを標準化する。

そのうえで、標準化された各コホート研究のデータの項目のメタデータを公開するコホート横断検索システムカタログの研究開発を行う。

3. 企業向け相談窓口の Web システムの開発とデータ説明書の研究開発

企業向け相談窓口の整備として Web システムの開発を行い、データの利活用を促進するため、データの説明書の研究開発を行う。企業向け相談窓口の Web システムは、Web 上で伴走支援を行うシステム(企業向けWeb相談申請システムプラットフォーム)及び企業向け情報提供ウェブサイトの研究開発を行う。データの説明書は、東北メディカル・メガバンク計画のコホートの収集データについて、企業向けの説明書の研究開発を行う。

(倫理面への配慮)

本研究はヒトゲノム・遺伝子解析、臨床研究、ヒトを対象とする医学系研究、動物実験等の実施はない。したがって倫理面の問題はないと判断した。

C. 研究結果

1. コホート横断検索システムの高度化の研究開発

コホート横断検索システムに対する産業界のニーズを把握し、コホート横断検索システムの高度化を行った。

(1) コホート横断検索システムの高度化に対する産業界のニーズ

株式会社ちとせ研究所による「コホート・バイオバンクの産業利用促進のための調査研究」を通じて、コホート横断検索システムに対する産業界のニーズを把握した。

関心のあるデータ項目としては、睡眠やストレス・うつ、運動、飲酒、喫煙などの生活習慣、食習慣、罹患歴、検体検査値、マイクロバイーム情報、メタボローム情報、体重、BMI、体脂肪率などであり、これらのデータについて時系列の検索にニーズがあることを受けて、時系列検索機能を拡充することとした。

また、コホートの収集データの項目のカタログも閲覧できるようにしたいというニーズがあることがわかった。これは「コホート横断検索システムカタログの研究開発」で取扱う。

(2) コホート横断検索システムの高度化の研究開発

把握したニーズに基づき、コホート横断検索システムの高度化の設計を行った。

コホート横断検索システムは、横断検索を実行するハブと、各コホート研究のデータベースのノードにより構成される。ユーザがフロントエンドで検索条件を組み立て、検索クエリをハブ API に渡すと、ハブ API は各ノード API に横断的に検索リクエストを送信する。各コホートのノード API は検索リクエストを処理し、ハブ API に検索結果を返す。ハブ API は各ノード API から返された検索結果をとりまとめて、ユーザに対してフロントエンドで検索結果を表示する。コホート横断検索システム(図 1,2)は、2020 年度の研究開発では、同一イベントのデータについて 2 時点の時系列検索まで可能としていた。産業界からのニーズを受けて、2021 年度の高度化では、2 時点から 4 時点までの同一または異なるイベントのデータの項目を指定し、そのイベントの日付の順序と日数の間隔を条件にした時系列検索の機能を開発した(図 3)。



図1 コホート横断検索システム



図2 ログイン後のトップページ

時系列検索機能

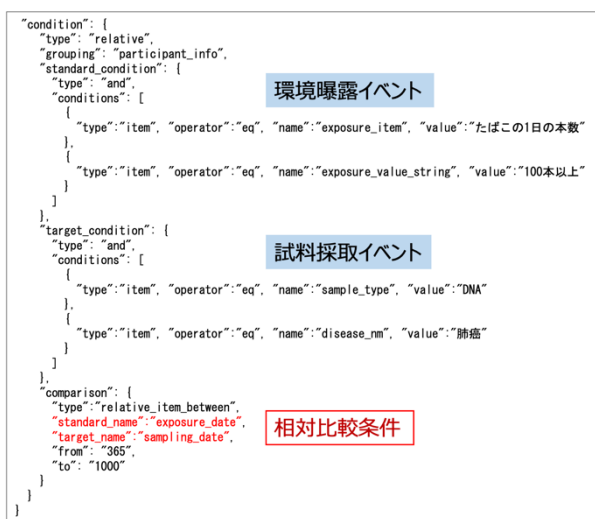


図3 時系列検索機能の拡充

2. コホート横断検索システムカタログの研究開発

コホート研究により収集されるデータの項目のメタデータを標準化し、各コホート研究のデータ項目のメタデータを公開するコホート横断検索システムカタログの研究開発を行った。

(1) 収集データ項目のメタデータの標準化

連携する官民研究開発投資拡大プログラム (PRISM) バイオ技術領域「糖尿病個別化予防を加速するマイクロバイオーーム解析 AI の開発」、「認知症に関与するマイクロバイオーーム・バイオマーカー解析」から協力を得て、各コホート研究で収集したデータの項目のメタデータを標準化した。

