

戦略プロポーザル

健康・医療リアルワールド データ利活用基盤の構築と 生成 AI への展開

STRATEGIC PROPOSAL

Health & medical real world data infrastructure

- Catalyst of generative AI development in Japan -

エグゼクティブサマリー

本戦略プロポーザルは、わが国における健康・医療RWD¹（リアルワールドデータ）の収集、流通を円滑に行い、健康・医療RWDに基づいた知の創出とその活用を加速する健康・医療RWD利活用基盤の構築を目指すものである。さらに、知の創出と活用における最重要課題として、医療版LLM²（大規模言語モデル）/基盤モデル³と診療支援ツールの研究開発戦略を提案する。

わが国は世界一の長寿国であり高い健康水準を実現しているが、それを支えているのは国民皆保険制度をはじめとした医療・ヘルスケアシステムである。一方、臨床現場の疲弊、硬直的な医療資源の配置、個別化医療の研究開発と創薬の遅れ、医療情報産業の停滞などの危機的状況が新型コロナ禍で顕在化した。増大を続ける国民医療費は令和3年度には45兆円を超え、わが国の医療・ヘルスケアシステムの持続可能性、ひいては将来的な国民の健康水準の維持が危ぶまれている。この状況に陥った要因の1つに、健康・医療RWD利活用基盤の不備がある。

わが国における健康・医療RWDに関する主な問題は、健康・医療RWDの①収集システムの不備、②流通システムの不備、③解析による知の創出とその活用の遅れ、である。特に③について、海外ではLLMをはじめとした生成AIが医療・ヘルスケア分野においても急速に実装されつつある。わが国が医療・ヘルスケア分野に生成AIを活用する際、海外で開発されたLLMや基盤モデルに依存すると、経済安全保障上のリスクや個人の医療情報流出などの懸念があるため、国内での医療版LLM/基盤モデルの開発が喫緊の課題である。それらの開発に必要となる大規模な健康・医療RWDを蓄積するには、まずは①②の問題解決が不可欠である。

以上に鑑み、本戦略プロポーザルではわが国が取り組むべき3つの提案を行う。

【提案①】健康・医療RWD収集を加速する技術の開発

臨床現場の負担を軽減しつつ高品質な診療データ収集を実現するため、従来のアプローチである電子カルテ入力用のテンプレート整備と並行して、医療版LLMおよびそれをベースとした診療記録作成支援ツール（診療記録の下書き、用語統一や自動コード⁴化など）を開発する。国立情報学研究所（NII）を中心とした国内での汎用LLM開発と連携し、例えば医療オントロジー（医療特有の概念と用語の対応関係）の組み込みなど、汎用LLMを医療分野に適応するための最適な手法を検討する。その前提となる、HL7 FHIR⁵準拠のデータ交換規格整備とそれに対応した電子カルテの開発、傷病名・医薬品・検査などのコード整備を合わせて実施する。

【提案②】一次利用と二次利用を前提とした健康・医療RWD流通の仕組みや制度の整備

【提案①】で収集した健康・医療RWDを円滑に流通させるための仕組みや制度を整備する。その際、患者への医療・ヘルスケア提供に患者個人のデータを直接的に活用する一次利用の観点からは、医療機関間のデータ共有やPHR⁶へのデータ提供が求められる。一方、新規医療技術開発や医療費のマク

- 1 Real World Dataの略語。診療現場などさまざまな情報源から日常的に収集された、患者の健康状態や医療の提供に関するデータのこと。付録3参照。
- 2 Large Language Modelの略語。生成AIの一種で、大量かつ多様な言語データを用いて訓練され、ファインチューニングなどによりさまざまなタスクに適応できる大規模な深層学習モデルのこと。2.1.4参照。
- 3 Foundation Modelの日本語訳。LLMと同様に大規模データを学習させたモデルであるが、言語以外にも画像、音声などマルチモーダルなデータを大量に学習させ、より幅広いタスクに適応できることが特徴である。2.1.4参照。
- 4 病名・検査・薬品などを分類・指示するためのコード。付録2.2参照。
- 5 HL7（Health Level Seven）はコンピュータ間での医療文書情報のデータ交換を標準化するための国際規格で、FHIR（Fast Healthcare Interoperability Resources）はHL7の最新の規格。付録2.2参照。
- 6 Personal Health Recordの略語。医療機関や自治体などで医療・ヘルスケアサービスの中で収集した患者の情報を、患者自身が閲覧できる仕組み・ツールのこと。付録3参照。

口分析などにデータを活用する二次利用の観点からは、用途に応じた品質と規模のデータが必要となる。一次利用と二次利用の両方のニーズを満たす健康・医療RWD流通の仕組みを、NDB⁷（レセプト情報・特定健診等情報データベース）と次世代医療基盤法⁸で規定された仕組みをベースに整備する。二次利用で創出された知の活用を促進するため、研究開発だけでなくその成果の商業化への健康・医療RWDの利用も前提とした同意プロセスと審査体制を構築する。

【提案③】医療版基盤モデルの開発および社会実装のための法規制の整備

【提案①】【提案②】で構築する健康・医療RWD利活用基盤を活かして、健康・医療RWDを大規模に蓄積するデータウェアハウス⁹を構築し、医療・医学の知的基盤として言語と画像のマルチモーダルな医療版基盤モデルを開発する。医療版LLM/基盤モデルをベースとした医療AIの法規制の検討を、【提案①】および【提案③】における医療版LLM/基盤モデルの開発・実証実験と並行して進める。匿名加工情報やAIモデルからの個人推定など、生成AIを含む医療ビッグデータには悪用への懸念があり、健康・医療RWDの利活用の妨げとなりうる。代表的な懸念である遺伝情報・ゲノム情報による差別に関して、その防止に向けた具体的な対応策を検討する。

以上の提案全体を実現するために、医療版LLM/基盤モデルの開発や社会実装の難易度に基づいたマイルストーンの設定、およびその実現に向けた各提案の推進テーマと時間軸の設計が必要である。推進方法としては、【提案①】～【提案③】のそれぞれについて、アカデミアを中心に関連する府省や企業が一体となり、内閣府SIP「統合型ヘルスケアシステムの構築」や医療DXなど関連する既存事業とも連携した体制を構築することを提案する。健康・医療RWDの利活用促進に対する社会受容・社会との協働に向けては、【提案①】～【提案③】により解決すべき問題と合わせて実現を目指す効果を訴求することにより、健康・医療RWDの提供元かつ利活用の最大の受益者となる個人・市民、および収集・提供の窓口となる医療従事者の理解醸成を図ることが望ましい。

上記の健康・医療RWD利活用基盤は、わが国の医療関連産業の発展、医療制度・提供体制の改善や診療の高度化、ヒトを対象としたデータ駆動型医学研究の加速に不可欠である。特に、言語と画像を組み合わせるマルチモーダルな医療版基盤モデルは、わが国における画像診断機器開発での強みを生かす。本プロポーザルの実行を通じて、わが国の医療における蓄積を活用しながら、医療・ヘルスケアシステムの持続可能性を高め、ひいては国民の健康に寄与することが期待される。

7 National Databaseの略語。外来・調剤・入院レセプト情報や特定健診等の情報を二次利用するためのデータベースのこと。付録3参照。

8 正式名称は「医療分野の研究開発に資するための匿名加工医療情報及び仮名加工医療情報に関する法律」。付録2.4参照。

9 さまざまなデータソースからデータを集め、分析を行うために構造化したデータを収納したデータベース。付録3参照。

Executive Summary

This strategic proposal outlines:

- 1) The building of a data infrastructure in Japan to accelerate health & medical real world data (RWD) collection, efficient distribution of collected data to stakeholders and end users while maintaining patient privacy, utilization of RWD for medical & clinical intelligence creation and application;
- 2) Development of a medical version of “large language model (LLM)” / “foundation model” and medical treatment supporting tools based on them.

Japan has the unique distinction of being the world's longest-lived country and a corresponding attainment of a high level of its people's health standard, which are sustained by its medical and healthcare system, including the universal health insurance system. On the other hand, Japan's medical and healthcare system has begun showing signs of stress and stagnation, such as exhaustion at clinical sites, inflexible allocation of medical resources, delays in R&D and drug discovery of personalized medicine, and inefficiency in medical information industry, became apparent during the COVID-19 pandemic. National medical expenses have exceeded 45 trillion yen in FY2021 and is further increasing year by year. The sustainability of Japan's medical and healthcare systems is declining, and it will have a negative impact on the health standards of the people in the future. One factor that has led to this situation is ineffective utilization of the health and medical RWD due to lack of the solid data infrastructure in Japan.

Some of the major problems related to the utilization of health and medical RWD in Japan are (1) an inadequate health and medical RWD collection systems, (2) an inadequate health and medical RWD distribution systems, and (3) delay in creating intelligence through analysis of health and medical RWD and its social implementation. In regard to the latter, the development of generative Artificial Intelligence (AI) such as LLM and its application are rapidly progressing in development worldwide. Economic security and leakage of personal medical information become added risks as domestic medical services come to rely on LLMs / foundation models developed a non-Japanese ecosystem. As a result, the development of a medical LLM / foundation model in Japan is an urgent issue. In order to accumulate large-scale health and medical RWD required for its development, it is essential to first solve problems (1) and (2).

Toward this end, we propose the following three subjects should be tackled in Japan.

(1) Investment in technology to accelerate health and medical RWD collection.

To simultaneously realize the acceleration of medical care data collection and reduction of the burden on clinical sites, [Subject 1] proposes:

- Develop a medical LLM and tools for medical-recording support (drafting medical treatment records, unifying medical terminology, automatic coding, etc.) based on the LLM, in parallel with the development of templates for electronic medical record input, which is a conventional approach.

- Study optimal domain adaptation of a general-purpose LLM in conjunction between a working group initiated by National Institute of Informatics (NII), and the medical community, e.g., the incorporation of medical ontology (correspondence between medical concepts and terms).
- Develop HL7 FHIR-compliant medical data exchange standards and an electronic medical record tool conforming to it, and codes for names of injuries / diseases / drugs / clinical tests etc., which are the preconditions of the development of medical-recording support tool.

(2) Development of health and medical RWD distribution system for its primary use and secondary use.

From the perspective of primary use, in which individual patients' data are directly utilized for their own medical care and healthcare, data sharing among medical institutions and the processes for provisioning to PHR (Personal Health Record) are required. From the perspective of secondary use, in which health and medical RWD is utilized for R&D such as the development of new diagnosis/treatment technology and macro analysis of medical expenses, a particular quality and scale of RWD that is commensurate with the research goal is required. For smooth distribution of health and medical RWD collected through the system established in [Subject 1], [Subject 2] proposes:

- Establish a health and medical RWD distribution system that meets the requirements of both primary and secondary use based on the receipt information / specific medical examination information database (NDB: National Data Base) and the Next Generation Medical Infrastructure Law.
- Build a consent process and review process on the premise that health and medical RWD will be utilized for academic discovery as well as commercialization to promote utilization of intelligence created by secondary use.

(3) Development of a medical foundation model and the laws and regulations for its implementation.

By making use of the health and medical RWD infrastructure built in [Subject 1] and [Subject 2], [Subject 3] proposes:

- Build a data warehouse to accumulate health and medical RWD on a large scale, and develop medical version of multimodal (language, image) foundation model as a groundwork for medical/clinical intelligence.
- Study the adequate regulation of medical AI based on the medical LLM / foundation model in parallel with its development and implementation.
- Discuss concrete measures to prevent discrimination based on genetic and genomic information as one of the most serious consequences by misuse/abuse of medical big data and AI such as personal estimation from anonymously processed information or AI models, so that concerns about such misuse/abuse won't hinder the utilization of health and medical RWD.

Milestones for the overall project will be set based on the degree of difficulty in developing and implementing a medical LLM / foundation model, and its concrete sub-subjects and timeline

for each subject will be designed to achieve these milestones. For each of [Subject 1] to [Subject 3], academia and other related ministries, agencies, and companies will work together to promote the project, in collaboration with related existing projects such as the Cabinet Office SIP "Establishment of an Integrated Health Care System" and "Medical DX". In order to achieve social acceptance and collaboration with the broader public, the effects to be achieved in conjunction with the problems to be solved through [Subject 1] - [Subject 3] will be appealed. This is expected to foster understanding among individuals and citizens, the primary providers and the greatest beneficiaries of health and medical RWD, as well as medical professionals, the contact points for the collection and provision of RWD.

By building a health and medical RWD infrastructure envisioned in this strategic proposal, the development of medical related industries, the improvement in medical treatment outcomes, and acceleration of human data-driven medical research would be realized. In particular, the medical foundation model of language and image could further strengthen Japan's presence in diagnostic imaging equipment. By these effects, it is expected that the sustainability of Japan's medical and healthcare system would be enhanced and consequently the health of the people would be maintained and further improved.

目次

1	提案の内容	1
2	提案を実施する意義	4
2.1	現状認識及び問題点	4
2.2	提案が実施された時の効果	12
3	具体的な提案の内容	14
3.1	提案の全体像	14
3.2	個々の提案の内容	14
4	提案実施上の推進方法及び時間軸	20
4.1	推進方法	20
4.2	時間軸	23
付録 1	検討の経緯	25
付録 2	わが国における問題点の補足説明、国内外の関連する状況など	28
付録 3	専門用語説明	43
	参考文献	45

1 | 提案の内容

本戦略プロポーザルは、わが国における健康・医療RWDの収集、流通を円滑に行い、健康・医療RWDに基づいた知の創出とその活用を加速する健康・医療RWD利活用基盤の構築を目指すものである。さらに、知の創出と活用における最重要課題として、医療版LLM/基盤モデルと診療支援ツールの研究開発戦略を提案する。具体的には、次の3つを提案する。

【提案①】健康・医療RWD収集を加速する技術の開発

【提案②】一次利用と二次利用を前提とした健康・医療RWD流通の仕組みや制度の整備

【提案③】医療版基盤モデルの開発および社会実装のための法規制の整備

医療版LLM/基盤モデルの開発には大規模な健康・医療RWDが不可欠であり、その観点からも【提案①】【提案②】による健康・医療RWD利活用基盤の整備は重要である。

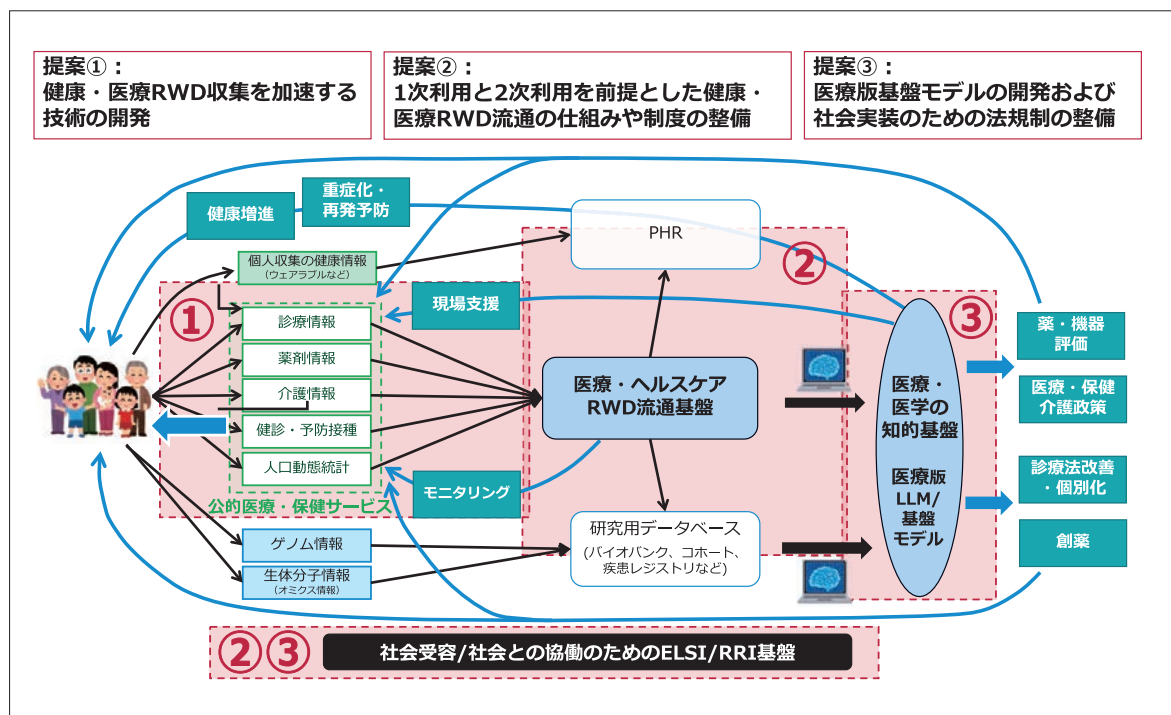


図1-1 提案の全体像

【提案①】健康・医療RWD収集を加速する技術の開発

従来の健康・医療RWD収集改善のアプローチである1)と並行して、2)の研究開発を提案する。

- 1) 電子カルテのデータ交換規格対応および各種コードの整備、臨床現場が効率的に入力可能な電子カルテのテンプレート整備

国内におけるHL7 FHIR準拠のデータ交換規格の整備とそれに対応した電子カルテ開発・導入、および傷病名・医薬品・検査等において優先度の高い項目のコード整備。一般的な疾患・症例に対応する標準的なデータ項目の設定、およびそれをベースとした電子カルテベンダーによらない共通テンプレート開発

想定する推進主体：関連学会を中心に、厚生労働省と電子カルテベンダーなどが一体になって推進

- 2) 国内開発のLLMを活用した診療記録作成支援ツールの開発（診療記録の下書きや用語統一、自動コー

ド化などを想定)

国立情報学研究所（NII）などで開発中の汎用的な日本語 LLM を医療分野に適応させた医療版 LLM の開発、医療オントロジー（医療特有の概念と用語の対応関係）の活用による効率的な開発の検討

想定する推進主体：1) の推進体制を軸に、NII 中心に進める LLM 勉強会などと連動して進める

【提案②】一次利用と二次利用を前提とした健康・医療 RWD 流通の仕組みや制度の整備

二次利用の用途に応じて、悉皆性の高い（患者、医療機関等のカバー率が高い）データおよび高品質な（対象者1名に対して収集するデータが充実している）データが必要になることから、それらのデータ流通の促進に向けた仕組みをそれぞれ1)、2) で提案する。3)、4) は、より幅広いデータとつなぐ仕組み、民間事業者による商業化を想定した仕組みの構築のための提案である。

1) 悉皆性の高いデータプラットフォームの整備・拡充

NDB をベースとした、医療費や医療資源などのマクロ分析の解析機能、解析環境の整備

想定する推進主体：厚生労働省を中心に推進

2) 診療情報データベース間で診療情報を連携する規格・仕組みの整備

次世代医療基盤法をベースとして、情報取扱事業者（図1-2の『データ交換ネットワーク』機能に相当）と加工事業者（同『データキュレーター』機能）が独立した制度の整備、および情報取扱事業者間でのデータ交換が可能な仕組みの実装

想定する推進主体：内閣府を中心に、データ交換の仕組みの構築は関連学会が協力して推進

3) 本人同意に基づき PHR や研究用データベースに情報を提供する仕組みの整備

オンライン資格確認の仕組みを用いた情報提供プロセスを実装

想定する推進主体：厚生労働省を中心に文部科学省、総務省などが連携し実施

4) 民間利用を前提とした審査体制と同意プロセスの整備

利活用目的に応じて解析用データ提供・アクセスを規制する制度整備と、それを適正かつ円滑に審査する仕組みの構築、および民間事業者による商業化を前提とした健康・医療 RWD 収集への同意プロセスの導入検討

想定する推進主体：生命科学・医学系研究に関する倫理指針の議論の場を中心に検討

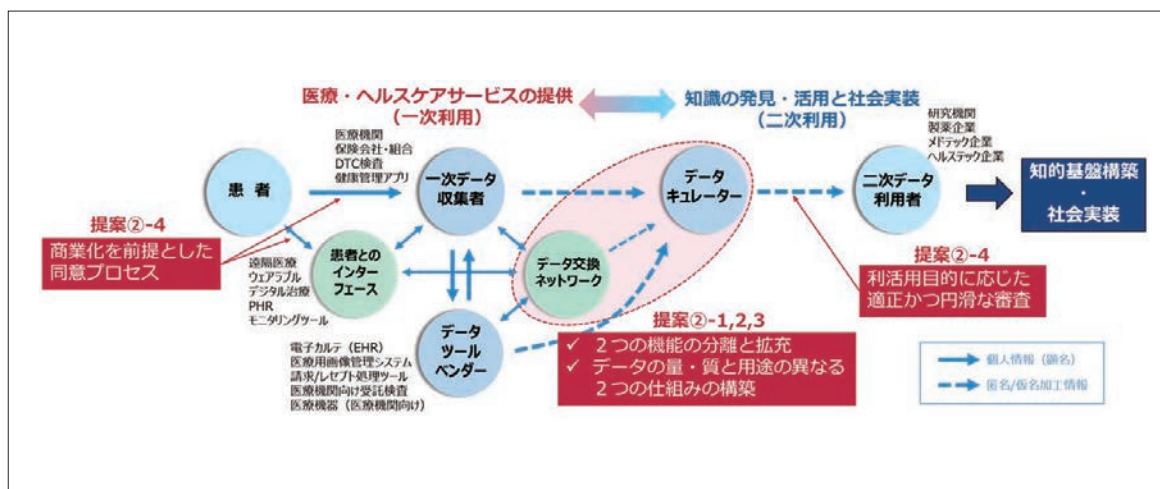


図1-2 医療・ヘルスケアRWD流通の概要と【提案②】の位置づけ

【提案③】医療版基盤モデルの開発および社会実装のための法規制の整備

【提案①】【提案②】を通じて構築する健康・医療 RWD 収集・流通の仕組みを最大限に活用し、1) で大規模な健康・医療 RWD 蓄積および医療の知的基盤としての基盤モデル開発を提案する。2) 3) は、医療ビッグデータや基盤モデル、それをベースとした医療 AI の利用への懸念に対応した法的・倫理的基盤整備の提案である。

1) 医療版基盤モデルの開発とそれに向けたデータウェアハウス構築

基盤モデルを含めた AI の学習に幅広く活用できる大規模データの蓄積、大規模データを活用した【提案①】の医療版 LLM の改良とそれをベースとした言語 + 画像のマルチモーダルな基盤モデルの開発および社会実装（画像検査サマリの作成、画像と所見を踏まえた疾病候補提示などを想定）

想定する推進主体：【提案①】2) の推進体制を軸に、既存の画像データベースと連携

2) 臨床現場での LLM/ 基盤モデルベースの医療 AI 活用における規制の検討

臨床現場で医療版 LLM/ 基盤モデルをベースとした医療 AI を実装する上で必要となる薬事承認、予期せぬエラーやハルシネーション¹⁰などに対応した制度の設計

想定する推進主体：規制当局（PMDA）などを中心に、国内の関連する取り組み（汎用 LLM 開発、医療版 LLM 開発など）を共有しながら議論を進める

3) 倫理的に不適合な医療ビッグデータ・AI 活用の防止

遺伝情報・ゲノム情報に基づく差別を防止するための具体的な対応策の検討

想定する推進主体：遺伝子情報に基づく差別禁止において先行する諸外国の実態をアカデミアと民間事業者で調査し、わが国における検討に生かす

【提案①】～【提案③】の全体の推進方法・時間軸として、短期的、中期的なマイルストーンをベースに各提案で推進すべきテーマをとして設定し、例えば健康・医療戦略推進本部などの司令塔の下で各提案の進捗や提案間の連携を統合的に管理する。各提案は、内閣府 SIP「統合型ヘルスケアシステムの構築」など関連するプロジェクトと連携しつつ、各提案間で産官学のステークホルダーが一体となって推進する。データの提供源かつ最大の受益者となる個人・市民および収集・提供の窓口となる医療従事者の理解醸成を図りつつ、PPIP¹¹などを活用した懸念の把握・解消も合わせて実施することが望ましい。医学、情報科学・データ科学両面の科学的意義を明示して双方の研究者を引き込むのと同時に、構築後の健康・医療 RWD 利活用基盤の長期的な運用を担う人材確保も念頭に、医学、データ科学、法学、倫理など、異なる学術分野を跨いだ人材育成が必要となる。健康・医療 RWD 流通における経済的な持続性のあるコスト負担の在り方、医療版 LLM/ 基盤モデルの継続的な改良を前提とした運用方法、データ提供・キュレーションに対する適切なインセンティブの設定など、本戦略プロポーザルで構築する基盤の継続性についても併せて検討すべきである。

10 Hallucination（幻覚、妄想）。人工知能の分野では、AI が学習データに基づかない回答を生成し、もっともらしい嘘を出力してしまう現象を意味する。

11 Patient and Public Involvement Panel の略語。医学研究・臨床試験プロセスの一環として、研究者が患者・市民の知見を参考にすること。付録 3 参照。

2 | 提案を実施する意義

2.1 現状認識及び問題

わが国の医療・ヘルスケアシステムには根深い問題が数多く存在し、危機的と言える状況であるが、本戦略プロポーザルで述べる健康・医療 RWD 利活用基盤の構築は、問題解決に向けた重要な取り組みとなりうる。本章では、わが国の医療・ヘルスケアシステムの問題を述べた後、健康・医療 RWD の利活用（データ収集、流通、知の創出とその活用）に関わる具体的な問題を述べる。なお、それら問題について、付録2においても詳細を掲載しているため、適宜ご参照頂きたい。

2.1.1 わが国の医療・ヘルスケアの問題

わが国は世界一の長寿国であり高い健康水準を実現しているが、それを支えているのは国民皆保険制度をはじめとした医療・ヘルスケアシステムである。わが国の国民皆保険制度は、先進国の中では安価で公平でアクセスしやすい医療を実現しており、国際的にも高い評価を得ている^[1]。こうした優れた医療・ヘルスケアシステムを今後も維持・発展させていくことが望ましいが、主に次の5つの問題が深刻さを増している。

◆臨床現場の疲弊

わが国に限らず、臨床現場の疲弊が問題となっている。電子カルテの導入が進んだ結果、医師は電子カルテの入力作業に多くの時間を割かれ、患者に向き合う時間や新たな知識を得る時間が減少している。これは医師の燃え尽き症候群の原因の1つでもある。さらに、わが国において電子カルテを研究利用する場合、電子カルテに入力されたデータを研究用データベースへと医療従事者が手作業で転記をせねばならず、負担が非常に大きい。

健康・医療 RWD の収集だけでなく、データ解析で得られた知識の臨床現場における利活用にも問題がある。近年、医学・診療に関する知見が急激に増加しており、多忙な医療従事者がそれら全てを把握することはほぼ不可能であることから、AIを活用してこのような状況を打破しようとする試みが行われているが、成功しているとは言い難い。例えば、約10年前からIBM社がAIプログラムWatson（ワトソン）を医療分野へ展開し、過去の論文などから抽出した知識を基にがんの治療方針を提案するプロジェクトを進めてきたが、電子カルテの記載様式がバラバラでデータキュレーションを人海戦術に頼らざるを得ず、さらに医療機関への実装が進まず収益が上がらなかったことから、商業的には失敗に終わった。医療機器として開発・認可された一部のAI医用画像診断などを除き、医師の知識習得とその実践は今なお人の手で行われている状況である。ただし、生成AIの医療分野への応用例が2023年ごろより見られるようになり、閉塞した現状のブレイクスルーとなる可能性がある。

◆地域格差解消、持続的な医療提供に向けた資源再配分の遅れ

わが国は広く薄く医療資源を配置しており、集中治療が必要な救急患者や重症患者のケアは手薄となる。その限界が露呈したのがCOVID-19まん延時の医療ひっ迫であり、わが国は人口当たり病床数が世界最多であるにもかかわらず^[2]、緊急入院ができず自宅で死亡する重症者が相次いだ。人口減少や高齢化が進み、地域ごとの医療ニーズが変化する中、限られた医療資源の効果的な再配置は遅れている。わが国は、行政が医療費の総額と個々の診療報酬を管理している一方で、医療を提供する病院の8割が民間経営であり^[2]、国や

自治体の指示だけでは医療資源を動かすことができない。平素より医療現場の状況を示すデータに基づいて関係者が議論し医療資源配分の最適化を進めることが重要であるが、そのためのデータが限られている点が問題である。

◆これからの医療・創薬の潮流への対応遅れ（データに基づく個別化・層別化医療）

医療提供にあたり、平均化した疾患発症・重症化モデルでは、患者1人1人に最適な対処ができない。例えば虚血性心疾患は、数十年にわたる時間経過の中で徐々に進行し、発作を何度も起こし、最終的に重篤な発作を起こし死に至る疾病であり、発作のコントロールがきわめて重要だが、発作を起こす年齢、発作の頻度、重症化のスピードなどは患者1人1人で異なる。大量の医療データを継続的に収集し、その解析により患者を層別化・個別化することで、適切なタイミングで適切な治療技術の提供が可能となるが、わが国ではそのためのデータ基盤の整備が遅れている。創薬においても、世界の医薬品市場が100兆円規模を超える中、これから大きな市場を形成すると考えられる個別化・層別化医療におけるわが国の遅れは深刻な問題である。

◆国内の医療情報産業の停滞

わが国の医療情報の電子化率は伸び悩んでおり、さらに国内の電子カルテ大手3社の相互情報交換や標準化の歩みも遅い。海外の多くの国々ではそれらが一気に進展し医療データ基盤が構築され、例えばCOVID-19対策においても医療データが活発に活用されたのとは対照的である。近年、国内の医療情報産業に海外企業の参入が進む。例えば、聖マリアンナ医科大学病院にezCaretech社（韓国）、国立がん研究センター東病院にFlatiron社（米国/スイス、Roche子会社）の電子カルテが導入された。手術ロボットDa Vinci（Intuitive Surgical社、米国）の導入も進み、わが国の外科医の手技データは同社に蓄積している。わが国の医療情報が海外に流出、あるいは流出しうることが見られ、安全保障、経済安全保障の観点からも見過ごせない問題である。

◆医療費、介護費の増大

わが国に限らず、高齢化などに伴う国民医療費・介護費の増大が、国家財政を圧迫している。例えばわが国の医療費は、令和3年度には45兆円にまで膨れ上がった。医療費・介護費を含む社会保障費は、30年前と比べると3倍以上に増大したが、税金などの収入は1-2割しか伸びていない。これからもさらなる高齢化が予想されており、このままでは国民医療費・介護費と税金の不均衡はさらに悪化しかねない。

2.1.2 健康・医療RWD収集における問題

わが国で収集される主な健康・医療RWD（表2-1）の中で、特に有用性が高いと考えられるのは、医療機関で取得される診療情報（電子カルテ）である。診療情報には、診断病名や医師の所感、診断の根拠となる検査値など、健康・疾病状態に関する詳細な情報およびアウトカムを示す情報が含まれていることから、診療・医療提供の改善および新たな知の創出などにおいて重要な情報源となる。

表2-1 わが国で収集される主な健康・医療RWD

	データの種類	データ保管先	データベース
医療情報	レセプト（外来）	医療機関、保険者[健保組合、自治体]	ナショナルデータベース（NDB）
	DPCLレセプト（入院）	医療機関、保険者[健保組合、自治体]	DPCデータベース
	診療情報（電子カルテ）	医療機関	
	医用診療画像	医療機関	

薬剤情報	レセプト（調剤）	薬局、保険者[健保組合、自治体]	ナショナルデータベース（NDB）
	処方箋	医療機関、薬局	
介護情報	介護保険情報	介護施設、保険者[健保組合、自治体]	介護保険総合データベース
健診情報	定期健診	事業者（雇用主）、自治体（保健所）、学校	
	特定健診（40-74歳）	保険者[健保組合、自治体]	
	後期高齢者健診（75歳-）	広域連合[都道府県]	
	妊婦・乳幼児健診	自治体（保健所）	
予防接種情報	予防接種記録	自治体（保健所）	
人口動態情報	戸籍に関する記録（出生・死亡届など）	自治体	

これまで、疾患レジストリなどの臨床研究用のデータベース構築や、個人が健康情報を閲覧・活用するためのプラットフォームであるPHRへの組み入れのために、電子カルテに収載されている診療情報を抽出し収集する取り組みが行われてきた。しかし、いずれも小規模な取り組みに留まる。新型コロナ禍での医療逼迫の反省も踏まえ、2022年に「電子カルテ情報の標準化」や「医療DX推進本部の設置」などが政府の方針として打ち出され、診療情報の利活用の重要性が幅広く認識されつつある。

医療機関におけるRWD収集の問題点として、電子カルテの普及率の伸び悩み（データがデジタル化していない）、入力項目の標準化の遅れ（データが不均一）、データ構造化の遅れ（解析できる形式ではない）、オンラインで利用可能なデータ交換規格の不備（データが共有できない）などが挙げられる。データ交換規格については厚労省がHL7 FHIRの導入を進めていることから、本項では最初の3つの問題点を概説する。これらに加えて、臨床現場の疲弊が深刻化しており、データ収集に係る臨床現場の負担の大きさも問題となっている。

【問題①-1】電子カルテの普及率の伸び悩み

わが国では、病院内で収集される情報は早くから紙媒体から電子化への移行が進んでいた。中でもレセプトは1980年代後半の時点で90%近くが電子化した。一方、電子カルテの電子化は徐々に進んでいるものの、令和2年時点で400床以上の病院では9割、200床未満の病院や一般診療所では5割に留まる¹³⁾。レセプトは悉皆性の高いデータといえるが、電子カルテは現時点では道半ばである。

【問題①-2】入力項目の標準化の遅れ

データの質を均一化するためには、多様な診療課題（疾病・症例・処置法）に対して、電子カルテに記録すべき項目が予め設定されていることが望ましい。しかし、現時点では標準的な記録項目の策定に向けた活動はあるものの、全国的に医療機関や医師に認知され徹底されたものはなく、バラつきが大きい。例えば「血清総コレステロール」「総Cholesterol」「T-コレステロール」などの表記揺れが見られ、これらはデータ分析を行う際に異なる項目と誤認される可能性があるため、それを防ぐための分類コードが必要である。現状、病名については、疾患分類コードの国際標準であるICD-10があり、それに準拠する形でわが国の「標準病名マスター」が制定され、70%以上の医療機関で使われている。病名・検査・薬品コードは厚生労働省推奨となっているが、全ての医療機関で使われてはいない。

【問題①-3】データ構造化の遅れ

電子カルテは、フリーテキストで書かれていることが多い。電子カルテ入力時に構造化テンプレートを作成・導入し電子カルテデータの構造化を図るケースもあるが、病院内に限った構造化に留まっている。

その背景として、国内の有力な電子カルテベンダー各社の足並みが揃っていないことが挙げられる。ハーバード大学関連病院や東京大学のCOVID-19研究の事例を見ると、それぞれの施設で開発した自然言語処理モデルを活用した、電子カルテデータの構造化と情報抽出は実現しつつある。しかし、わが国は電子カル

ベンダー間で共通化・標準化されている項目が少なく、さらに同じベンダーでも病院ごとにカスタマイズが進んでおり、1つの自然言語処理モデルを適用してさまざまな病院の診療情報の抽出と構造化を実施することは今なお困難である。

2.1.3 健康・医療RWD流通における問題

米国の健康・医療RWD/RWE（Real World Evidence）の流通スキームの概要図を図2-1に示す^[4]。本人の情報閲覧や医療・ヘルスケアサービスの提供に必要な本人データの利用である一次利用においては「データ交換ネットワーク」、データに基づく知識発見とその活用を目的とした集団のデータの利用である二次利用においては「データキュレーター」が健康・医療RWD流通の要である。米国では関連するスタートアップが数多く登場し、健康・医療RWD利活用が加速している。一方、わが国では「データ交換ネットワーク」「データキュレーター」を担うプレイヤーが少ない。

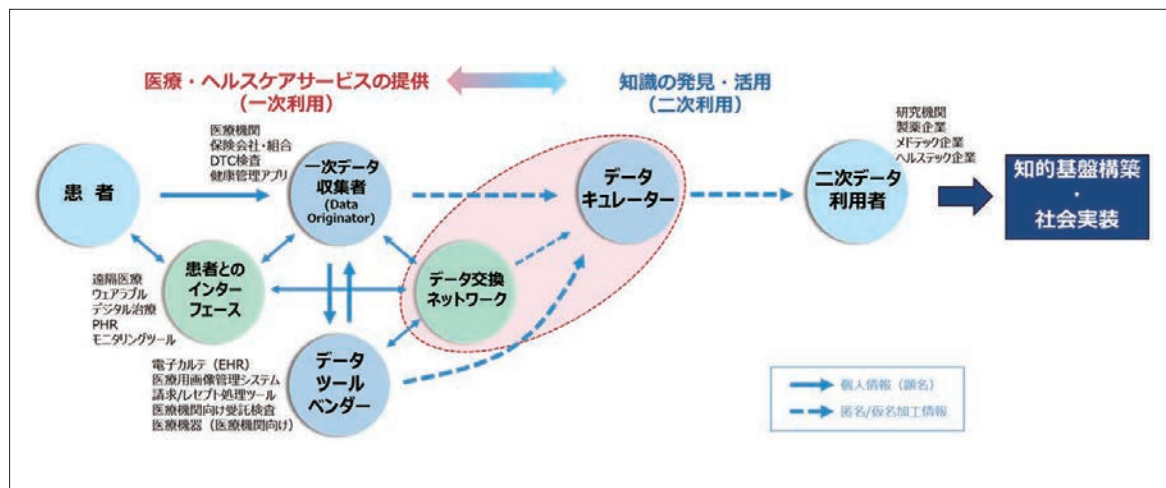


図2-1 米国における健康・医療RWD/RWEの一次利用・二次利用の流れ

出典：文献^[4]の図を基に一部改変

二次利用を通じて新たな価値を創出するためには、データの量（健康・医療RWD収集の対象者数）と質（対象者1名に対して収集するデータの充実度）が重要となるが、用途に応じて、必要とされるデータの質と量は異なる。医療費、医療資源のマクロ分析を行う際は、悉皆性の高いデータ（量の優先度が高い）が特に求められ、個別化・層別化した診断・治療技術開発においてはデータの質の高さが特に求められる。それぞれにおいてわが国には次の通り問題点が存在する。

【問題②-1】悉皆性の高い健康・医療RWD流通・集約のための仕組みの不備

わが国では、外来、入院、処方、介護のレセプトおよび特定健診の情報が審査支払機関に集約され、名寄せ・匿名化がなされた上でNDBに保管されている。しかし、NDBが匿名化を前提とするため、二次利用における制約が大きい。保険番号を利用するため名寄せの精度に限界がある、診療情報など他の情報と組み合わせた分析ができない、などの問題もある。また、データモニタリング、ダッシュボード表示といった分析機能は有しないため、国や自治体による政策立案への利活用も進んでいない。

コラム1

英国、米国における医療データ流通・集約の仕組み
(事例)

悉皆性の高い医療データ流通・集約については、英国NHS（National Health Service）の仕組みが参考となる^[5]。英国では、医療サービスは公的プロバイダー（NHS傘下の病院およびGP：かかりつけ医）により提供されるが、公的プロバイダー間のデータ交換ネットワークの構築・運営・管理をNHS傘下のNHS Digitalが担っている。さらにデータ蓄積およびデータキュレーターの役割もNHS Digitalが担う。ただしデータの種類は診療サマリ情報、電子処方箋など必要最低限に限定されている。電子カルテは民間ベンダーが供給するが、英国政府が定める規格への準拠が実質的に必須となっている。

質の高い医療データ流通・集約については、米国の仕組みが参考となる^[4]。米国保健省の支援の元、診療データ交換の促進を進めるためのSequoia PJが2012年より開始した。Sequoia PJでは、FHIRに対応したネットワーク間データ交換のフレームワーク「Carequality」、およびそれを実装したネットワーク（中小規模医療機関ネットワークをつなぐ）である「eHealth Exchange」を構築し、それぞれ2018年に独立した組織となった。現在、全米のヘルスケアネットワークのうち70%以上が「eHealth Exchange」に接続している^[6]。医療機関同士だけでなく、民間保険者やデジタルヘルス開発者などが診療データに接続する場合もこうしたネットワークを介することができるほか、異なるデータ交換規格のネットワーク間を変換・接続する機能を有する事業者も存在する。データキュレーターについては公的なイニシアティブはないが、スタートアップ/民間企業が、上記ネットワークなどを活用してデータ集約、加工と提供を担っている。また、Mayo Clinicのような大規模医療機関グループや大学病院（一次データ収集者）、Epic社のようなシェアの高い電子カルテベンダー（データツールベンダー）によるネットワークなど、質が高くかつ一定規模以上の量もあるデータソースが存在するのが米国の強みであるといえる。