コホート横断検索システムに対する産業界のニーズによると、関心のあるデータ項目としては、睡眠やストレス・うつ、運動、飲酒、喫煙などの生活習慣、食習慣、罹患歴、検体検査値、マイクロバイオーーム情報、メタボローム情報、体重、BMI、体脂肪率などであったため、生活習慣・食習慣は環境曝露イベント、罹患歴は協力者の既往症・併存症、マイクロバイオーーム情報、メタボローム情報は解析情報、体重、BMI、体脂肪率などについて重点的にメタデータを標準化した。

収集データ項目のメタデータとしては、項目の名称、単位、データ型、項目のカテゴリの4種類とした。このうち、項目のカテゴリは、データの情報源を示すものである。単位は、検査値など単位をもつものについて、項目のもつ値の意味を示す。データ型は2値分布、離散分布など、項目がどのような分布をとるものなのかを示す。また、収集データの階層分類は、項目を一意的コードでコーディングし、項目とそのカテゴリの階層関係を分類するものであり、項目の意味的関係を提供し、コホートで収集された多種多様なデータのなかから、必要なデータを発見することを可能とするものである。

(2) コホート横断検索システムカタログの研究開発

標準化された各コホート研究のデータの項目のメタデータについて、これを公開するコホート横断検索システムカタログの研究開発を行った。

コホート横断検索システムカタログにおいて、メタデータを公開するため、

- JSON Web Token (JWT) による認証
- 収集データ項目の階層分類
- 収集データの標準化されたメタデータ

について横断的な発見のための API を設計し、開発した。収集データの階層分類を返す API は、指定されたコホートの収集データの階層分類が JavaScript Object Notation (JSON) 形式で返す。

返される内容は、項目を識別するコード、その項目の親の階層の項目のコード、項目名、階層構造上での並び順である。

この機能により、コホートで収集された多種多様な項目について収集データ項目のすべての項目のリストと項目間の階層構造から項目を発見可能とする。収集データのメタデータは、指定されたコホートの収集データの、メタデータが JSON 形式で返す。

API を使用するメリットは、リクエストアドレスポンスおよびデータベースの構造の仕様を共有可能なことである。現在、国際的にはコホートを連携し、メガコホートを形成する流れがあり、API を使用することで、今後、国外のコホートとの連携した、コホートの収集データのデータ発見につながると考えられる。

連携する官民研究開発投資拡大プログラム (PRISM) バイオ技術領域「糖尿病個別化予防を加速するマイクロバイオーム解析 AI の開発」、「認知症に関与するマイクロバイオーム・バイオマーカー解析」から協力を得て、各コホート研究で収集したデータの項目のメタデータをコホート横断検索システムカタログに収載し、公開した (図 4)。各コホートの概要説明 (図 5, 6, 7) が公開され、コホートのデータ項目の階層検索とキーワード検索によるメタデータの公開がなされた (図 8)。



図 4 コホート横断検索システムカタログ トップページ



図 5 コホート概要説明
東北メディカル・メガバンク計画概要説明

PRISM認知症プロジェクト 認知症に関与するマイクロバイオーム・バイオマーカー解析

プロジェクト代表者 山本 (前田) 万里
プロジェクト代表機関 国立研究開発法人農業・食品産業技術総合研究機構

プロジェクト概要

認知症の早期発見・予防技術開発、認知症予防食品の開発は、決定的なものがなく健康寿命延伸を阻止する大きな要因となっていた。そこで、500名以上の健康者(軽度認知障害 MCI、認知症患者を含む)のバイオマーカー(血中成分、サイトカイン80項目、アミロイドβ等)、エビゲノム、腸内マイクロバイオーム、認知機能等のデータを収集・解析する。2年間の被験者データを統合し、加齢や認知機能と関連のある心身状態や生活習慣などの探求を行う。さらに、認知症患者からのデータを比較することで、認知機能との関連因子を探索する。また、認知症患者を含む約300名の被験者の口腔内細菌叢のメタゲノム・メタボローム・プロテオーム解析により、当該疾患に特異的に関連するマイクロバイオームの特定を行う。加えて、認知症患者を含む高齢被験者の腸内細菌叢(糞便)を無菌マウスに定着させて作製したノドバイオーム(無菌動物に認知症の生物を定着させ、そこに存在するすべての生物が分かっている動物)あるいはADモデルマウスを用いた認知症患者に特異的な病理(神経新生、炎症、アミロイド沈着等)あるいは行動・認知機能に関わる腸内マイクロバイオームの同定を行い、その分メカニズムを解明する。これらを通じて、軽度認知障害/認知症発症の早期発見と重症度を高精度に評価する技術・製品・サービスを開発し、認知機能の維持に資する食品やケア、介入方法の官民の研究開発を加速化する。