【問題②-2】質の高い診療情報の流通・集約のための仕組みの不備

わが国では、ローカルな地域レベルで基幹となる大学病院などを中心に診療情報のデータ交換ネットワークが構築されている事例があるが、国家レベルのものは無い。データキュレーター機能についても、学会主導の疾患レジストリなど小規模なものしか無く、外部へのデータ提供は共同研究が中心で、データ利活用も限定的であった。2018年に施行された次世代医療基盤法で、認定を受けた民間業者が、連携する医療機関同士でのデータ交換および匿名加工情報を提供するためのデータキュレーター機能を担うことが可能となった。2023年5月に成立した改正により、認定を受けた利用者に対して仮名加工情報の提供が可能となり、ゲ

ノムや希少疾患の情報の取り扱い、NDBとの名寄せ、薬事承認での利用ができるようになった。このように法制度面の土台は整いつつある。

しかし、現時点では認定事業者がカバーする医療機関、自治体が限定的であるため、大規模な健康・医療RWD収集は困難である。現時点で収集の対象になっているのは各地域の基幹病院が中心であり、各患者の一連の受診行動を反映した健康・医療RWD収集ができず、研究開発への二次利用だけでなく、PHRなどを通じた個人の一次利用においても十分とは言えない。

【問題②-3】RWDを用いる臨床研究の倫理審査に要するユーザー側の負担

医学的介入を伴う臨床研究では個別の倫理審査が必須である。一方、介入を伴わない観察研究について、ハーバード大学関連病院では包括的な研究申請を実施済みであれば、個別の研究に対する病院への研究計画の提出や倫理申請は不要である。つまり、包括的な研究申請に沿っている研究であればアイデアが生まれた段階で速やかに研究に着手できる。一方、わが国では、個別の研究に対して研究計画の提出や倫理審査が必要なため、手間や時間を要し、萌芽的なアイデアを試しにやってみようという研究者、データ科学者のチャレンジを阻害し、結果的に新たな知の創出の機会が失われている。

【問題②-4】研究成果の商業化におけるデータ利用の制約

個人情報保護法ならびに次世代医療基盤法に基づいて、診療情報の提供・利用の同意を得る際、使用目的は研究開発が主に想定されている。一方、診療データなどを用いた研究開発の成果として、アルゴリズムの特許を出願、製品化するケースがこれから増加すると思われる。アカデミア単独で社会実装（製品化）まで進めることは困難であるため、開発途中で企業に導出され企業側で研究開発が進むことになるが、この段階で目的が学術研究ではなくなる。要配慮個人情報である医療情報は、原則として本人の同意（オプトイン）を得て取得する。実際に、製品化の段階で企業側が使用した全てのデータについて個別同意を取得し直したケースもあり、企業側にとって大きな負担となっている。次世代医療基盤法についても、研究利用が主に想定されているため、商業・営利目的と見なされうる利用は制限されやすい。商業・営利目的に過度に制限がかかってしまうと企業側の利活用が進まず、データベース維持に必要な収益を確保する上でも不利であると考えられる。

2.1.4 知的基盤構築とその活用における問題

健康・医療RWD利活用を妨げる要因として、医療ビッグデータ・AI解析結果の臨床現場への実装における懸念（エラー、バイアスなど）、生成AI（LLM/基盤モデル）の活用における懸念、倫理的に不適合な利用・悪用への懸念が挙げられる。

【問題③-1】医療ビッグデータ・AI解析結果の臨床現場への実装における懸念（エラー、バイアスなど）

健康・医療RWDには、データの質の不均一性や欠損が避けられない。そのため、AIなどを活用したビッグデータ解析結果や、そこから生まれたプロダクトは、ある程度エラー・誤りを含む。人命に関わる医療においては、エラーを嫌う傾向が特に強いため、臨床現場への実装が進みにくい。AI解析のバイアスやロバスト性も懸念される。新型コロナ禍で、世界中で数多くのAIツールが開発されたが、その多くは、個々の研究者が収集可能な小規模なデータセットを元にしており、社会実装に耐えられるロバスト性とバイアスフリー性が不足し、使い物にならないとの報告が見られる^[7]。

【問題③-2】生成AI（LLM/基盤モデル）の活用における懸念

GPT-4をはじめ、生成AIの進化が著しく、新たな基盤モデルの開発と利活用が急速に進んでいる^[8]。生

成AIで代表的なモデルであるLLMは、大量かつ多様な言語データを用いて訓練され、ファインチューニングなどによりさまざまなタスクに適応できる大規模な深層学習モデルである（例えばChatGPTは、GPTを対話インターフェースに適応させたもの）。言語以外にも、画像を大規模に学習した画像生成AIや、言語、画像、音声などマルチモーダルなデータを学習させた基盤モデル（LMM: Large Multimodal Modelと呼ばれることもある）が開発されている。医療分野でも、米国を中心にLLMの利活用が始まっている。



図2-2 LLM/基盤モデルと用途特化型タスク・アプリの関係

個別の用途に特化したAIモデルをそれぞれ開発するのと比較して、汎用的なLLM/基盤モデルをベースとして用途・分野特化型モデルを構築し、その上で特定の用途のタスク・アプリを開発するアプローチは効率的である。一方で、基盤とするLLM/基盤モデルが使用できなくなった場合、用途・分野特化型モデルや個別タスク・アプリも機能なくなってしまう。医療をはじめとしたライフラインにおいて、海外で開発・運営される基盤モデルに依存することは、経済安全保障上のリスクとなりうる。また、海外で開発・運営される基盤モデルの利用や適応学習を通して、個人の医療情報が流出するおそれもある。

基盤とするLLMの変更により、個別タスク・アプリの動作が変化する可能性がある。医療利用において、動作が変わってしまうのは安全上のリスクとなる。特にソフトウェアの医療機器（SaMD: Software as Medical Device）として承認を受ける場合、現状の制度では承認が無効になる可能性もある。

LLM/基盤モデルをベースとする生成AIが、誤った情報の拡散やハルシネーション（生成AIがもっともらしい虚偽情報生成すること）によるエラーを引き起こす可能性が懸念されており、国際的に生成AIの規制に関する議論が行われている。世界保健機関（WHO）が医療・ヘルスケアへの生成AI実装において、LLMのリスク評価が不十分であるとの懸念を表明している。米国医師会（AMA）は2023年の年次総会において、LLMを含むAIのもたらすメリットや予期せぬ結果について適切な規制案を提案すると公表したが^[9]、わが国では議論があまり進んでいない。

コラム2

医療分野における LLM/ 基盤モデルの実装へ（事例）

昨今、様々な分野・業界に影響をもたらしている LLM/ 基盤モデルであるが、医療分野においても同様に、社会実装事例も生まれつつある。

①医療分野への汎用 LLM の適用：

Microsoft 社は、OpenAI 社の GPT-4 を使用して患者の診断と治療計画を要約・文書化することにより、医療の効率化を図ろうとしており、米国の電子カルテ大手 Epic 社と協力してパイロットスタディを実施中。

②医療分野に特化した LLM 開発

Google 社/DeepMind 社は、医学的質問により正確かつ安全に答えるために、医療分野向けに調整された LLM の Med PaLM2 を開発した。米国医師免許試験形式の問題の MedQA データセットに対し専門家レベルのパフォーマンスを示した。

このほか、言語だけでなく画像なども含めたマルチモーダルな大規模モデルである基盤モデルの医療用途での開発を目指す動きも見られる^[10]。医療分野でも、診療サマリ、レポートなどの文書情報だけでなく、医用画像、心電図などの時系列の信号、オミクスデータなど多種のデータを扱っていることから、医療分野に特化した基盤モデル開発と利活用により医療が大きく進化しうると期待される。

【問題③-3】倫理的に不適合な利用・悪用への懸念

ビッグデータに基づく判定、特に深層学習による判定の根拠を人間が理解することが難しい。技術的には説明可能 AI、因果推論といった対処方法があるが、合理的な層別化なのか不当な差別と見なされるのかは倫理的な検証も不可欠である^[11]。研究倫理審査で目的の正当性については評価ができるものの、倫理的に不適合な結果を完全に排除することはできない。匿名加工情報であっても、多数の情報を名寄せすると個人を推定できる確率は高まり、プライバシー侵害など悪用される可能性が否定できない。さらには、セキュリティ対策を万全にしてもさまざまな要因でデータ漏洩は発生しうる。情報収集や研究開始時に規制をかけたとしても、データ利活用によって個人が不利益を被るような、倫理的に不適合な事象が起こる懸念を完全に排除することは容易ではない。

2.2 提案が実施された時の効果

2.2.1 社会・経済的な効果

健康・医療RWD利活用基盤を構築することで期待される最も重要な効果は、健康寿命の延伸である。加えて、医療資源配分の最適化、臨床現場の環境改善、医療産業の活性化などにもつながりうる。結果として、わが国の医療保障制度の根幹をなす国民皆保険制度の持続性、ひいては医療・ヘルスケアシステム全体の持続性の確保につながる。

●健康寿命の延伸

健康・医療RWD利活用基盤から治療履歴とアウトカムの情報を抽出し解析することで、医薬品や医療機器のリアルワールドにおける安全性や有効性の評価が可能となり、より安全性・有効性の高い医療技術が選択され、健康寿命の延伸につながる。診療情報をPHRで閲覧可能とすることで、国民1人1人が自分自身の状態を把握でき健康増進の意識が高まり、国民1人1人の行動変容が加速すると期待される。また、心疾患（特にペースメーカー使用の場合）や妊娠などで経過観察が必要な個人において、それぞれの医療ニーズに固有なデータ種までPHRを介して閲覧可能になることで、より効果的な疾病・健康状態管理につながる。

●医療資源配分の最適化

高齢化に伴う国民医療費の高騰と医療ニーズの増大・変化に対応するため、医療資源配分の最適化は極めて重要な課題であり、健康・医療RWD利活用基盤の貢献は大きい。医療技術の費用対効果が注目を集めるが、健康・医療RWD利活用基盤を構築することで、膨大な種類の既存の医療技術、およびこれから登場する新規医療技術の費用対効果の評価が進むものと期待される。医療提供体制の観点では、健康・医療RWD利活用基盤を通じて自治体・地域などの単位で医療資源の配分実績を把握し、さらに自治体の人口動態などの社会的データも踏まえて分析することで、医療ニーズの予測ときめ細かな医療・介護・保健政策の実現が期待される。

●臨床現場の環境改善

臨床現場でのデータ入力・分析支援、AI診断支援技術などが実装されることで、電子カルテなどへのデータ入力に費やされる労力が軽減する。さらに、高度化・複雑化の一途を辿る医学知識を医療従事者がより効果的に習得することが可能となり、患者への医療提供の最適化が実現すると期待される。昨今、医療従事者の働き方改革が深刻な課題となる中で、解決につながる有効な手立てとなりうる。

●創薬/医療関連産業の活性化

創薬/医療関連企業が、健康・医療RWD利活用基盤によって医薬品・医療機器の市販後調査をより効率的かつ網羅的に行うだけでなく、製品のアウトカムデータを用いた治験（対照群を置くのが困難な疾病、適用拡大・リポジショニングなど）の実施も可能となる。患者群をより細やか、的確に層別化・個別化した治療法開発が可能となり、医薬品・医療機器（特に国内メーカー）の活性化が期待される。また、健康・医療RWD利活用基盤および生成AI活用は、電子カルテやPHRなど医療情報産業の活性化にも繋がる。

わが国では、MRIやCT、内視鏡などの画像診断機器を開発する世界有数の企業が複数存在し、そうした企業が開発した優れた製品が国内に幅広く配備され、検査数も多い。画像と言語の医療版基盤モデル開発における、わが国の強みとなる。臨床現場以外においても、患者同士の情報共有・コミュニケーションのサポート、モバイル医療機器を活用した遠隔・非常時の診断支援、など幅広い活用が期待される。

2.2.2 科学技術上の効果

●健康・医療 RWD を活用した医学・医療研究の活性化

従来、健康・医療 RWD を用いた研究は、データ収集にかかる負担があまりにも大きく、既存のいくつかのデータベースも多く研究者のニーズに合致した使い勝手の良い基盤とはいえなかった。従って、実臨床に従事しない研究者が健康・医療 RWD を用いた研究を実施することは、ハードルが非常に高い。健康・医療 RWD 利活用基盤により、臨床医だけでなく幅広い分野の研究者にとって健康・医療 RWD へのアクセスの敷居が下がることで、多様な切り口の研究が活発に推進され、新たな知識が次々と生み出されると期待される。

健康・医療 RWD 利活用基盤は、主に診療情報を中心とした公的医療・介護・保健データが対象である。これらの情報の収集・集約を促進するために開発した技術・ツールや開発過程で得られた知見は、看護師の処置に関するデータ、訪問診療データ、ウェアラブルデータなどの他の種類の健康・医療 RWD 収集・集約にも展開できる可能性があり、多様な種類のデータ利活用の促進が期待される。医療機関内外での健康・医療 RWD の収集・蓄積が進めば、ある個人のライフコースでの追跡とそのデータを使った健康状態予測シミュレーション/デジタルツインが実現する可能性がある。また、収集される多様なデータを活用することで、認知症、精神疾患といった医療機関外での長期にわたる RWD 収集が必要となる疾病のデータ蓄積が可能となり、治療・緩和や予防、そのための体制・制度整備に向けた医学的・社会的研究が進展することが期待される。

●ヒトを対象とした研究の加速

既存のコホートやバイオバンクなどの研究用データベースにおいて、ゲノム、オミクスなど大量のヒトのデータが集約されている。しかし、それらデータは遺伝因子もしくは環境因子に関連するデータが主であり、病態の詳細や治療結果・アウトカムを示す臨床データが含まれていないか、含まれている場合も限定的であり、医学的に有益な知の創出に向けた解析は困難である。研究用データベースと、健康・医療 RWD 利活用基盤で得られるリアルワールドでの診療情報、薬剤情報やエンドポイント（死亡データ）を組み合わせることで、ゲノムの違いの意味をはじめヒトの医学的理解が深まる。こうして得られた新たな知見に基づき、新たな診断・治療技術開発のシーズが生まれる可能性がある。

ヒトを対象とした健康・医療 RWD 収集は、技術面、コスト面、倫理面の観点から制限がある。そのため、通常の臨床研究においては、何らかの仮説に基づいてヒトを対象に健康・医療 RWD の収集を開始する。一方、その仮説構築のために利用可能な RWD がなく、仮説構築だけの目的で多種多様なヒトのデータを集めることも難しい、というジレンマがあった。健康・医療 RWD 利活用基盤により集約されるデータを利活用することで、ヒト研究における新たな仮説構築が加速しうる。

3 | 具体的な提案の内容

3.1 提案の全体像

本戦略プロポーザルは、わが国における健康・医療RWDの利活用を加速するための方策を提案するものである。具体的には次の3つを提案する。

【提案①】健康・医療RWD収集を加速する技術の開発

【提案②】一次利用と二次利用を前提とした健康・医療RWD流通の仕組みや制度の整備

【提案③】医療版基盤モデルの開発および社会実装のための法規制の整備

生成AIの活用については、米国を中心として海外で急速に開発・実装が進んでいることから、喫緊の課題である。OpenAI社のGPTをはじめ海外で開発された基盤モデルが先行している状況を踏まえると、商業的観点からは海外の基盤モデルを使うことが合理的である場合が多いと考えられる。一方、2.1.4で述べた通り、特に医療分野での活用については経済安全保障リスクや個人の医療情報流出の懸念が大きいことから、本戦略プロポーザルではその観点を重視して、国産の医療版LLM/基盤モデルの開発を提案する。LLM/基盤モデルの開発には大規模な健康・医療RWDが不可欠であり、その観点からも【提案①】【提案②】による健康・医療RWD利活用基盤の整備は重要である。

関連する国内の主な取り組みとして、戦略的イノベーション創造プログラム（SIP）第三期で採択された「統合型ヘルスケアシステムの構築」プロジェクトがある。これは医療情報による知識発見の促進を目指すものであり、2023年度から始動している（詳細は付録2.3参照）。この中では重要な研究開発テーマが数多く進められており、インパクトの大きな成果を挙げると期待される一方、同プロジェクトが当初発足時に掲げていた取り組みと、本プロポーザルが提案するゴールとにはギャップもある。最新の技術動向も踏まえた研究開発、および法制度や社会受容に関する取り組みなど、わが国でなすべき事は多い。同プロジェクトの進捗も踏まえて、本章で掲げる3つの提案を推進することが重要である。

3.2 個々の提案の内容

3.2.1 【提案①】健康・医療RWD収集を加速する技術の開発

臨床現場における負担の大きさが、診療情報の収集が進みづらい主な要因である。RWD収集・構造化・統合のための技術やツールを開発し、データ蓄積を加速しようとする取り組みがこれまでなされてきたが、臨床現場に幅広く実装可能な汎用的なツール開発は進んでいない。LLMは、現時点では開発の実現性や必要なデータ規模、開発期間など未知数の部分が多いが、臨床現場において必要な情報を構造化された状態で収集かつ臨床現場の負担を軽減できる、さらに将来的には診療支援の基盤となりうる技術である。そこで、従来の健康・医療RWD収集改善のアプローチである【提案①-2】の電子カルテ入力用のテンプレート整備と並行して、【提案①-3】において医療分野に適応したLLM開発とそれをベースとした診療記録作成支援ツールの開発を提案する。【提案①-1】は、その前提となる、HL7 FHIR準拠のデータ交換規格整備と対応した電子カルテの開発、傷病名・医薬品・検査などのコード整備の提案である。

【提案①-1】電子カルテのデータ交換規格対応および各種コードの整備

HL7 FHIR 準拠のデータ交換方式の整備、HL7 FHIR 準拠の電子カルテ機能拡充、日本においてFHIRを使用する場合の共通ルール整備、厚労省標準にある医療文書や項目についてFHIRで記述するためのユースケースごとの仕様の整備、これらを実装した主に中小医療機関向けの電子カルテの開発を推進する。

傷病名・医薬品・検査コードについては、標準コード・マスターの標準規格文書への採用と格納の要件化、複数の医療機関で行われる可能性が高い検査などの優先度の高い項目のコード化や、コードの周知徹底に向けた取り組みを推進する。

【提案①-2】臨床現場において医療従事者が効率的に入力可能な電子カルテのテンプレート整備

入力する項目やベースとなる電子カルテシステムが異なる場合、それぞれに対応したテンプレートを作成する必要がある、いかに効率よくテンプレートを整備するかが課題となる。テンプレートに収載する項目を仕様として定めるため、一般的な疾患・症例に対応する標準的なデータ項目の設定を実施する。各診療科に関連する学会などと協議の上で決める必要がある。全ての疾患・症状について設定するのは困難であると考えられるため、まずは生活習慣病などの一般的な疾病を対象に、データ収集するミニマム項目を検討する。電子カルテベンダーによらないテンプレート開発も行う。例えば、国立高度専門医療センター（いわゆるナショナルセンター）6拠点でのデータ収集・共有用に開発したJapanese Standard Platform for EHRs (JASPEHR) は、共通テンプレートから各ベンダーの電子カルテにおけるHL7 FHIR対応テンプレートに変換する仕組みである。こうした仕組みを活用し、テンプレート開発を効率的に実施する。

【提案①-3】国内開発のLLMを活用した診療記録作成支援ツールの開発

LLMをゼロから開発する場合、大規模な学習データや大量の計算資源が必要となりハードルが非常に高い。そこで、汎用の日本語LLMをベースとして、追学習などにより医療分野向けに適応させることで、より効率的に医療版LLMの開発を推進する。医療分野への適応についても、質の高い医療向け学習データを用意することで必要なデータ量の削減を目指す。教科書レベルの質の高い学習データを用いることで少ないパラメータ数、データ量でも同等以上の性能を発揮するLLM開発が可能、という研究報告^[12]があるため、同様のアプローチにチャレンジする。質の高い学習データとしては、医療の専門用語の辞書である医療オントロジーが挙げられる。例えば、国内で開発されてきたJ-CaseMap辞書や、国際的な医療用語辞書SNOMED-CTの翻訳版などがある。一方、国産で性能が高いLLMは開発中の段階であり、分野適応などの手法についても国内に知見・ノウハウがまだ蓄積されていない。国立情報学研究所（NII）を中心に、日本語に強いGTP-3レベルの汎用LLM構築が始まっており、それと合わせて国内の自然言語処理・計算機システム研究者が参加し研究成果・知見を共有するLLM勉強会が開催されている。医療分野への適用についても議論に挙がっていることから、こうした国内の動きと連動して推進することが望ましい。

構築した医療版LLMを基盤として、臨床現場の負担軽減に寄与する診療記録作成支援ツールを開発する。例えば、ChatGPTのような医療従事者との対話型のデータ入出力ツールが開発・実装されることで、支援診療記録の下書きなどの業務効率化と、病名などの用語統一、コードの自動付与などが進むと考えられる。診療記録作成だけでなく、患者のデータ入力の支援により診療の効率化や質の向上につながるAI問診票^[13]、患者の診療情報を一元的に表示することで診療中の患者の情報検索に要する時間・手間を劇的に減らすことが期待されるダッシュボード^[14]など、臨床現場にとって有用なツール開発も進める。医療版LLMをそのまま利用するだけでは実現できない機能もあると想定されるため、【提案③-2】で後述するレベル分けを元に、開発の中で蓄積される知見を基に、実現に向けた手段を検討する。

3.2.2 【提案②】一次利用と二次利用を前提とした健康・医療RWD流通の仕組み・制度の整備

二次利用の用途に応じて、悉皆性の高いデータと、高品質なデータが必要になることから、それらのデータ流通の促進に向けた仕組み、より幅広いデータとつなぐ仕組み、民間事業者による商業化を想定したデータ利用同意プロセスと審査体制の検討・整備を提案する。

【提案②-1】の悉皆性の高いデータは、一次利用においても医療機関を跨いで確認したり、災害時などに参照したりする基本的なデータセットである。行政サービスのデータインフラを担うデジタル庁の計画でも健康・医療・介護は準公共サービスと位置づけられていることも踏まえると、データ流通の仕組みや規格は公的に整備・運用すべきである。一方、【提案②-2】の高品質なデータは、基礎研究のみならず創薬や医療関連技術の開発などに資するものであることから、公的に整備しつつも運用は民間に委ねるのが望ましいと考える。本提案は、こうした長期的なあるべき姿に近づくための次のステップであり、提案実施後も継続的な改善が必要となる。

【提案②-1】悉皆性の高いデータプラットフォームの整備・拡充

NDBをベースに、想定する用途（医療費、医療資源のマクロ分析）に適した機能拡充を実施する。データ漏洩のリスクを低減しつつ、データベース上の必要十分なデータを簡便にモニタリング・解析できるツール開発と環境構築を進める。具体的には、NDBヘリモートでアクセスし（Data Visiting）、クラウド環境上でデータ解析を行い、解析結果だけを取り出す解析ツール・環境の整備を行う。データ解析をNDB上で完結させることで、さまざまな解析を迅速に試行錯誤することができる。例えば英国のGenomics Englandでは日本からもリモートで解析できる環境が実装されており、東北メディカルメガバンクでも、制約は大きいものの同様のシステムが稼働していることから、これら既存のシステムも参考に解析環境を整備する。NDBにおいて、改正次世代医療基盤法を根拠として診療情報との名寄せが可能となったため、同一の医療IDを用いたNDB内での名寄せの精度向上などの研究も推進する。

【提案②-2】診療情報データベース間で診療情報を連携する規格・仕組み

改正次世代医療基盤法では、認定事業者間でのデータ連携の仕組みはなく、各認定事業者が構築した医療機関ネットワーク内で収集した診療情報のみを加工・提供するため、提供できるデータの量を増やすことは容易でない。また、次世代医療基盤法/改正次世代医療基盤法では、情報取扱事業者（データ交換ネットワーク機能）と加工事業者（データキュレーター機能）がセットで認定されている。一方、一次利用の観点では、日常的に患者がアクセスしうる範囲で地域を限定した医療機関ネットワークが効率的であるが、医療研究などへの二次利用では、各地域の基幹病院を中心に広範なネットワークで大規模にデータが集まることが望ましいため、一次利用・二次利用を同じ事業者でカバーするのは合理性が低い。

まず取り組むべきことは、特定の加工事業者が複数の情報取扱事業者の診療情報を取り扱うことを認め、より多様なデータ（例えば、特定地域の大小の医療機関から集約した診療情報）や大規模なデータを加工事業者が提供できる仕組みの実現である。一次利用に必要な診療情報を共有できるよう、情報取扱事業者間でのデータ交換（Network of networks）が可能な基盤整備も重要である。情報取扱事業者がデータ交換ネットワーク機能を実装する上で必要な、標準仕様の策定を進める。これは【提案②-3】で述べる診療情報をPHRに提供するための基盤にもなりうる。

【提案②-3】本人同意に基づきPHRや研究用データベースに情報を提供する仕組みの整備

一次利用においては、PHRを通した本人の診療情報の閲覧・活用が重要である。コホート・バイオバンクなどに紐付いた研究用データベースで収集したゲノムデータ、オミクスデータや生活・環境情報を個別化医療

や創薬に活用するためには、患者の状態や治療アウトカムなどの診療情報と統合したデータ解析が重要である。こうした利活用を可能とするデータ提供の仕組みを整備する。

具体的には、マイナンバーカード上の電子証明の機能（署名用電子証明書、利用者証明用電子証明書）を利用し、本人同意と本人証明を行なった上でデータ提供を可能とする仕組みを導入する。コホート・バイオバンクなどに紐付いた研究用データベースは、研究目的でのデータ収集がなされているため、データ提供先の条件を定めた上でオプトアウトに含めるなど、より包括的な同意プロセスにできないか検討する。

これらは、技術的には厚生労働省が推進する「オンライン資格確認」の仕組みを用いることで可能である。何に対して同意を求めるかについては、例えば、データ種類を選択させるか（PHR向けは医療機関の同意、研究用データベース向けは本人の同意）、日々追加されるデータの提供において都度同意を求めるか、データ保持の期限を設けるか（研究用データベース向け）、などの観点からの検討が必要である。

【提案②-4】民間利用を前提とした審査体制と同意プロセスの整備

米国や欧州では、健康・医療RWDの第三者提供における制限に関して、学術機関か民間企業などの属性ではなく、利用目的で規定している。医療機関などからデータキュレーターへ情報提供する際の規制、すなわち入口における規制ではなく、出口において、データキュレーターがデータ提供・データへのアクセスを許可する際に利活用の目的に応じて規制する、という枠組みである。わが国における健康・医療RWD流通でも同様の枠組みの整備を検討する。検討すべき点として、一元化もしくは地域などでの集約化（審査のノウハウ蓄積とバラツキの抑制）、利用主体に合わせた柔軟な基準設定（学術機関に対してはある程度包括的な目的での利用を認め、民間企業に対しては特定の目的に絞った利用に限るなど）、利用目的に適した解析に使用するデータ加工法の判断（顕名情報/仮名加工情報/匿名加工情報）、データ利用条件の標準化（GA4GH:The Global Alliance for Genomics and Healthが提唱するDUOコード^[15]などを活用し円滑で公平な審査へ）などが挙げられる。医薬品・医療機器開発、特に承認申請などへの健康・医療RWD利用は公共の利益に資するとも言え、純粋な学術研究目的からは外れるが原則として個別同意を必須としない運用が望ましい。しかし、商業利用・営利目的と解釈しうる利用に対して忌避的な意見も存在すると思われるため、想定される商業利用ケースを規定し、商業利用の可能性のある旨を健康・医療RWD収集時に告知した上で、個人にオプトアウトでの同意を求める、などの同意プロセスを次世代医療基盤法などに盛り込むことを検討すべきである。

3.2.3 【提案③】医療版基盤モデルの開発および実装のための法規制の整備

【提案②】を通じて構築する健康・医療RWD収集・流通の仕組みを最大限に活用し、医療の知的基盤としての基盤モデルを開発する。わが国の強みである画像診断機器を生かし、言語と並んで臨床現場で活用される画像と言語のマルチモーダルな基盤モデルを開発する。そのために不可欠な大規模な健康・医療RWDを蓄積するため、まずは診療情報と画像情報を中心に蓄積するデータウェアハウスの構築を行う。基盤モデルやそれをベースとした医療AIの利用をはじめとして、健康・医療RWDの利活用にはさまざまな懸念もあるため、社会受容に向けた議論と必要な法的・倫理的基盤の整備に取り組む。

【提案③-1】基盤モデル開発に向けたデータウェアハウス構築

【提案①】【提案②】で構築した健康・医療RWD基盤を活用して、基盤モデル開発での利用を想定し、診断・検査結果の文書、医用画像データ、心電図をはじめとした検査結果など、さまざまな種類の健康・医療RWDを蓄積するためのデータウェアハウスを構築する。基盤モデルの開発だけでなく特定用途への適応の学習や個別用途のAI開発などにも活用できるよう、【提案②】も参考に、幅広い包括的な用途を想定してデータアクセスの環境や審査体制を整備する。大規模な健康・医療RWDの需要を生み出されることになるため、

【提案②-2】の施策に加えて健康・医療 RWD 流通プレーヤーを増やす呼び水にもなると期待される。

【提案③-2】医療版基盤モデルの開発と社会実装

データウェアハウスに蓄積される大規模データを活用し、【提案①-1】で開発した医療版 LLM の改良、および LLM をベースとして医用画像と言語のマルチモーダルな基盤モデルを開発する。大規模な医用画像データの蓄積が必要となるため、新たに収集する健康・医療 RWD だけではなく、過去に蓄積された画像データベースに含まれる放射線画像や病理画像の活用、因果関係や数値シミュレーションを踏まえた学習用画像生成といった検討も行う。構築した医療版基盤モデルをベースに、放射線画像や病理画像からの検査サマリ作成、画像と所見を踏まえた疾病候補提示、などの臨床現場支援ツールを開発する。【提案①-3】と同様に、NII を中心とした LLM 国内開発と連動させ、知見や経験を共有しながら進める。特に、今後国内で言語と画像の汎用的な基盤モデル開発が進むのであれば、その分野適応により開発する方が早い可能性があるため、動向を踏まえて開発方法を判断する。

因果や時系列も踏まえた総合診断支援、学術論文等から得られる最新の知見を反映した診断支援・個別化医療支援など、本提案で開発する医療版 LLM や医療版基盤モデル（医用画像＋言語）をそのまま活用するだけでは実現できないものの、実現すると医療・ヘルスケアの効率や質の向上にインパクトをもたらさうる項目が数多くある。これらの実現には、中長期での技術開発やデータ蓄積、実証の積み重ねなどが必要となる一方、基盤モデルの知見が不十分である現段階ではその実現手段を確定することはできず、研究開発を進める中で最適な手法が見えてくると考えられる。さらなる改良や実装拡大に向けては、有識者の議論をベースに実現したいことのレベル分けを設定し（自動運転レベル分けのアナロジー）、それをマイルストーンとした中長期的な研究開発が重要である。

表 3-1 医療版 LLM・基盤モデルのレベル分け（イメージ）

	実現したいこと	推進すべきテーマ
医療機関	・ 診療記録作成の補助 ➢ 診療記録の下書き、用語統一・自動コード化など ・ 診療記録に基づく疾病候補提示	・ 汎用 LLM をベースとする医療版 LLM の開発 ・ 医療版 LLM の API 利用 or 外部ツールの連携
	・ 画像検査記録作成の補助 ➢ 画像検査サマリの下書きなど ・ 診療記録と画像検査に基づく疾病候補提示	・ 言語と画像のマルチモーダルな医療版基盤モデルの開発 ・ 医療版基盤モデルの API 利用 or 外部ツール連携
	・ 診療記録、画像検査、経過情報などに基づく診断支援	・ 医療版基盤モデルに加え、因果や時系列を検証する方法論の開発 ➢ Chain of Thought（生成 AI への依頼の例示や手順分解）での検証など
	・ 最新の臨床研究の知見に基づく診断支援、個別化医療支援	・ ゲノム/オミクス、論文 Fig などの学習データに基づく、独自の基盤モデルの開発 ➢ 理研など開発中の科学技術基盤モデルとの連携
	・ 患者/消費者向けの対話型の健康・医療情報提供	・ 臨床現場での活用事例を踏まえた、医療版基盤モデルの挙動の制御、管理体制の検討
一般		

【提案③-3】臨床現場での LLM/基盤モデルベースの医療 AI 活用における規制の検討

広義の医療 AI は、診断や外科的治療、PHR も活用したデジタル治療や健康増進アプリなど、さまざまな活用形態が想定されるが、ここでは臨床現場における AI 医療機器としての活用を想定する。AI 医療機器の規制については、FDA などの規制当局を中心に議論や導入が進む。FDA 承認済の AI を用いた医療機器は 500 を超えるが（2022 年 10 月時点）、これらは全てモデルが固定されたものであり、市販後の再学習は許されていない。学習し続けるのが AI の強みであることを踏まえると、健康・医療 RWD 利活用基盤を通じて随時集約される健康・医療 RWD の利活用において、学習し変化し続ける AI モデルが望ましいが、海外でも法

規制が追いついていない。わが国では、2020年9月施行の薬機法改正で市販後の再学習が可能になったが、何をもって一部変更とするか、変更によりどれくらいの影響が出るのか、データの質をどう担保するかといった点について、実証実験が行われていない。特に、医療版LLM/基盤モデルの変更に際して、それらを基盤とした用途適応型のタスク・アプリの承認をどのように考えるかは、医療版LLM/基盤モデルの開発と並行して議論と検証が不可欠である。また、医療版LLM/基盤モデルに起因する予期せぬ結果やエラーについては、LLM/基盤モデルをベースとしたツールの実装を検討する中で検証し、医療AI規制の議論に反映させる。

【提案①-3】で開発する診療現場支援ツールなどをモデルケースとして、具体的な規制の整備に向けた検討・検証を行う。アウトカム向上や臨床現場の負担軽減などの有効性が大きいことが確認できた場合、単独での安全性・リスクのみに着目するのではなく、人間と同様にビッグデータ・AIもエラーを起こしうることを前提として、フェイルセーフなどヒューマンエラー対策と同様の考え方に基づいた規制・運用方法を検討する。患者側の期待や懸念についても規制や現場での利用プロセスなどに反映させるため、PPIPなどを活用して具体的な事例における期待や懸念を調査する。

さらに、医療従事者ではなく患者や一般消費者がツールを利用するケースを想定し、医師が責任をもつ診療行為との責任の所在の違いを加味した上で、ツール開発の規制のあり方を検討する。

【提案③-4】倫理的に不適合な医療ビッグデータ・AI活用の防止

倫理的に不適合な利用・悪用への懸念がデータ活用の妨げになることを防ぐための手段として、不利益を生じさせる行為そのものの防止や、不利益を被った時の救済策の整備が考えられる。そうした方策の一環として、健康・医療データの中で特に悪用の懸念が高い遺伝情報（ゲノム情報、家族病歴など）による差別を禁止する法律や規制が、欧米各国で導入されている^{[16][17]}。

わが国でも、2023年6月にゲノム医療推進法が成立し、遺伝子差別の防止が明記された。今後は具体的な計画策定に向けた議論が必要である。例えば米国で、GINA（Genetic Information Nondiscrimination Act）が出来たことで人々に問題が意識され関連訴訟が増えたという報告^[18]があることから、法規制導入による効果や副作用が顕在している海外の状況を調査し、それをベースに国内での影響と対応を議論する。

4 | 提案実施上の推進方法及び時間軸

4.1 推進方法

健康・医療RWD利活用基盤の整備と幅広い利活用に関わるステークホルダーは産・学・官、そして健康・医療RWDの提供元かつ利活用における最終的な受益者になるべき個人・市民と多岐にわたることから、いかにステークホルダー間で連携し、同じ方向を目指して推進できるかが重要である。また、3章で説明した各提案は関連するステークホルダーが異なるため、それぞれについて必要なステークホルダーを巻き込みつつ推進することが望ましい。生成AIの開発と実装は日進月歩であることから、最初の段階で中長期の方針を決めることはできず、段階的に進む方向を決定する必要がある。これら状況を踏まえ、本戦略プロポーザルでは、【提案③-2】で設定する医療版LLM・基盤モデルのレベル分けを基に、段階的に中長期でのマイルストーンを設定し、司令塔の下で各提案の進捗や提案間の連携を統合的に管理する、という推進方法が適切であると考えられる。各提案は、内閣府SIP「統合型ヘルスケアシステムの構築」など関連するプロジェクトと連携しつつ、各提案のステークホルダー間で協業して実施する。

医学、データ科学、法学、医療/データ倫理など、学術分野を横断した検討となることから、健康・医療RWD利活用基盤構築後に運用を担う人材を確保することも念頭に、分野を跨ぐ人材の育成が重要である。

4.1.1 推進体制

3章で述べた【提案①】【提案②】【提案③】について、各提案を推進する中核となる産学官民のステークホルダーおよび関連する既存プロジェクトを挙げる。本章では、各項目の具体的な担い手となりうる組織・団体名などを明示しているが、実際にプロジェクトを開始することとなった段階で、関係者間での議論の上で最新状況も踏まえた最適なメンバーで推進されることが望ましい。

【提案①】 これまでも電子カルテ入力補助テンプレートや診療情報の構造化技術などの開発を主導してきたアカデミア、例えば医療情報学会が中心となって推進することが望ましい。加えて、医療版LLM活用の前提となる電子カルテの導入率向上や収集項目の標準化・コード化を進める厚生労働省、実際に電子カルテを開発・供給するベンダーもボードメンバーとなって進める。【提案①-1】【提案①-2】に関連するプロジェクトとして、SIP「統合型ヘルスケアシステムの構築」、厚生労働省が中心で進める「データヘルス改革」や「医療DX」があることから、進捗状況を相互に共有し、本提案のマイルストーンを達成できるよう連携や分担を行う。【提案①-3】医療版LLMの開発では、NIIが中心となって進めるLLM勉強会と連動して進めることが望ましい。

【提案②】 NDBを所管する厚生労働省を中心に悉皆性の高い健康・医療RWD流通の仕組みを構築し、分析機能についてはSIP「統合型ヘルスケアシステムの構築」とも連携して推進する。質の高い健康・医療RWD流通の仕組みの構築は、次世代医療基盤法を所管する内閣府を中心に推進し、技術的な面では医療情報学会と連携して推進する。これらについて、現行の認定事業者やその他既存の診療情報データベース、主な二次利用者である製薬企業、研究機関などの意見を踏まえて進めるのが望ましい。標準仕様の策定では、医療情報学会などを中心に策定しているHL7 FHIR記述仕様をベースに、一次利用を担う医療機関やPHR事業者のニーズを踏まえる必要がある。PHRや研究用データベースなどへのデータ提供は、厚生労働省と内閣府を中心に、マイナンバーカードの活用を想定するため総務省やデジタル庁とも連携を図る必要がある。ウェアラブル機器の開発企業との協業に加え、データ項目や形式の標準化を視野に、PHRの推進を進める業界団体

(PHR普及推進協議会、PHRサービス事業協会、PHR協会)との連携が望ましい。研究用データベースについては、既存のバイオバンク・コホートと連携が必要である。審査体制や同意プロセスについては、医療・医学と医療倫理、法学の有識者と行政が協議を重ねる「人を対象とする生命科学・医学系研究に関する倫理指針」に個人情報扱いも含めた指針が定められている。省庁横断であり個人情報保護委員会などとも連携していることから、まずはこの指針の改定の一環で有識者間の議論を行い、必要に応じて次世代医療基盤法などに反映させる。検討にあたっては、医療・医学、医療倫理、法学に加えて、情報工学、民間(製薬企業、デジタルヘルス事業者など)、患者会など、多様なステークホルダーの参画が望ましい。

【提案③】4つのサブ提案のうち【提案③-1】【提案③-2】は、基本的には【提案①】と同様の体制で推進すべきである。データウェアハウスについては、【提案②-2】の改正次世代医療基盤法におけるデータキュレーター機能を有する機関の設立が望ましい。各データベースで蓄積した医用画像データの活用については、各データベース構築を主導する学会と協業しつつ、必要となるデータキュレーションやアノテーションの研究者の協力も仰ぐ必要がある。基盤モデルをベースとしたツール開発・実装に当たっては、医用画像を利用する医療機器やヘルスケア機器メーカー、電子カルテ以外の医療情報システムベンダーなどの参画を促す。医療AIの規制に関しては、規制当局であるPMDAや厚生労働省を中心に議論を進める。AI医療機器を実際に開発するメーカーの取り組みや本戦略プロポーザルにおける医療版LLM/基盤モデルをベースとしたツール開発の取り組みを共有し、医療従事者も交えて、臨床現場に導入する上で安全性を担保しつつも開発を停滞させないような規制・ガイダンスの議論が必要である。遺伝差別の防止については、医療制度だけではなく、保険、雇用など関係する分野が多岐に渡り、国による仕組みや習慣の違いが影響するため、それら広範な情報をアカデミアに加えて民間事業者も協力して収集・分析した結果を基に、【提案②-4】と同様、倫理指針の議論の場を中心に検討を進める。