本プロジェクトで収集している情報のカタログとなります。ご活用ください。

情報種別	
基本情報	30項目
検体検査	72項目
バイタル	6項目
調査票 (食事、睡眠日誌、採便時等)	3457項目
認知機能検査 (Q-ESD, Mini-Cog)	19項目
認知機能	17項目
睡眠変数	112項目
自律神経	68項目
マイクロバイオーム	34432項目
サイトカイン解析	76項目
ゲノム・オミクス情報	2項目

図 6 認知症に関与するマイクロバイオーム・バイオマーカー解析 (PRISM 認知症プロジェクト) 概要説明

NIBIOHNマイクロバイオームプロジェクト 「健常人マイクロバイオーム情報基盤の構築」ならびに「糖尿病個別化予防を加速するマイクロバイオーム解析AIの開発」

プロジェクト代表者 國澤 純
プロジェクト代表機関 医薬基盤・健康・栄養研究所

プロジェクト概要

医薬基盤・健康・栄養研究所のマイクロバイオームデータベースには、健常人ならびに糖尿病など疾患患者などを対象に、食事、睡眠、運動などの生活習慣や健康診断データ、疾患・服薬履歴などのメタデータと共に、腸内細菌叢の構成や機能に関するマイクロバイオーム・代謝物（メタボローム）、免疫因子（イムノーム）など多様なデータを収蔵統合している。さらに、収集データは、独自に開発した統合解析プラットフォームMANTAを用いて可視化、解析することが可能である。これら一連のデータやシステムを、個別化/個別化AI栄養システム、高機能食品、ヘルスケア製品の開発につなげるプラットフォームとして多くの企業に活用していただいている。

本プロジェクトで収集している情報のカタログとなります。ご活用ください。

情報種類	
基本情報	6項目
調査票（食事、対象者特性、ライフスタイル、当日排便状況）	1282項目
メディカル	363項目
フィットネス	9項目
分子情報 代謝物	228項目
分子情報 免疫因子	29項目

図 7 「健常人マイクロバイオーム情報基盤の構築」ならびに「糖尿病個別化予防を加速するマイクロバイオーム解析 AI の開発」（NIBIOHN マイクロバイオームプロジェクト）概要説明

階層より検索	キーワード検索	階層より検索	キーワード検索
- 基本情報 - 年齢 - 性別 - 健康調査情報 - 生化学的検査 - 免疫学的検査（アレルギー検査等） - 血液学的検査（血球検査・白血球分画等） - 尿検査 - 特定健康診査 - 調査票 - 調査票（生活） - 調査票（食） - タブレット調査票		- 血圧 - 随時血圧測定：1回目 測定日付時刻 20071000 - 随時血圧測定：2回目 測定日付時刻 20071001 - 随時血圧測定：1回目 収縮期血圧 20072000 - 随時血圧測定：2回目 収縮期血圧 20072001 - 随時血圧測定：平均値 収縮期血圧 20072002 - 随時血圧測定：1回目 拡張期血圧 20073000 - 随時血圧測定：2回目 拡張期血圧 20073001 - 随時血圧測定：平均値 拡張期血圧 20073002	

図 8 コホートのデータ項目の階層検索とキーワード検索

3. 企業向け相談窓口の Web システムの開発とデータ説明書の研究開発

企業向け相談窓口の Web システムとして、Web 上で伴走支援を行うシステム（企業向け Web 相談申請システム）および企業向け情報提供 Web サイトの研究開発を行った（図 9）。企業向け相談窓口機能の実証でとりまとめた、企業によるコホートのデータの利用までのワークフローについて、各プロジェクトを確実に効率的に進捗させるための進捗管理ツールを、asana を用いて開発した。企業がデータを利用できるようになるまでの工程は複雑で、企業向け相談窓口機能の実証にて報告しているように 16 もの工程からなる。asana のワークフロー管理により、それぞれの工程で企業、企業向け相

談窓口、コホートの担当者が対応すべきタスクを管理し、進捗を共有できるようにした。企業からのニーズに、データを利用できるようになるまでの手続きがわかりにくいというものがあった。今回、企業がコホートのデータを利用できるようになるまでの手続きをワークフロー化し、これをワークフロー管理ツールで進捗管理することで、複雑な手続きが整理されて、「見える化」された。また、企業とのメールや電話でのやりとりは非常に多く、このやりとりを管理するカスタマー関係管理のツールとして、pipedrive を採用し、integromat（make）を用いて、asana と連携させた。さらに、企業向けに情報を提供する Web サイトを webflow を用いて構築した。連携する官民研究開発投資拡大プログラム（PRISM）バイオ技術領域「糖尿病個別化予防を加速するマイクロバイオーム解析 AI の開発」、「認知症に与えるマイクロバイオーム・バイオマーカー解析」の企業向けの情報も含めて提供する Web サイトとして、情報を公開する。

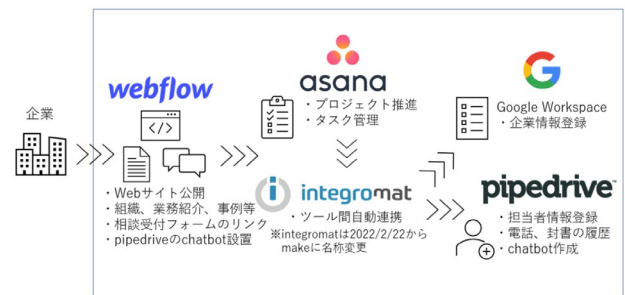


図 9 企業向け相談窓口の Web システム

データの説明書は、東北メディカル・メガバンク計画で、分譲に供されている地域住民コホート、三世代コホートの収集データについて、企業向けの説明書の研究開発を行った。コホート調査の概要を説明し、収集データとして基本情報、コホート調査による調査票情報、健診情報、検体検査情報、収集した試料のゲノム・オミックス解析によるゲノム配列情報、メタボローム解析情報、プロテオーム解析情報について説明書を作成した（図 10）。ゲノム・オミックス解析については、実験のプロトコルや解析パイプラインの詳細についてもまとめた（図 10）。

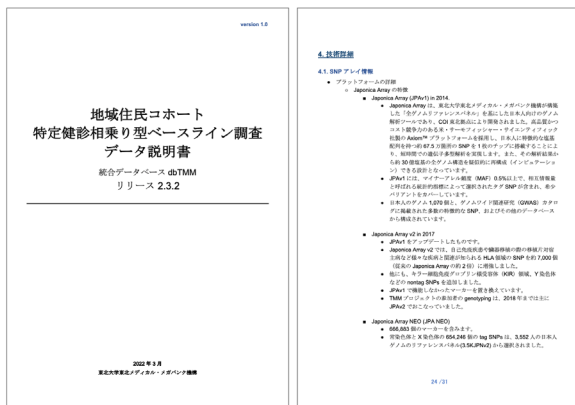


図 10 データ説明書

D. 考察

コホート横断検索システムに対する産業界のニーズを把握し、時系列検索機能の拡充による高度化を行った。また、コホートの収集データの項目のカタログも閲覧できるようにしたいというニーズを受けて、コホート横断検索システムカタログを研究開発した。連携する官民研究開発投資拡大プログラム(PRISM) バイオ技術領域から協力を得ての施策の状況を踏まえて、各コホート研究で収集したデータの項目のメタデータを標準化した。また、コホート横断検索システムカタログを、API を設計して開発した。この API を使用することで、今後、国外のコホートとの連携した、コホートの収集データのデータ発見につながると考えられる。

企業向け相談窓口の整備として Web 上で伴走支援を行うシステム(企業向けWeb相談申請システム)及び企業向け情報提供ウェブサイトの研究開発を行った。さらに、企業向けのデータの利活用を促進するため、コホートのデータの説明書として、東北メディカル・メガバンク計画のコホートの収集データの説明書の研究開発を行った。

本研究により研究開発したコホート横断検索システムやカタログ、企業向けWeb相談申請システムを 2022 年度に実証し、2023 年度で実装及び運用することにより、コホート・バイオバンクの産業利用が促進し、多様な分野において科学的エビデンスに基づいたサービスを提供できる環境を整備し、エビデンスに基づいたデータ駆動型のヘルスケア産業を創出するとともに、データを統合・強化する

大規模健康常人コホート・バイオバンクの構築を通じて、国民の QOL 向上に資する疫学研究の発展に寄与することが期待される。

E. 結論

次世代バイオデータ基盤の構築に向けて、産業界のニーズを踏まえて、各コホート研究のデータを横断して検索するシステムの高度化を行った。また、コホート研究のデータ項目の標準化されたメタデータを公開するコホート横断検索システムカタログの研究開発を行った。さらに、企業向け相談窓口の整備として Web システムの開発を行い、データの利活用を促進するため、データの説明書の研究開発を行った。2022 年度からは、官民研究開発投資拡大プログラム(PRISM) バイオ技術領域「糖尿病個別化予防を加速するマイクロバイーム解析 AI の開発」、「認知症に関与するマイクロバイーム・バイオマーカー解析」と連携して、コホート横断検索システムやカタログ、企業向けWeb相談申請システムの実証に取り組む。

F. 健康危機情報 なし

G. 研究発表

- | | |
|---------|----|
| 1. 論文発表 | なし |
| 2. 学会発表 | なし |

H. 知的財産権の出願・登録状況

- | | |
|-----------|----|
| 1. 特許取得 | なし |
| 2. 実用新案登録 | なし |
| 3. その他 | なし |