以上で述べたように、各提案でステークホルダーや中心となるプレーヤー、省庁が異なる。各プレーヤーや省庁でそれぞれ目指す方向がある一方で、健康・医療RWD利活用基盤により収集・流通・知の創出と活用という一連の流れを加速するという目的を達成するには、その方向性を本戦略プロポーザル全体として目指すものと整合させなくてはならない。また各提案を進める上では、2章で述べた既存の各問題の解決が基本的な方針となるが、問題解決だけでは多様なプレーヤーが連携・協働する上で訴求力として十分とは言えず、2.2の期待される効果を基に、多様なプレーヤーが共感できるような中長期的な目指す姿をプロジェクトとして明示する必要がある。

また4章冒頭に記載の通り、段階的にマイルストーンを設定する場合、各提案の実施のタイミング、レベル感をマイルストーン達成から逆算して設定する必要がある。これらを鑑みて、本戦略プロポーザルでは、省庁横断の司令塔機能の下で、各提案の進捗や提案間の連携を統合的に管理する、といった推進方法を採用するのが望ましい。

4.1.2 分野間連携・人材育成

健康・医療RWD利活用基盤の構築と幅広い利活用を推進するためには、医療・医学と情報科学・データ科学、法学、倫理などの異分野連携が重要であり、そのために複数分野に精通した人材育成が必要となる。例えば、医療と法学の連携について、臨床と法曹の実務的なレベルでは、従来は医療訴訟など限られたトピックでの連携に留まっていた。医師と弁護士のダブルライセンスを持つ人材が少しずつ増えてきていることから、こうした人材を軸に両分野の共通理解を深めつつ議論を重ねることが期待される。【提案②】で述べた倫理審査を円滑に運用していく上でも、こうした人材は重要である。

医療版LLM・基盤モデルの構築においては、医学と情報科学・データ科学の研究者の協働・連携が不可欠となる。そのためには、双方の分野の研究者を訴求するための科学的な意義を提示する必要がある。例え

ば、医学に対しては基礎医学と臨床情報が一体となった新たなヒト科学（臨床データ・アウトカムからゲノムの意味を探るなど）、情報科学・データ科学に対しては国内におけるLLM・基盤モデルの知見蓄積や機序解明、といったポイントが考えられる。

健康・医療RWD利活用基盤が運用段階に入った段階で、中長期的な人材育成にも取り組む必要がある。臨床現場へのAIの実装が今後ますます進んでいくと考えられ、医療系人材を教育する際、AIやデータ科学を必須項目とすることが重要である。これから、AI診断支援の臨床現場への実装が進むと、医療従事者と患者がAI診断支援結果を参照しながら一緒に意思決定を行う（Shared-decision making）ケースが増えると考えられる。ゲノム医療における遺伝カウンセラーに相当するような人材育成が必要になると考えられる。

4.1.3 社会受容・社会との協働

健康・医療RWD利活用の施策が社会に受容されるためには、個人情報保護とのバランスをいかにとるかが課題となる。【提案②-4】で実施する法的・倫理的基盤の整備に加えて、健康・医療RWDの提供元となる個人・市民の理解醸成を進めていく必要がある。医学研究への健康・医療RWDの利活用に対して、個人・市民の理解は比較的高いとの報告がある（詳細は付録2.4を参照）。一方、民間企業が医薬品・医療機器開発と社会実装を主導する際に健康・医療RWDを使用するケースも増えると考えられるため、個人・市民の受け止めに留意しつつデータ利用の同意プロセスを慎重に検討し進める必要がある。医薬品・医療機器開発が医療の改善や医療産業の発展、ひいては個人・市民の健康につながることを、丁寧に説明をしていくことが求められる。オプトアウトによる選択権を設定し、その旨を明確に伝えることも重要である。

構築した健康・医療RWD基盤が実際に機能する上では、社会との協働、特に医療・ヘルスケアサービス等の提供を通じて健康・医療RWDを収集する医療機関や自治体（図2-2における『一次データ収集者』）の協力が不可欠となる。特に診療情報については、大規模なグループ病院や保険者が存在する米国や、国が管理する欧州と異なり、日本で大規模なデータベースを構築するには、各機関でサイロ化した健康・医療RWDの個人情報や仮名加工情報の形での第三者提供が避けられない。そこを円滑に進める上では、【提案①】におけるデータ収集に係る現場の負担軽減に加え、データ提供への適切なインセンティブの設計が求められる。また、前述の通り医学研究目的のデータ利活用に対して個人・市民の理解は比較的高い一方で、医療機関・医療従事者側が過度に個人情報提供のリスクを避けている可能性が指摘されていることから、一次データ収集者が負うリスクの軽減が必要である。【提案③-4】の倫理的に不適合なデータ活用の防止は、データ提供者に提供したデータの悪用に対する責任を負わせない意味もあることから、そうした取り組みと合わせて理解醸成を図る。

医療AIやLLM/基盤モデルの活用においては、それらを活用する臨床現場だけでなく、対象となる個人・市民の懸念を具体的に把握する必要がある。そのためには、医療AIや遺伝情報利活用の意義の明確化、想定されるリスクの評価、法規制により想定される影響などの評価を行い、個人・市民への説明・対話を行った上で懸念事項をリストアップする必要がある。これは、PIIPなど適切な検証手法を用いる。生成AI規制に関する国際的な議論の方向性と個人・市民に与える影響についても配慮する必要がある。

4.1.4 プロジェクト終了後の継続運用体制

【提案②】で構築する健康・医療RWD流通の仕組みは、長期的に運用・改善を継続する。3章で述べた通り、悉皆性の高いデータ流通については公的に運用する一方で、質の高い診療情報の流通は民間を主体に運営するのが望ましいと考えられ、次世代医療基盤法では民間の認定事業者による運営が想定されている。ただし、データの特性上、セキュリティなどへの継続的な投資は不可欠であり、コストをデータ提供の収益だけで賄おうとすると、情報漏洩等のリスクが増大して個人・市民の協力が得られにくくなる、データ使用料の

高止まりにより分析データ利用が進まない、といったように、データ利活用のサイクルが滞る結果に陥る可能性がある。また、資金力のある企業によるデータ利活用が優先され、アカデミアによる利活用が妨げられるおそれがある。アカデミアのデータ利用料の公的な補填、共通する技術開発やインフラは公的資金で支え運用を中心としたコストのみを収益で賄う上下分離方式など、経済的に持続可能な運用ルールの整備が必要である。また、流通プレーヤーやRWD需要の呼び水として、【提案③-1】の基盤モデル開発向け大規模データベース構築のように、国が主導でRWD二次利用の大きな需要を創出することも有効と考えられる。例えば米国では、NIH傘下の国立老化研究所（NIA）が2023年4月に、アルツハイマー病や関連する認知症の研究を促進するために、診療情報、保険請求情報、個人収集の健康データなどのRWDを大規模に集約するデータベースを構築することを公表した^[19]。予算としては最大6年間で3億ドルと大きな額であるが、一からデータ収集を行うわけではなく、既存の健康・医療RWDの流通スキームを活用したデータ蓄積に相当の予算が投じられると考えられる。RWDなしには研究が難しい疾患、例えば認知症や精神疾患などについて、こうした施策をわが国でも実施することで、研究が進むだけでなく、健康・医療RWD流通基盤を利用する安定的な需要を生み出すことにもなる。

医療版LLM/基盤モデルも一度開発したら完成ではなく、流通するデータを活用して継続的に改良することが望ましい。プロジェクト終了後の運用方法（運用主体、コスト負担など）についても、本提案を実施する中で検討を実施する必要がある。

4.2 時間軸

本戦略プロポーザルを推進する際の時間軸のイメージを図4-1に示す。4.1で述べた通り、マイルストーンを設定して進めるのが望ましいことから、本提案では4つのマイルストーンを例示した。具体的なマイルストーンについては、図3-3のレベル分けの結果などを参考にして、医療と情報科学・データ科学を中心とした有識者で決定する必要がある。また、中長期でのマイルストーンやその達成のための実施項目については、生成AIの技術動向や国内外での活用状況、開発を通じて蓄積された知見やノウハウを基に適宜見直すのが望ましい。なお図4-1には、データヘルス改革など関連する政策・プロジェクトも図示したが、それらの詳細については付録2.3を参照いただきたい。

各提案における進め方の詳細は以下の通りである。

【提案①】 医療版基盤モデルの開発のベースとなる取り組みであり、診療支援に直結しうる取り組みであるため、早い段階で集中的に実施する。LLM開発については、国内の知見・ノウハウを蓄積しながらの実施となることから、マイルストーンを段階的に分けて反復し、最適な学習手法や必要な学習データとその処理方法などを見極めながら実施するのが望ましいと考える。データ交換規格や各種コード整備の実施時期や優先度は、マイルストーンから逆算して設定するのが望ましい。

【提案②】 医療版基盤モデルの開発においてデータ蓄積のカギとなる取り組みであることから、早い段階で完了させる必要がある。**【提案②-1】****【提案②-2】****【提案②-4】**を優先し、外部データベースへのデータ提供の仕組み整備は、基本的な仕組みが構築された後に実施する必要がある。構築後も継続的な改善・拡大が望ましい。

【提案③】 データウェアハウスについては、序盤に立ち上げた後は、健康・医療RWD収集・流通の仕組み整備状況やマイルストーンに合わせて蓄積するデータを拡大することが望ましい。医療版基盤モデルの開発は**【提案①-3】**のLLM開発の次のステップとして設定するのが望ましい。臨床現場への医療AI実装に向けた法規制や倫理的に不適合な利用の防止については、開発や実証実験と並行して前半から段階的に検討を実施する必要がある。

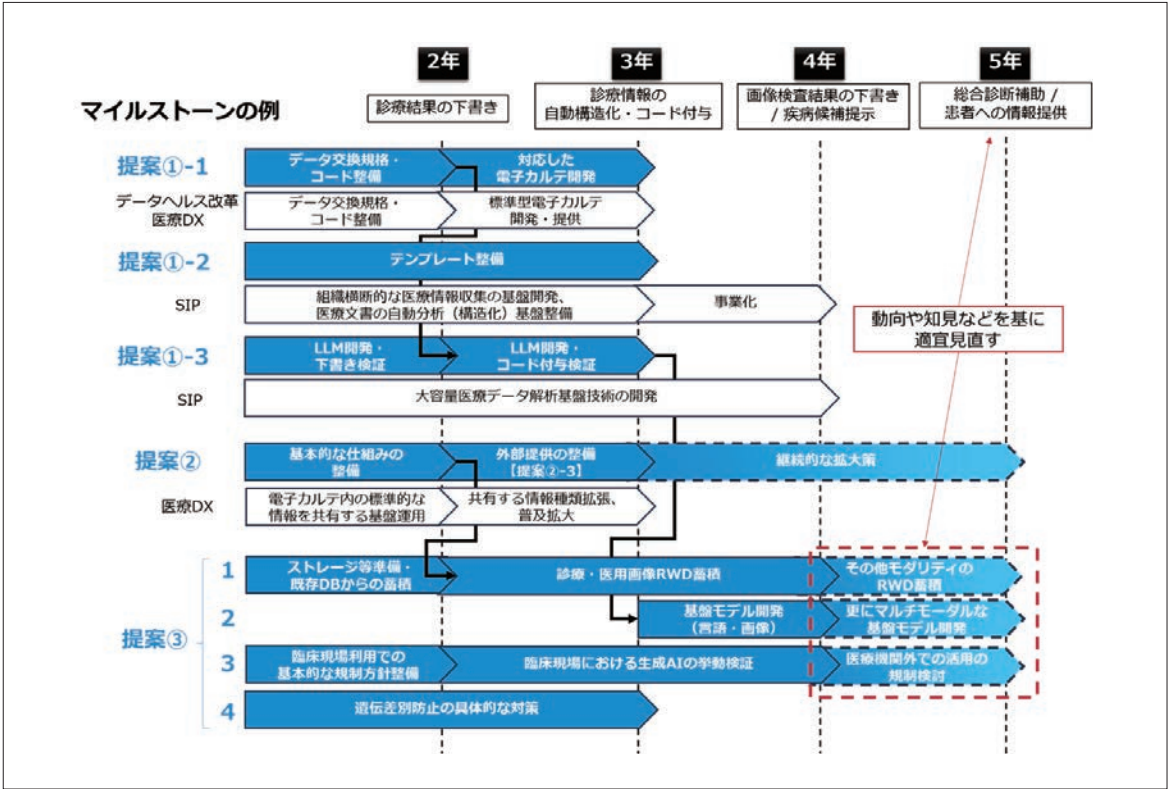


図4-1 時間軸イメージ（関連プロジェクト含む）

付録1 検討の経緯

国立研究開発法人科学技術振興機構（JST）研究開発戦略センター（CRDS）では、2021年度の戦略スコープ検討委員会において、戦略プロポーザルを作成すべきテーマ候補の1つとして本テーマを選定し、2021年6月に検討チームを発足させた。その後、検討チームにおいて提言作成へ向けた調査・分析・検討を進めた。

検討チームでは、国内外の研究開発動向・技術水準を調査し、わが国が推進すべき方向性を模索した。その上で、個人データ駆動医療・ヘルスケアに関する研究開発に関してCRDSが構築した仮説を検証する目的で、科学技術未来戦略ワークショップを開催した。ワークショップの結果は報告書としてCRDSより発行し、CRDSウェブサイトにて公開している^[20]。

付録1.1 インタビュー・意見交換を実施した識者

（五十音順、敬称略、所属は実施時点）

江花 有亮	（東京医科歯科大学 生命倫理研究センター）
大江 和彦	（東京大学大学院医学系研究科）
川上 英良	（理化学研究所 情報統合本部）
黒田 知宏	（京都大学医学部）
近藤 充弘（他6名）	（日本製薬工業協会 医薬品評価委員会）
佐藤 寿彦	（株式会社プレジジョン）
鈴木 正朝	（新潟大学法学部）
瀬戸山 晃一	（京都府立医科大学 医学生命倫理学）
田岡 和城	（東京大学医学部附属病院）
高橋 慧	（理化学研究所 革新知能統合研究センター）
武田 理宏	（大阪大学医学部）
田中 博	（東京医科歯科大学 医療データ科学推進室）
田宮 元	（東北大学大学院医学系研究科、理研AIP）
角田 達彦	（東京医科歯科大学 難治疾患研究所）
中路 重之 / 村下 公一（他2名）	（弘前大学健康未来イノベーションセンター / ICI株式会社）
中島 直樹	（九州大学病院）
長神 風二 / 荻島 創一	（東北メディカルメガバンク）
橋田 浩一	（東京大学 情報理工学系研究科 ソーシャルICT研究センター）
浜本 隆二	（国立がん研究センター研究所 医療AI研究開発分野）
彦惣 俊吾	（大阪大学医学部循環器内科）
松田 浩一	（東京大学医科学研究所）
松村 泰志	（国立病院機構 大阪医療センター）
宮内 恒（他1名）	（三井住友銀行 デジタル戦略部 / 株式会社プラスメディ）
森 健策	（国立情報学研究所 医療ビッグデータ研究センター）
森田 朗	（一般社団法人 次世代基盤政策研究所）

康永 秀生	(東京大学大学院医学系研究科公共健康医学専攻)
山本 隆一	(医療情報システム開発センター)
横野 恵	(早稲田大学社会科学総合学術院)
米村 滋人 (他3名)	(東京大学法学部)
脇 嘉代	(東京大学大学院医学系研究科)

付録1.2 科学技術未来戦略ワークショップ^[20]

「個人データ駆動医療・ヘルスケア」

【日時】2022年11月29日(火) 10:00～15:00

【場所】JST東京別館2階会議室A①およびオンラインのハイブリッド開催

【プログラム】

10:00-10:10 趣旨説明(宮菌侑也、CRDS特任フェロー)

10:10-10:20 開会挨拶(永井良三、CRDS上席フェロー)

話題提供①：総論

10:20-10:30 「医療DXの取り組みについて」

田中 彰子(厚生労働省)

10:30-10:40 「ヘルスケア・デジタル基盤としての電子カルテ」

美代 賢吾(国立国際医療研究センター)

10:40-10:50 「FHIR標準による健康医療情報管理」

大江 和彦(東京大学)

話題提供②：技術、国内外事例

11:00-11:10 「医療情報における匿名化・仮名化に関わる制度整備の現状と問題」

山本 隆一(医療情報システム開発センター)

11:10-11:20 「CLIDASにおける多施設情報収集と活用」

的場 哲哉(九州大学病院)

11:20-11:30 「AIを有する統合がん診療支援システムの開発」

佐野 武(がん研有明病院)

11:30-11:40 「製薬産業を守る<日本を守る>

青木 事成(日本製薬工業協会 / 中外製薬)

11:40-11:50 「医療デジタルツインの社会実装における海外動向」

反田 篤志(マッキンゼー・アンド・カンパニー)

11:50-12:00 「医療・保健政策のための医療情報利活用」

満武 巨裕(医療経済研究機構)

話題提供③：技術、法律・医療システム

13:00-13:10 「医療データ利活用のための制度基盤」

児玉 安司(新星総合法律事務所)

- 13:10-13:20 「米国における電子カルテデータ活用の現況：研究者の視点から」
八木 隆一郎（Brigham and Women's Hospital）
- 13:20-13:30 「医療情報に関するデータガバナンス制度整備の必要性」
森田 朗（次世代基盤政策研究所）
- 13:30-13:40 「医療デジタルツイン実現に向けた情報科学の課題」
喜連川 優（国立情報学研究所）
- 13:40-14:40 総合討論（ファシリテーター：永井上席フェロー）
コメンテーター 坂田 恒昭（大阪大学共創機構、CRDS 特任フェロー）
佐藤 寿彦（株式会社プレジジョン）
春原 典彦（富士通Japan）
- 14:40-14:45 閉会挨拶（谷口 維紹、CRDS 上席フェロー）

【ワークショップ参加識者】

（五十音順、敬称略、所属はワークショップ開催日時点）

発表者

- ・青木 事成 日本製薬工業協会 医薬品評価委員会 / 中外製薬株式会社
- ・大江 和彦 東京大学大学院医学系研究科 教授
- ・喜連川 優 国立情報学研究所 所長 / JST-CRDS 特任フェロー
- ・児玉 安司 新星総合法律事務所
- ・佐野 武 がん研究会有明病院 病院長
- ・反田 篤志 マッキンゼー・アンド・カンパニー パートナー
- ・田中 彰子 厚生労働省医政局 参事官
- ・的場 哲哉 九州大学病院循環器内科 診療准教授
- ・満武 巨裕 医療経済研究機構 副部長
- ・美代 賢吾 国立国際医療研究センター 医療情報基盤センター長
- ・森田 朗 次世代基盤政策研究所 代表理事
- ・八木 隆一郎 Division of Cardiovascular Medicine, Brigham and Women's Hospital Research fellow
- ・山本 隆一 医療情報システム開発センター 理事長

コメンテーター

- ・坂田 恒昭 大阪大学共創機構 特任教授 / JST-CRDS 特任フェロー
- ・佐藤 寿彦 株式会社プレジジョン 代表
- ・春原 典彦 富士通Japan 電子カルテソリューション事業部長

付録2 わが国における問題点の補足説明、国内外の関連する状況など

付録2.1 「わが国の医療・ヘルスケアの問題」の補足説明

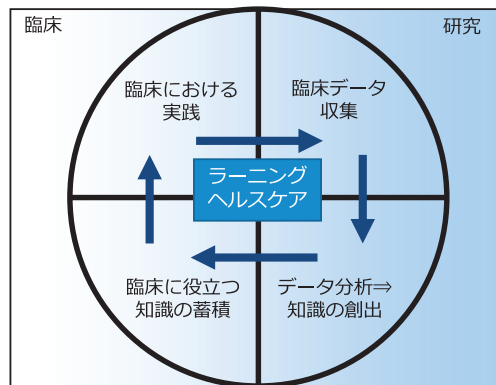
本節では、「2.1.1 わが国の医療・ヘルスケアの問題」で述べた概要の補足、ならびに海外での状況について述べる。海外の状況、特に日本と比べて進んだ事例については、医療・介護制度が国によって大きく異なることに留意して解釈すべきである。例えば、米国やイスラエルなどは民間の医療機関/保険会社が医療・介護提供の主体であり、政府の関与はセーフティーネット的な位置づけに留まる。英国や北欧は、公的機関/公的保険が医療・介護提供の主体である。一方、わが国は公的保険（国民皆保険）でありつつも、医療・介護提供は公的なものと民間のものが混在する、比較的ユニークな仕組みとなっている。従って、医療・介護制度の改革において、諸外国の仕組みをそのまま導入することはできず、精緻なデータに基づいて、公的/民間の関係者が協議し進めていくことが肝要である。

①わが国の臨床現場の疲弊

臨床の現場の疲弊が深刻な問題となっている。医師は、診療に加えて電子カルテの入力に時間を取られているとの指摘が、世界各国で見られる。英国での民間企業による調査^[21]によると、医療従事者は臨床文書の作成に週平均で13.5時間、正規労働時間の1/3相当を費やしており、過去7年間で25%増加していた。一方、必要な情報を採るのに要する時間（1日当たり約1時間）は、過去7年間で微増であった。電子カルテ導入により情報がデジタル化し可読性が向上したにもかかわらず、情報入力や検索に要する時間は増えており臨床現場の効率化に寄与していない。これは患者に向き合う時間や新たな知識を得る時間が減ることを意味しており、医師の燃え尽きが発生する要因の1つになっているとの指摘もある。

電子カルテへのデータ入力の作業が直接的に研究用データベースの構築につながるわけではない。付録2.2で後述するような問題により、臨床医が電子カルテのデータを研究用データベースに転記する必要があり、臨床研究を行う上での大きな負担となっている。わが国のCOVID-19対策でも、感染患者の登録作業に多大な労力が必要なために、感染者の全数把握ができないという状況に陥った（例えば、電子カルテの基本情報とHER-SYSとの連携もなされていなかった）。

RWD収集だけでなく、データから得られた知識を臨床現場で活用する上でも問題がある。これまでの医学の発展は本質的には経験に基づいており、従来より診療情報は医療の評価や改善に用いられてきた。近年のIT技術や計測・解析技術の進化を背景に、診療情報の収集と知識の産出が加速しており、臨床研究論文の報告数は増加の一途である。また、昔より行われている現場での医学知の蓄積に対して、IT技術を活用することで、臨床現場で情報収集と知識創出を一体的に進めて現場を改善しようというlearning healthcareという概念がある^[22]（図A2-1）。



図A2-1 ラーニングヘルスケア

出典：文献[22]の図を基にCRDS作成

このように、知識の積み上げがデータ駆動となって高速・大量化しつつあるのに対して、実践する主体である医師側の、知識のアップデートや習得が高速化しているわけではないため、キャッチアップし切れないという状況が発生している。そうした状況をAIで解決すべく、AI（ワトソン）が過去の論文などから抽出した知識を基にがんの治療方針を提案する、というプロジェクトをIBMが立ち上げたが、少なくとも商業的には失敗に終わった。医療機器として開発・認可された一部のAI医用画像診断などを除き、知識の習得とその実践は人の手で行われている状況である。

こうした状況は、患者にとっても、既知の希少疾患の診断ができない・診断に時間がかかる、などの弊害を生じさせている。例えば、希少疾患「遺伝性血管性浮腫（HAE）」について、わが国には2,000人から3,000人の患者が存在すると推定されているが、診断されている患者は約450名に留まっており未診断の患者が多くいると考えられているのに加え、初めての発作が起きてから確定診断までに平均13.8年かかるといわれている^[23]。

②わが国の地域格差解消、持続的な医療提供に向けた資源再配分の遅れ、および海外の関連する事例

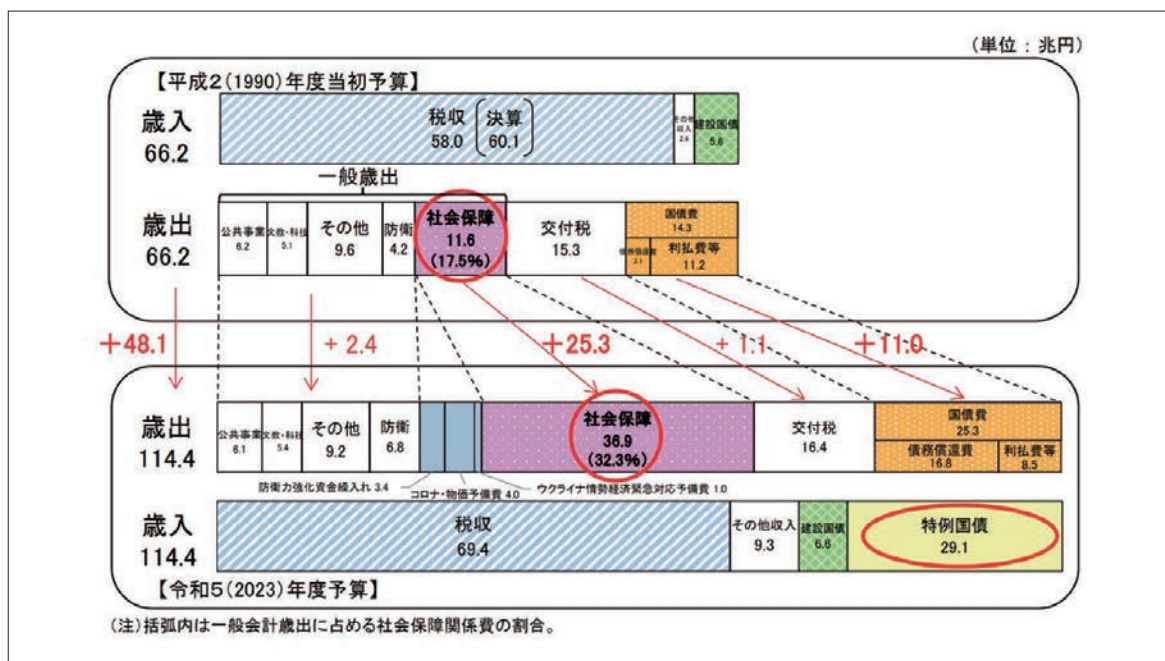
わが国の医療制度では、医療のアクセス、コスト、医療の質という3つの相反するものを高いレベルで共存させることに成功しており、世界的にも高い評価を得てきた。しかし、20世紀後半に構築されたわが国の医療制度にも限界が見られつつある。わが国の人口当たりの病床数は米国の5倍、ヨーロッパの2-3倍である。一方、100床当たりの医療スタッフは米国の1/5、ヨーロッパの1/2から1/3である^[2]。いわば広く薄く医療資源を配置している。このため病床利用は必ずしも効率的ではなく、集中治療が必要な救急患者や重症患者のケアは手薄となる。その限界が露呈したのがCOVID-19まん延時の医療ひっ迫であり、わが国は病床数が世界最多であるにも拘らず、緊急入院ができず自宅で死亡する重症者が相次いだ。また、医療だけでなく、社会・経済活動にも制限がかかり、大きな負の影響をもたらした。

こうした有事対応に加え、人口減少や高齢化が都市部よりも顕著なへき地の医療の維持といった課題に対し、医療資源を有効に配置することが必要である。わが国では、行政が医療費の総額と個々の診療報酬を管理している一方で、医療を提供する病院の8割が民間経営であり、国や自治体の指示だけでは動かすことができない。そのため、平素より医療現場の状況を示すデータに基づいて関係者が話し合い、相互理解を深めておく必要がある。平成24-25年度に開催された社会保障制度改革国民会議では、「医療データに基づく医療システムの制御」が謳われ、データに基づいて関係者が協議によって方向性を示すとされた。それを機能させるには、地域毎の医療状況の可視化と医療の有効性の分析、とくに医療提供シミュレーションのためのシステム開発が必要である。しかし、その土台となるデータが医療現場や自治体などに散在しており分析できておらず問題である。

COVID-19対策における英国の事例を挙げる。英NHS（国民保健サービス）と電子カルテベンダーの協力の元、オックスフォード大の研究者らが英国の人口の40%をカバーする電子カルテデータと統計局管理の死亡情報をリンクさせたデータベースOpenSAFELYを構築し、10,926人の死亡者の診療情報から、COVID-19による死亡に関連する臨床的要因を、英国での感染発生から半年足らず（2020年8月）で解析した^[24]。RWDからこうした知識が産出できれば、診療の優先度を判断するトリアージや医療資源の再配分などの検討が可能である。しかし、わが国ではこうした知識の産出において必要な各種情報をリンクさせた大規模データを迅速かつ効率的に取得する仕組み・体制が未整備で、対応が大きく遅れた。

③国民医療費、介護費の増大

わが国は世界に先駆けて少子高齢化が進んだ結果、社会保障費の増加による国家財政の圧迫、労働人口の減少が深刻な問題となっている。わが国の高齢化率（総人口に占める65歳以上人口の割合）は2021年に28.9%、2025年には30%となる見込みである。また、2021年の75歳以上の人口の割合は14.9%である。2040年頃には、いわゆる団塊ジュニア世代が高齢者となり、高齢者人口がピークを迎える一方、現役世代が急激に減少する。高齢化による医療・介護ニーズの増大が、国民医療費・介護費の高騰の主たる要因である。厚生労働省「国民医療費の概況」によると、2021年度の国民医療費は45.0兆円、うち11.4兆円は国庫、5.7兆円は地方自治体の負担である。医療保障費を含む社会保障費がわが国では約30年前と比べると3倍以上に増大する一方で、税金などの収入は1-2割程度しか伸びておらず、国民医療費・介護費の高騰は極めて深刻な問題である。



図A2-2 歳出・歳入の比較（1990年度-2023年度）

出典：財務省、日本の財政関係資料（2023年10月）

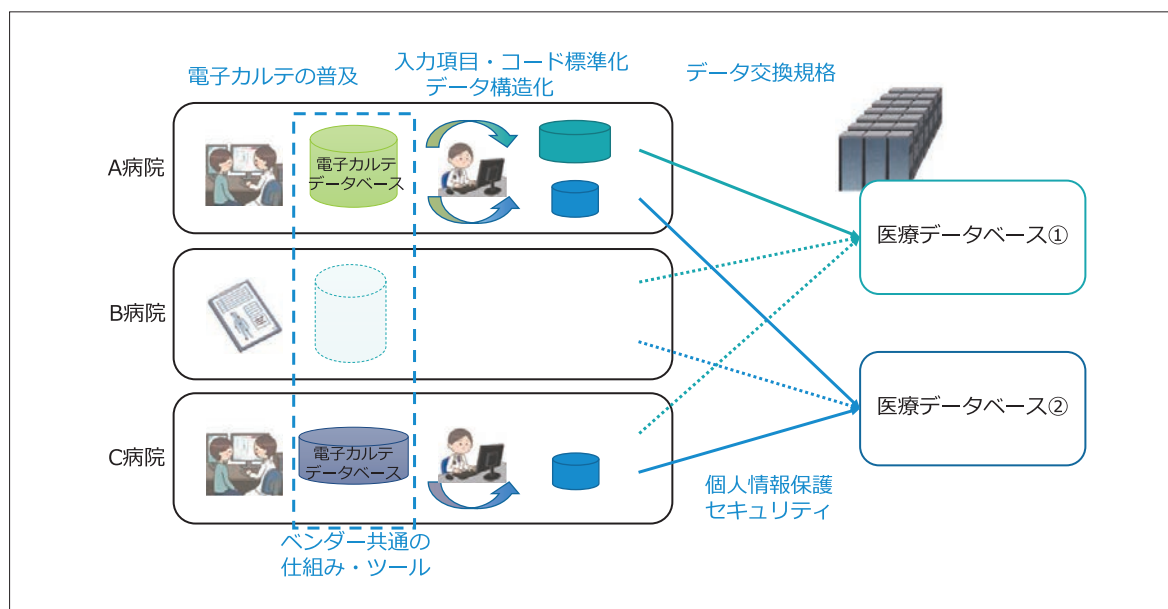
国民医療費の高騰のもう一つの要因が、高額な医療技術の登場である。およそ2000年代後半以降に抗体医薬、2010年代後半以降には核酸医薬や遺伝子治療などの、新しいタイプの医療技術が次々と上市されている。これら医療技術は、投与対象となる患者数は限られているものの、数百万～数億円と高額である。

医療技術評価と医療政策への実装が世界で最も活発に実施されている英国では、NICE（The National

Institute for Health and Clinical Excellence) が医療技術の多面的な評価を実施し、公的保険である NHS (National Health Service) への保険収載可否を事実上決定している。一方、ドイツでは、IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) が新薬の承認後1年以内に有用性評価を実施し、その結果によって保険償還額を減額するという仕組みが実装されている。またドイツでは、同様の考えに基づいたデジタルヘルス法 (DVG Digitale Versorgung Gesetz) により、安全性や情報セキュリティ等の一定の要件を満たしたデジタルヘルスアプリについては、有効性が未証明でも暫定的な承認申請が可能なファストトラックを設定し、1年間保険償還される状態で、RWDを用いた有効性確認の実施が可能な仕組みを導入している。こうした仕組みにより、事前に臨床・実社会でのアウトカムを予測することが難しい医療・ヘルスケア技術に対しても、市販後にアウトカムのデータに基づきそれに見合った適正なコストを設定することが可能となる。デジタルヘルスは疾病予防や健康増進への効果が期待され医療費削減にもつながる可能性があるものの、アウトカムを示すデータを実社会で効率的に収集するための基盤が未整備の日本では、合理的なコスト設定が難しい状況である。

付録2.2 「RWD収集・流通・知的基盤の構築とその活用における問題」の補足説明

以下、①～③においてRWD収集、④でRWD流通、⑤で知的基盤の構築とその活用における問題について、本文の補足を述べる。



図A2-3 診療情報収集における問題まとめ

①電子カルテの普及率の伸び悩み

医療機関内のデータを外部に集約させるには、そもそもデジタル化されている必要がある。病院内で収集される情報の中でも、レセプトの電子化は非常に早く進み、1980年代後半の時点ですでに90%近くの診療報酬請求書が電子的に処理され、皆保険制度を支えるシステムとなった^[25]。一方で、電子カルテのシステム

の普及状況としては、令和2年時点で一般病院では57.2%である。400床以上の病院は9割を超えている一方で、200床未満や一般診療所では5割を切っている。電子カルテ導入のコストを支払うだけのメリットやインセンティブがないのが主な理由と考えられる。そのため、現状では電子カルテの情報をいくら集めても、悉皆性の点ではレセプトのデータにはまだまだ及ばない。

	一般病院 (※1)	病床規模別			一般診療所 (※2)
		400床以上	200～399床	200床未満	
平成23年 (※3)	21.9% (1,620/7,410)	57.3% (401/700)	33.4% (440/1,317)	14.4% (779/5,393)	21.2% (20,797/98,004)
平成26年	34.2 % (2,542/7,426)	77.5 % (550/710)	50.9 % (682/1,340)	24.4 % (1,310/5,376)	35.0 % (35,178/100,461)
平成 29年	46.7 % (3,432/7,353)	85.4 % (603/706)	64.9 % (864/1,332)	37.0 % (1,965/5,315)	41.6 % (42,167/101,471)
令和 2年	57.2 % (4,109/7,179)	91.2 % (609/668)	74.8 % (928/1,241)	48.8 % (2,572/5,270)	49.9 % (51,199/102,612)

【注 釈】

(※1) 一般病院とは、病院のうち、精神科病床のみを有する病院及び結核病床のみを有する病院を除いたものをいう。

(※2) 一般診療所とは、診療所のうち歯科医業のみを行う診療所を除いたものをいう。

(※3) 平成23年は、宮城県の石巻医療圏、気仙沼医療圏及び福島県の全域を除いた数値である。

図 A2-4 電子カルテ普及率

出典：厚生労働省「医療施設調査」

②分析可能なデータ収集に向けた入力項目・コード化、および構造化の現状

- ・入力項目：データの質を均一化する観点では、ある診療課題（疾病・症例・処置法）に対して、検査・問診を行い電子カルテに記録する項目が決まっていることが望ましい。しかし、診療課題によって必要な項目が異なる、必要十分な項目数の合意形成が難しい（数が多すぎると現場で使われない、数が少なすぎると解析の価値が低減）といった要因から、広く認知された記録項目セットの標準は策定されておらず、実際には医療機関や医師によって記録する項目がバラついている。
- ・コード化：例えば同じ検査項目をそれぞれの医療機関で記録する際に、「血清総コレステロール」「総 Cholesterol」や「T-コレステロール」といった表記揺れが生じ、違う項目として認識されてしまう可能性がある。それを防ぐためのツールがコードで、検査であれば日本臨床検査医学会によりJLACが整備されている。また病名については、疾患分類コードの国際標準としてICD-10があり、それに準拠する形で日本の「標準病名マスター」が制定され、70%以上の医療機関で使われている。病名・検査・薬品コードについては厚生労働省推奨となっているが、全ての医療機関で統一的に使われているという状況ではなく、また現場で正確に病名や検査のコードを付与するのは難しい。また、項目がコード化されていても、検査値の単位（標準的と思われる検体検査の単位が、大学病院や国立循環器医療研究センターといった主要な病院間でも同一でなかった事例もある）などのバラつきがあると、データが相互参照できない要因となる。
- ・構造化：電子カルテは、基本的に人間の医師や看護師などが紙に記載していたものが電子化したものであり、フリーテキストで書かれていることが多い。これを解析可能・機械可読な形にするためには構造化が必要である。標準化・コード化された項目は構造化された状態で記録されるが、画像・生理検査レポート、経過記録、退院時サマリなどの患者の状態を記録したデータはフリーテキストで記録されている。こうしたデータの登録のためには、現状は電子カルテから臨床医などが個別にデータを書き出して登録す

の必要があり、大変な人手を使っている。電子カルテ入力時に構造化した状態でデータを入力するためのテンプレートを作成・導入している例があるが、電子カルテシステムの違いや入力項目・様式の違いにより、病院内の展開に留まっているケースが多い。自然言語処理による構造化も試みられているが、さまざまな病院や診療科で広く使えるモデルの構築には至っていない。

③オンラインでデータ交換可能な規格の不備

HL7は、コンピュータ間での医療文書情報のデータ交換を標準化するための国際規格である。現在の厚労省の標準規格となっているのは、HL7 ver2.5 に対応したSS-MIXである。SS-MIXはオフラインでの使用を前提としており、SS-MIXにより医療機関内のストレージに出力したデータを適切な媒体で共有ストレージに出力し、外部から閲覧したい場合はそちらを参照する、といった設計になっている。SS-MIXによる診療情報のデータベース構築は、循環器疾患のレジストリ研究であるCLIDASを始めとしてこれまでも行われて、優れた成果を上げている。一方で、医療機関外とのデータ交換には向いておらず、またHL7 FHIRのように汎用技術をベースとしていないことから標準準拠ツール作成やメンテナンスの手間も大きく、全ての病院で採用されている状況ではない。

ポータブルな医療機器やウェアラブル端末といったデータ取得手法、PHRによるモバイル端末上でのデータ参照のニーズが出てくる中で、オンラインでのデータ交換の必要性が高まり、web通信技術（REST）を採用した規格であるHL7 FHIR（Fast Healthcare Interoperability Resources）が、2010年代半ばに登場した。米国、英国を始め、世界各国でHL7 FHIRの導入が進んでいる。わが国でも、国内におけるFHIRを使用する場合の共通ルール整備や、厚労省標準にある医療文書や項目についてFHIRで記述するためのユースケースごとの仕様の整備、実装実験については、日本医療情報学会を中心に厚労科研などで実施されている。実装や普及に向けて、電子カルテベンダーなども巻き込んで更なる推進が求められる。

④ヒトデータを用いた臨床研究の倫理審査に要する負担

ヒトやヒトデータを対象とした臨床研究では、原則として事前の倫理審査が必要である。介入なし（身体への重大な有害事象の可能性がない）にヒトデータを用いる研究、いわゆる観察研究について、ハーバード大学関連病院においては、個別の研究に対する病院への研究提出や倫理申請も不要であり、包括的な研究申請のみで研究を実施できる。つまり、包括的な研究申請に沿っている研究であれば思い付いた段階で研究に着手できる。一方、わが国では、個別の研究に対して計画作成や倫理審査が必要なため、手間や時間がかかる。試しに調べてみようという研究者、データ科学者の好奇心を削いでいると考えられ、新たな知識を産出する上での妨げになっている。また、わが国は施設毎に倫理審査委員会があり、わが国の倫理審査委員会は2,500～3,000個ほどあり、人口比換算では米国の2倍、英国の30倍以上であると言われている^[26]。医療データを利活用した研究の場合、個人情報に関わる法律の解釈や情報漏洩・セキュリティの観点が必要で、従来倫理審査委員会を対象としていた医療倫理とは異なる専門性が求められる。こうした有識者をすべての倫理審査委員会で揃えることは難しく、審査に時間がかかりやすいのに加え、判断のブレが大きくなりやすい。

⑤倫理的に不適合な利用・悪用への懸念

ビッグデータに基づく判定、特に深層学習による判定は、その根拠を人間が理解することが難しくなる。技術的には説明可能AI、因果推論といった対処方法があるが、合理的な層別化なのか不当な差別と見なされるのかは倫理的な検証も不可欠である。研究倫理審査で目的の正当性については評価ができるものの、倫理的に不適合な結果を完全に排除することはできない。また、匿名加工情報であっても、多数の情報を名寄せすると個人を推定できる確率は高まり、プライバシー侵害など悪用される可能性が否定できない。CT画像の学習済み深層学習モデルから学習に用いた画像データのある程度再構築できた、という研究結果も報告されている^[27]。さらには、セキュリティ対策を万全にしてもさまざまな要因でデータ漏洩は発生しうる。つまり、

情報収集や研究開始時の規制により、データ利用によって生じうる、個人に不利益を被るような倫理的に不適切な事象が起こる懸念を完全に排除することはできない。こうした不利益に対して、悪用されたデータを提供した側に責任が及んでしまうと、責任を問われるリスクを避けるためデータ提供に消極的になってしまうと考えられる。

そうした懸念がデータ活用の妨げになることを防ぐための手段として、不利益を生じさせる行為そのものの防止や、不利益を被った時の救済手段の整備が考えられる。特に、遺伝情報（ゲノム情報、家族病歴など）は、生まれながらに持っているもので生涯変わらないこと、医療以外にも保険、雇用、結婚などさまざまな場面で不利益をもたらしかねないことからその懸念が強く、遺伝情報による差別を禁止する法律や規制が、欧米各国で導入されている。差別的扱いによる影響が大きいと考えられる保険分野での規制がメインとなっている。各国がこういった規制を導入した背景の1つとして、人々が差別のリスクを理由に医学研究へ協力しなくなるおそれがあり、わが国でもデータ活用を妨げる個人・市民の懸念の1つとなりうる。わが国では現時点ではそういった法規制はないものの、わが国でも一定数（3%程度）の人が遺伝情報による不適切な扱いを受けた経験があるとの調査結果^[28]が報告されており、2022年4月に日本医師会が遺伝情報による差別防止に関する声明を出す^[29]など、患者団体を中心に導入を望む声がある。わが国は国民皆保険であるため、優先度が高いと考えられてこなかったが、保険以外の雇用・婚姻などの影響が大きい可能性がある。

こうした状況に対して、2023年6月に成立したゲノム医療推進法において、遺伝差別の防止が明記され国に計画策定を義務付けていることから、具体的な対策についての議論が進むものと期待される。

付録2.3 国内の関連する政策・プロジェクト

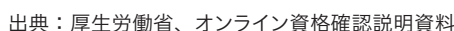
本戦略プロポーザルと関連するわが国のプロジェクトとして、厚生労働省が推進する「データヘルス改革」と医療DX推進本部主導の「医療DX」、内閣府SIP第三期「統合型ヘルスケアシステムの構築」について、関連部分を中心に説明する。また、地域は限定されているものの、国内における健康・医療RWD利活用の先行事例として、弘前COI/COI-NEXTの取り組みを説明する。

データヘルス改革/医療DX

厚生労働省は、健康、医療、介護分野の有機的な連結やICTとの技術革新の利活用の推進を目指して、「データヘルス改革」の取り組みを2017年より続けてきた。データヘルス改革では、以下の4つの取り組みを行っている。

- ①自身の保険医療情報を閲覧できる仕組みの整備：健診・各種検診・レセプト等の情報閲覧の仕組み
- ②医療・介護分野での情報利活用の推進：医療機関間の情報共有、介護におけるデータ収集・共有
- ③ゲノム医療の推進：全ゲノム解析等実行計画
- ④基盤の整備：マイナンバーカードを利用したオンライン資格確認の仕組み構築・運用、など

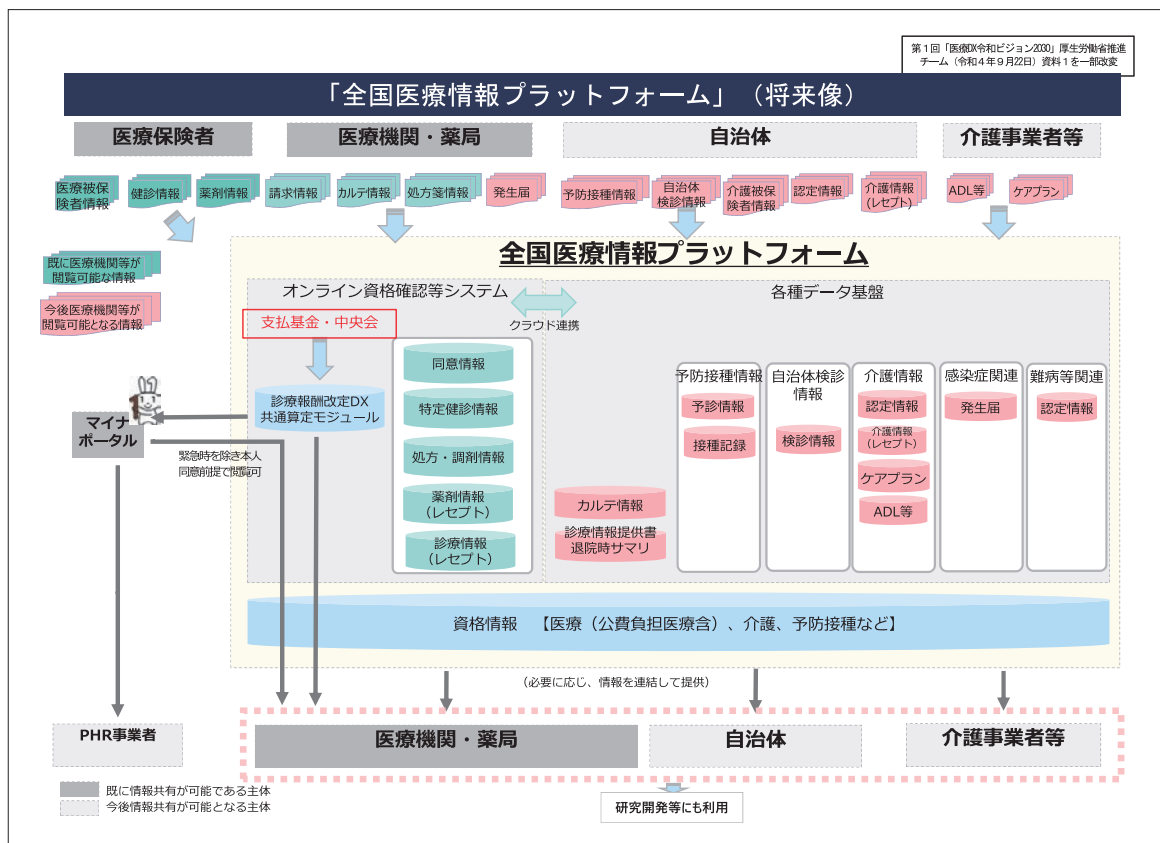
基本的には③を除くと、本人閲覧や（本人同意に基づく、医療・ヘルスケアサービス提供のための）医療機関同士の共有など、健康・医療RWDの一次利用を促進・効率化するための施策である。①②では電子カルテデータも閲覧・共有の対象として挙がっているが、想定されているのは傷病名やアレルギー情報など基本的・汎用的なデータである（二次利用の観点では『悉皆性の高いデータ』に該当する）。④で整備する「オンライン資格確認」は、1）全国の医療機関や薬局が安全かつ常時に接続できる、2）医療情報を個人ごとに把握して、本人の情報を確実に提供することが可能、それから、3）患者、利用者の同意を確実に電子的に得ることが可能、といったシステムであり（図A2-5）、①②を推進・拡大する上での前提条件となるものである。



医療DXの推進に関する工程表〔全体像〕



「全国医療情報プラットフォーム」では、前述のオンライン資格確認システムをベースに、自治体や介護事業者からの情報についても、それぞれのデータを集約するクラウドの間を連携させることにより、共有できるようにすることを目指す。また情報提供の対象についても、医療機関と薬局に加えて自治体や介護事業者にも拡大するものである。扱うデータ種については、外来/入院/調剤などのレセプト、受診歴や診療・手術実績などの基本的な診療情報、特定健診、予防接種、乳幼児健診などの情報は対応している。事業主健診・学校健診は、提供を求める法的根拠も含め整備が予定されている。



図A2-7 全国医療情報プラットフォーム

出典：内閣官房、医療DX推進本部幹事会資料

「電子カルテ情報の標準化」については、データヘルス改革の中でも、前述の①②における診療情報の本人閲覧・医療機関間での共有の前提として、電子カルテ情報や交換方式の標準化に取り組んできた。具体的には、傷病名・医薬品・検査などのコード、診療情報提供書や退院時サマリといった一般的な文書のマスターの整備、さらにそれらをHL7 FHIR交換規格で交換するための仕様の公表を順次実施している。それに加えて、医療DXでは実装と普及を進めるため、普及率の低い小規模な医療機関向けのクラウドベースの電子カルテ開発や基金による導入支援の実施が予定されている。

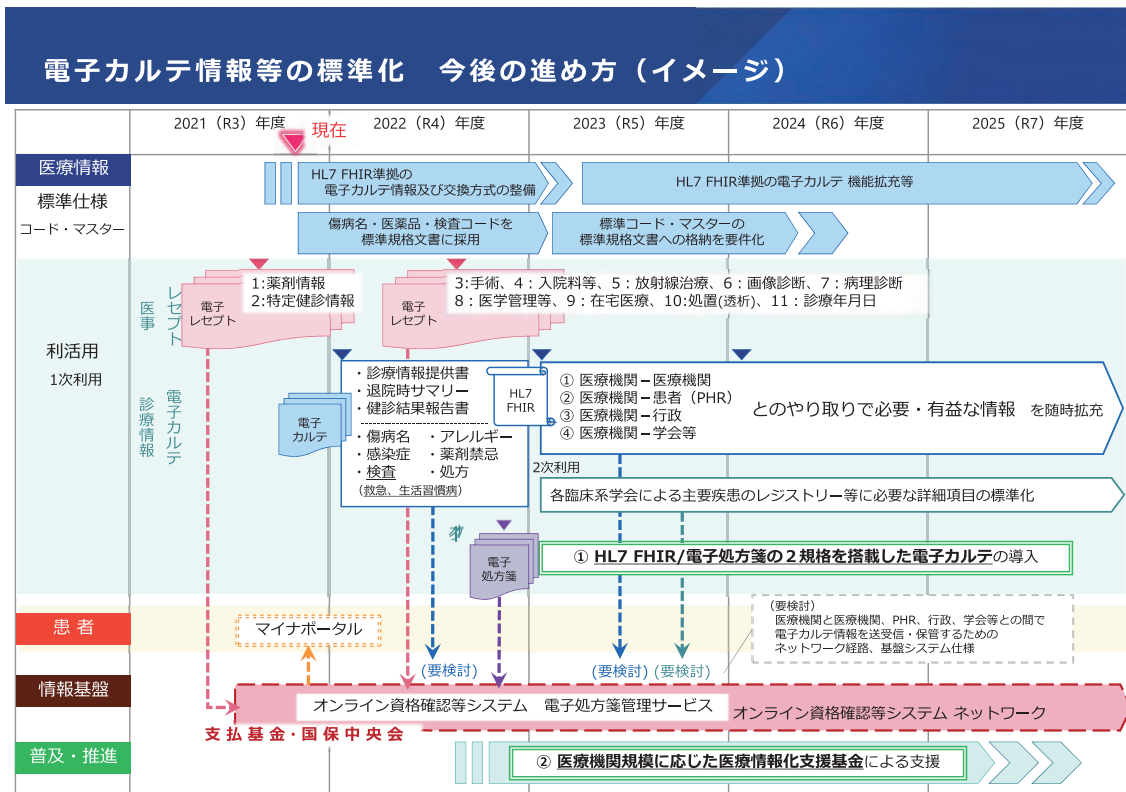


図 A2-8 電子カルテ情報の標準化 進め方イメージ

出典：厚生労働省、健康・医療・介護情報利活用検討会資料

SIP第三期「統合型ヘルスケアシステムの構築」

データヘルス改革、医療DXが健康・医療RWDの一次利用を促進・効率化するための施策であるのに対して、SIP「統合型ヘルスケアシステムの構築」では二次利用による知の発見とその活用に焦点を当てたプロジェクトである。本SIPのミッションは、医療、ヘルスケア、研究開発、医療政策のそれぞれの現場で、実態を可視化し、新たな気づきに基づいて複雑な医療健康システムを制御することである。本SIPの中核的技術開発は、現場の医療・ヘルスケアデータを標準化してデジタル空間に投影するための「医療デジタルツイン」であり、これを基に、現実に近いシミュレーションを行い、社会にフィードバックして制御する。そのために、医療現場の個々の情報、日常生活における健康情報、国と自治体が管理する情報を活用し、①知識発見、②医療実践・患者支援、③地域医療のための3つのデジタルツインを構築する。

具体的には、1) データ収集・統合・処理のための技術開発、およびそこで構築した基盤を生かして、2) 上記①～③における事業・ソリューションの開発、を実施する（図A2-9）。図A2-9に示した課題の中で、本プロポーザルに関連するサブ課題は、技術開発に関わる「D：デジタルツインのための先進的医療情報システム基盤の開発」および「E：大容量医療データ解析基盤技術の開発」であり、特に連携が必要と思われるのは以下の3つのテーマである。

- ・医療機関・ベンダー・システムの垣根を超えた医療データ基盤構築による組織横断的な医療情報収集の実現（テーマD-1）
- ・統合型の医学概念・知識連結データベースの構築及び医療文書の自動分析基盤の整備（テーマD-2）
- ・大規模医療文書の高精度解析基盤技術の開発（テーマE-2）

また、テーマD-1およびテーマD-3「僻地診療支援のためのクラウド型標準電子カルテサービスの研究開発」については、前述の「医療DX」の「電子カルテ情報の標準化」とも関連するテーマである。

 付録2
わが国における問題点
の補足説明、国内外の
関連する状況など

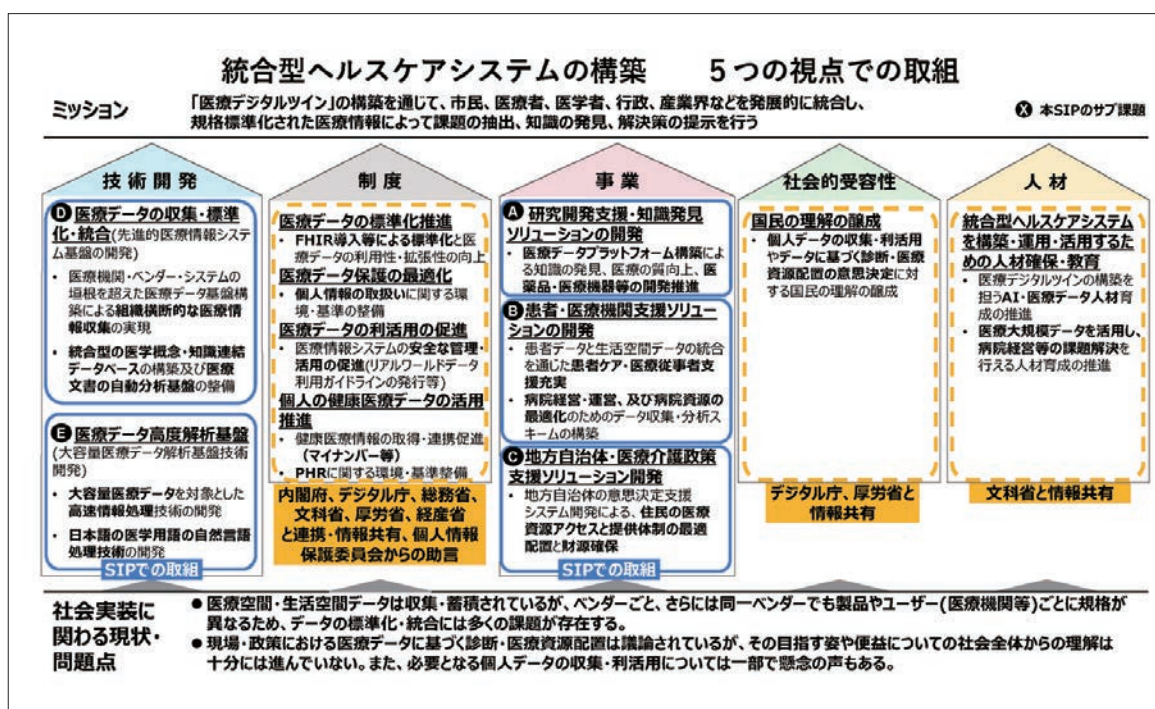


図 A2-9 取り組み概要

出典：SIP 第三期「統合型ヘルスケアシステムの構築」ウェブサイト

国内の先行事例：弘前 COI/COI-NEXT の取り組み

弘前大学では、青森県の「短命県返上」を旗印に、2005年から岩木健康増進プロジェクト（青森県弘前市岩木地区の住民約1000人を対象に毎年実施している大規模健康診断大規模住民合同健診）を実施してきた。2013年から2022年にはJSTのセンター・オブ・イノベーション（COI）プログラムの拠点、2023年からは「共創の場形成支援プログラム（COI-NEXT）」拠点（地域共創分野・本格型）に採択され、2020年には「イノベーションネットアワード2020（第9回地域産業支援プログラム表彰事業）」において「文部科学大臣賞」を受賞した。このプロジェクトの特徴は、地域住民の協力の下で超多項目（約3,000項目）の健康データを収集する健診、およびこの健診を通して蓄積された超多項目ビッグデータであり、さまざまな大手企業や大学がこうしたデータの取得や解析に参画する。また、データプラットフォームを生かして、弘前市・青森県をはじめとする地方自治体、医師会などの関係機関、市民と連携し、自治体による健康宣言や健康リーダーの育成、小中学校での健康授業、企業経営者による健康宣言などの、市民を巻き込んだ健康づくりの仕組みを構築している。

健診を通して蓄積した超多項目データは世界的に見てもユニークである一方で、健常者の健康情報のみであることから、エンドポイント（疾患への罹患やその治療結果）に関する情報が含まれていない。そのため、市民や自治体が興味ある医療経済についての解析などを行うことができなかった。エンドポイントの情報を取得するため、2021年に弘前大学は次世代医療基盤法の認定事業者であるICI株式会社と医療情報の提携契約を結んだ。それにより弘前大学は、同じくICI社と提携契約を結んだ弘前市や青森県後期高齢者医療広域連合が保有する、国保と後期高齢者のレセプトや要介護認定などの医療・介護情報を取得し、超多項目データと紐づけることができるようになった。次世代医療基盤法の枠組みで、医療機関以外の自治体や保険者の有するデータを扱うのは初めてであったが、弘前大学やICI社がオプトアウトに際したデータ提供の説明や議会や市民団体からの質問対応のサポートなどを実施したのに加え、岩木健康増進プロジェクトを背景とした住民との信頼関係もあったことから、オプトアウト比率は0.25%程度と低くなっているとのことである。このように、次世代医療基盤法を活用した医療RWDとの連携により、蓄積した健診データが持つ研究的価値、社

会的価値を更に向上させたという面で、先進的な取り組みであると言える。また、COI-NEXTで健康・予防医学研究の推進とそれに基づく革新的健康モデルの構築を目指すのに対し、蓄積したデータだけでなく研究成果の社会実験を実施するための場や地域住民・社会との信頼関係を有するのが強みであると言える。

こうした弘前における先進的な取り組みにも課題は存在する。1つは電子カルテデータの活用である。上記と同じく次世代医療基盤法の枠組みで、医療機関からICI社に提供された電子カルテデータを活用することは制度面では可能であるが、地域内における対象医療機関が弘前総合医療センターのみで病院名が一意に特定できてしまうことから、実際には利用できていないとのことである（2022年5月時点）。本戦略プロポーザルなどを通して次世代医療基盤法のRWD流通のすそ野が広がることにより、こうした課題の解決が期待される。また別な課題として、健康・医療RWDの活用により地域住民の健康状態や医療経済の改善が見込める一方で、収集したデータの解析や得られる知識の活用はロケーションを問わないため、単純なデータ活用だけでは地域の産業振興にはつながりにくいとの指摘も出ている。COI-NEXTでは、多様なステークホルダーが連携できるインキュベーション環境の形成も目指しているが、こうした課題への対処が念頭にあると考えられる。

付録2.4 個人情報保護と活用に関する国内外の状況

①各国の個人情報保護法制

高度にIT化された状況で、情報提供者の個人情報を明確に保護する必要性が高まり、1980年に採択されたOECDプライバシーガイドラインを元に、国ごとに個人情報保護法の整備が進められた。欧州（EU）では、個人情報保護の権利を基本的人権の一つとして位置づけ、2018年5月に施行されたGeneral Data Protection Regulation（GDPR）をはじめとする強力なデータ保護法制度が整備された。米国でも、カリフォルニア州消費者プライバシー法（California Consumer Privacy Act：CCPA）など、州法レベルのプライバシー保護法がいくつかの州で存在するほか、2022年には、最終的に連邦議会では審議されなかったものの、米国初の連邦レベルの個人情報保護法案としてAmerican Data Privacy and Protection Act（ADPPA）がエネルギー・商業委員会から下院議会に送付された。

日本では、2003年に個人情報保護法が制定され、2015年の改正により個人情報保護委員会が発足し個人情報保護法上の権限が一元化されるなど、GDPRを中心とした海外の制度との整合性を持たせてきた。一方で、過剰な保護により、必要な情報にアクセスしづらく研究開発が滞るなど、情報提供者である個人・市民が享受し得るメリットを失う可能性もあり、適切な保護と活用の衡平を踏まえた制度設計が求められ、改正が進められている。例えば、2021年改正で3本の法律（個人情報保護法、行政機関個人情報保護法、独立行政法人等個人情報保護法）が1本に統合されるとともに、地方公共団体間にも全国的な共通ルールを規定し全体の所管が個人情報保護委員会に一元化されることで、情報取得主体により個人情報保護のルールが異なる、いわゆる「2,000個問題」是正を進める足掛かりができた。また、2020年/2021年の改正で「他の情報と照合しない限り特定の個人を識別することができないよう加工された情報」である仮名加工情報が新たに創設されたことにより、それまでの匿名加工情報で問題となっていた、原情報との関連付けが不可能（薬事申請に使えない）、他のデータとの突合（名寄せ）が不可能、といった点を解消するための道筋ができた。

②医療情報活用に関わる法制度

個人情報の中でも、ライフサイエンス・臨床医学分野の研究開発で必要とされる医療情報は、差別や偏見につながるおそれのある機微な情報であることから、日本では、2017年5月30日に施行された改正個人情報保護法において要配慮個人情報（法第2条3項）に位置づけられた。要配慮個人情報は原則として、取得や第三者提供には本人の個別同意（オプトイン）が必要となる。大規模なグループ病院や保険者が存在する

米国や、国が管理する欧州と異なり、わが国では民間主体の各医療機関がそれぞれ医療情報を一次保管するため、大規模なデータベースを構築するには医療情報を第三者提供する形で集約することが避けられないが、医療機関において通常診療の中で個別同意を得ながら診療情報を収集・第三者提供することはハードルが高い。

米国や欧州の法制度や法解釈の議論を見ると、国際競争力の強化を念頭に、医薬品・医療機器開発も含めた医療の研究開発を目的とする、個別同意なしでの健康・医療情報の移転・利用を広く認めた法制度・法解釈となっている。

米国では、「健康に関する研究を行う上での健康情報の果たす重要な役割を認識し、プライバシー保護が研究者のデータへの持続的アクセスを妨げてはならない」といった趣旨を含んだ決議を連邦議会が行なっている^[30]。また、健康情報の保護と合法的な使用・開示の概要を規定するHealth Insurance Portability and Accountability Act (HIPAA) の中では、「研究」は「一般化できる知識の発展に貢献するようデザインされた系統的調査」と定義され^[31]、非常に広い範囲をカバーしている。

欧州についても、一般的にはGDPRはプライバシー保護重視と認識されているが、医療や公衆衛生の分野において広大な例外規定がみられる。前文において「特別な種類の個人データの取扱いは、データ主体の同意なしに、公衆衛生（健康に関する全ての要素）の分野における公共の利益を理由として必要となることがある」、例外として個別同意が不要となる規定として9条2(h) 項に「医療制度又は社会福祉制度及びそのサービス提供の管理のために取り扱いが必要となる場合」、9条2(i) 項に「医療及び医薬品もしくは医療機器の高い水準の品質及び安全性を確保するという公衆衛生の分野において、公共の利益を理由とする取り扱いが必要となる場合」が規定されている^[32]。EU内で国境を越えて共同利用するデータベースであるEuropean Health Data Space (EHDS) 構築といった施策はそうした法制度に基づいたものである。

わが国の個人情報保護法でも、第一条においてその目的を「個人情報の適正かつ効果的な活用が新たな産業の創出並びに活力ある経済社会及び豊かな国民生活の実現に資するものであること…＜中略＞…に配慮しつつ、個人の権利利益を保護する」と記載している。欧米と同様、医薬品・医療機器開発も含めた医療の研究開発目的での個人情報の活用が認められるとも解釈可能であるが、18条、20条、27条に規定されている個人情報に係る制限の例外（本人同意なしでの個人情報の目的外使用と第三者提供、要配慮個人情報の取得）が認められる条件として、「学術研究機関等が収集する／提供される、かつ学術研究目的での使用の場合」といった趣旨が規定されていることから、日本ではアカデミアが使用主体の場合のみ例外を認める運用となっている（共同研究であれば、民間でもアカデミアからの取得や提供を受けることは可能）。他の例外の条件として「公衆衛生の向上を目的とする場合」もあるが、本人の同意を得るのが困難な場合に限られることから、実態としては適用されることは少ない。また、27条2項にはオプトアウトでの個人情報の第三者提供を可能とする規定があるものの、要配慮個人情報である医療情報では認められていない。そのため、民間の医療機関も含む学術研究や成果の商業化を伴う創薬や医療機器開発における医療情報利活用にとっては、大きな阻害要因となっていた。

個人情報保護法のみ準拠すると医療分野の研究開発が滞るおそれがあったことから、2018年に施行された「医療分野の研究開発に資するための匿名加工医療情報に関する法律（次世代医療基盤法）」により、匿名加工情報の利用を前提として、個別同意なし（丁寧なオプトアウト）で認定匿名加工医療情報作成事業者（『認定事業者』）に医療情報を第三者提供できる仕組みが生まれた。匿名加工情報では、希少疾患患者の情報やゲノム情報が扱えない、原情報との関連付けが不可能なため薬事申請に使えない、といった問題があったが、2023年改定で仮名加工情報も扱えるようになり、創薬・医療機器開発への活用拡大が期待される。これらの法律により二次利用における流通は改善された一方で、医療情報の一次利用においても法的問題が指摘されている。例えば、専門家でないと医療情報の適切な利用かどうか判断が難しいにもかかわらず、本人同意さえ取れば使用制限を受けないという同意偏重の運用方法である、認知症高齢者の診療、災害時などに個人の医療情報の取得・共有が必要な場合でも「公衆衛生向上」「本人の同意取得が困難」の例外に当

てはまるかを医療機関等の現場で判断することは困難、といった問題点である。そこで、医療分野の特性に配慮して医療情報の保護と（一次利用・二次利用含めた）活用を規定した「医療情報基本法」が必要、との意見もある。

③健康・医療情報の活用に対する社会（個人・患者、医療従事者）受容

個人情報/プライバシーの保護と情報活用についての考え方は、欧州はプライバシーを人間の尊厳の現れの一つと捉え、米国はプライバシーを自由に根差すものと解すると指摘されるように、米国と欧州でも異なる^[33]。欧州は、個人情報保護に対する権利を基本的人権として扱う一方、公共の福祉・公益により個人の権利が制約を受けうるという考え方である。また、学問の自由も欧州を中心に発展した概念であり、学術研究利用を広く認める背景となっている。米国は、個人情報を含めた私的な問題について政府も含めた他社からの干渉なしに当事者が自由に選択できることを重視しており、欧州のGDPRのような包括的な個人情報保護法が存在しない要因にもなっている。一方、活用については、経済面含めた合理性を重視していると考えられる。そうした欧米の状況に対して、わが国の個人情報保護法制は基本的には欧州の考え方を踏襲している一方、個人や民間においてはプライバシー権を米国的に捉えられることも多い。こうした日本の状況が、一貫した議論や制度設計が難しい背景となっていると考えられる。

日本人の特性として安心・安全を重視すると言われているが^[34]、個人情報の保護と活用に関する意向においてもその特性が反映されている。例えば、総務省の調査によると^[35]、国が定めるプライバシー・パーソナルデータ保護に関する規制やルールでは、便利・快適さの享受より安心・安全を重視すべきとの回答が他国より多い。また、企業等のサービス・アプリにおいて個人情報提供を求められた場合、データ流出がないこと、プライバシー保護、悪用されないこと、データ二次提供先が明確なこと、など他国より多くの条件を求める傾向がある。

一方、健康・医療情報については、活用への意識・意向が必ずしも低いわけではないことを示唆する結果が数多くある。例えば、前記の総務省調査において、病歴・病状に関する情報提供への不安は他国と比べて低い（他のデータ種と逆の傾向）。また、研究への活用においても、「医学研究と医療の改善」を目的としたデータ活用の意思は欧州主要国と同程度に高いとの調査結果がある（『安全が確保される前提』との設問であるため、前記の総務省調査とは矛盾はない）^[36]他、バイオバンク（BioBank Japan）への診療情報提供について、同意しなかった患者は10%に満たない。こうした傾向は、日本の医療保険制度、および医療従事者への信頼が背景にあると推測される。次世代医療基盤法における診療情報提供に関して、一般の人にインターネットアンケートすると20%がオプトアウトするのに対し、ある医療機関において患者を対象に調べたところ0.05%だけがオプトアウト、99.95%が情報を提供すると回答した^[37]、との結果はこの推測を支持するものである。

こうした個人・患者側の状況に対して、むしろ医療機関・医療従事者側が過度に個人情報提供のリスクを避けていることが、活用が進まない理由の1つではないかと指摘する意見もある。この背景として、前述の法的問題（個人情報に係る制限の例外かどうか現場で判断できない）に加え、医療倫理が影響を及ぼしていると考えられる。医師の職業倫理について書かれた宣誓文である「ヒポクラテスの誓い」、ならびに世界医師会（WMA、1947年に創設）が1948年に採択した「ジュネーブ宣言」には、「患者の健康・利益を第一に考慮し」、「患者の秘密を尊重する」といった趣旨の記載があり、原則として患者の医療情報は外に提供すべきではないものであった。一方、2017年のジュネーブ宣言の全面見直しでは、医師が果たすべき役割として「患者の利益と医療の進歩のために自身の知識を他者に共有すること」との文言が加わり、インターネットの普及など科学技術の革新・著しい社会の変化を受けて、診療や臨床研究を通して得られた知識をお互いに共有することが新たに求められることとなった。健康・医療情報の活用に対する個人・患者の意向、時代に即した医師に求められる役割の変化、といった点を踏まえ、医療機関・医療従事者側に対して医療情報活用の意識醸成を行うことが必要ではないかと考えられる。それと合わせて、次世代医療基盤法が広がらない理由として

付録2
わが国における問題点
の補足説明、国内外の
関連する状況など

挙げられる、医療機関のIT設備のコスト、丁寧なオプトアウトの手間などを始めとして、医療機関・医療従事者側の負担を軽減することも求められる。

付録3 専門用語説明

RWD

Real World Data

FDAにおいては「様々な情報源から日常的に収集された患者の健康状態および/または医療の提供に関するデータ」と定義されるが、本戦略プロポーザルでは、対象は必ずしも患者に限定されず健常者も含んでいる。情報源としては、医療機関で収集される電子カルテやレセプトのデータのほか、ウェアラブル端末やデジタルヘルスツールなどから収集されるデータも含む。RWDの解析により抽出した、医薬品・医療機器の有効性やリスクに関連するエビデンスのことを、RWE（Real World Evidence）と呼ぶ。

PHR

Personal Health Record

狭義には、医療機関や自治体などで医療・ヘルスケアサービスの中で収集した患者の情報を、患者自身が閲覧できる仕組み・ツールを意味する。より広義の意味で使われることもあり、例えばPHR普及推進協議会では、（医療機関や自治体などで収集するデータに限らず）個人の生活に紐づく医療・介護・健康等に関するデータを本人の判断のもとで記録・閲覧・活用する仕組み、と定義している。本戦略プロポーザルでは、データ流通にフォーカスした狭義に近い意味で扱っている。

NDB

National Database

日本語の名称は「レセプト情報・特定健診等情報データベース」。「高齢者の医療の確保に関する法律」を根拠に、2009年から構築・運用されている。保険者より厚生労働省へ「電子化されたレセプト情報（外来・調剤・手術・介護の報酬請求明細）」「特定健診・特定保健指導情報」が提供され、個人を特定できない形でデータベース化されている。90%以上のレセプトをカバーしており、世界的に見ても高い悉皆性が特徴のデータベースである。汎用性の高い基礎的な集計結果を「NDBオープンデータ」として公表し誰でも自由に利用できるほか、利用申請により、研究者だけでなく民間企業も二次利用も可能である（申請した目的での利用のみ）。

匿名加工情報/仮名加工情報

個人情報保護法で定められている用語である。いずれも特定の個人を識別することができないように個人情報を加工したものであり、個人を特定できる記述（氏名、住所など）や個人識別符号（DNA配列、マイナンバー、被保険者番号など）を除くことは共通である。主な違いは、1) IDなどを利用して他のソースの情報と連結（名寄せ）可能かどうか、2) 希少疾患のような特異性のある記述を除くかどうか、である。1)の連結可能性については、患者の詳細な診療行動（一連の医療機関・訪問時間のセット）、精細な非圧縮医用画像のように、別な情報ソースと1:1で照合できる解像度の高いデータを含む場合は連結可能と解釈される可能性も指摘されている。そのため、匿名加工情報での質の高い医学研究開発は困難であり、そうした意見を反映した結果、2023年の次世代医療基盤法改正において仮名加工情報が扱えるようになった。

データウェアハウス

さまざまなデータソースからデータを集め、ある程度定めた用途のために構造化したデータを収納したデータベース。収納したデータの検索や分析の機能も含めることもある。比較でよく用いられる「データレイク」は、

生データと非構造化データを中心に、データ保管を主目的としたデータベースである。データウェアハウスはデータレイクと比較して、データ利用に必要な加工・処理が少なく高速化できる一方、用途の柔軟性には劣る。

PPIP

Patient and Public Involvement Panel

日本語訳は「研究への患者・市民参画」で、医学研究・臨床試験プロセスの一環として、研究者が患者・市民の知見を参考にすることを指す。研究参加者との意見交換やヒアリングが主な実施形態であるが、対象は研究参加者とは限らない。例えば、第二期SIP「AIホスピタル」と連携して、AIを利用したヘルスケアについての信頼構築を試みた大阪大学・オックスフォード大学共同の「AIDEプロジェクト」は、一般市民や患者を募ってワークショップを開催、医療AIに対する期待や懸念について特定・グルーピングを行った^[38]。研究者にとっては患者の不安・懸念点を理解・解消することになる一方、患者にとっても研究が身近になり関心を高めるにつながると期待される。

参考文献

- [1] The Lancet. 378, 9796(2011), 1049. DOI: 10.1016/S0140-6736(11)61223-3.
- [2] 厚生労働省、地域医療構想及び医師確保計画に関するワーキンググループ (2022).
<https://www.mhlw.go.jp/content/10800000/000905110.pdf>
- [3] 厚生労働省、健康・医療・介護情報利活用検討会 医療情報ネットワークの基盤に関するワーキンググループ (2022).
<https://www.mhlw.go.jp/content/10808000/000938862.pdf>
- [4] 科学技術振興機構研究開発戦略センター、「調査報告書：健康・医療トランスフォーメーション 科学技術・イノベーションの潮流」、CRDS-FY2023-RR-03 (2023年8月) .
<https://www.jst.go.jp/crds/report/CRDS-FY2023-RR-03.html>
- [5] 厚生労働省、諸外国における医療情報の標準化動向調査報告書 (2019).
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryoku/iryoku/johoka/207754.html
- [6] Blanca “Didi” Davis and Alan Swenson, Journal of Digital Imaging 35, 4(2022): 812–816. DOI:10.1007/s10278-021-00538-y.
- [7] M. Roberts, D. Driggs, M. Thorpe et al., Nat Mach Intell 3(2021): 199–217. DOI:10.1038/s42256-021-00307-0.
- [8] 科学技術振興機構研究開発戦略センター、「調査報告書：人工知能研究の新潮流2 ～基盤モデル・生成AIのインパクト～」、CRDS-FY2023-RR-02 (2023年7月) .
<https://www.jst.go.jp/crds/report/CRDS-FY2023-RR-02.html>
- [9] American Medical Association, Practice Management(2023).
<https://www.ama-assn.org/practice-management/digital/educate-patients-about-misleading-ai-generated-medical-advice>
- [10] M. Moor, O. Banerjee, Z.S.H. Abad et al., Nature 616(2023): 259–265. DOI:10.1038/s41586-023-05881-4.
- [11] 日本医師会 生命倫理懇談会、「令和2・3年度答申：『医療AIの加速度的な進展をふまえた生命倫理の問題』について」(2022) .
https://www.med.or.jp/dl-med/teireikaiken/20220309_3.pdf
- [12] S. Gunasekar, Y. Zhang, J. Aneja et al., arXiv (2023):2306.11644. DOI:10.48550/arXiv.2306.11644.
- [13] 佐藤智太郎『ITヘルスケア誌』17巻1号(2022): 12. DOI: 10.11204/ithc.17.12.
- [14] R. Koopman, K. Kochendorfer, J. Moore et al., Ann Fam Med 9, 5(2011): 398-405. DOI: 10.1370/afm.1286.
- [15] M. N. Cabili, J. Lawson, A. Saltzman et al., Cell Genomics 1, 2(2021): 100031. DOI:10.1016/j.xgen.2021.100031.
- [16] Y. Joly, C. Dupras, M. Pinkesz et al., Annu Rev Genom Hum Genet 21(2020): 491–507. DOI:10.1146/annurev-genom-111119-011436.
- [17] 現在の状況はGenetic Discrimination observatoryなどを参照
<https://gdo.global/en/gdo-map-approaches>
- [18] B. A. Areheart and J. L. Roberts, The Yale Law Journal 128, 3(2019): 710-790.

https://www.yalelawjournal.org/pdf/AreheartRoberts_a2gvpzai.pdf

[19] NIH Funding RFA-AG-24-009.

<https://grants.nih.gov/grants/guide/rfa-files/RFA-AG-24-009.html>

[20] 科学技術振興機構研究開発戦略センター、「科学技術未来戦略ワークショップ報告書：個人データ駆動医療・ヘルスケア」、CRDS-FY2022-WR-12（2023年3月）。

<https://www.jst.go.jp/crds/report/CRDS-FY2022-WR-12.html>

[21] Nuance 社ブログ記事 (2022)。

<https://whatsnext.nuance.com/en-gb/healthcare-ai-uk/tackling-clinical-documentation-more-efficiently/?cid=7016T000002DBDaQAO>

[22] 井上悠輔, 松井健志, 山本圭一郎『医学のあゆみ』276, 9(2021): 871-875.

[23] 武田薬品工業社ニュースリリース (2021)。

<https://www.takeda.com/ja-jp/announcements/2021/ubie>

[24] E. J. Williamson, A. J. Walker, K. Bhaskaran et al., Nature 584(2020): 430–436. DOI:10.1038/s41586-020-2521-4.

[25] 山本隆一、『情報管理』60, 9(2017): 619-628. DOI:10.1241/johokanri.60.619.

[26] 平成27年度日本医療研究開発機構医薬品等規制調和・評価研究事業、「治験活性化に資するGCPの運用等に関する研究」分担(2016)。

<https://www.amed.go.jp/content/000004760.pdf>

[27] G. Kaissis, A. Ziller, J. Passerat-Palmbach et al., Nat Mach Intell 3(2021): 473–484. DOI:10.1038/s42256-021-00337-8.

[28] K. Muto, A. Nagai, I. Ri et al. J Hum Genet 68(2023): 579–585. DOI:10.1038/s10038-023-01163-z.

[29] 日本医師会、「遺伝情報・ゲノム情報による不当な差別や社会的不利益の防止」についての共同声明(2022)。

<https://www.med.or.jp/dl-med/teireikaiken/p20220406.pdf>

[30] Institute of Medicine, "Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research." The National Academies Press(2009). DOI: 10.17226/12458.

[31] 45 CFR 164.501.

<https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.501>

[32] 日本語訳は個人情報保護委員会ウェブサイトを参照

<https://www.ppc.go.jp/files/pdf/gdpr-provisions-ja.pdf>

[33] 国立国会図書館、『調査と情報—ISSUE BRIEF—』1216(2023)。

<https://dl.ndl.go.jp/view/prepareDownload?itemId=info:ndljp/pid/12557940>

[34] 三菱UFJリサーチ&コンサルティング社、『季刊政策・経営研究』2008 vol.2など <https://www.murc.jp/uploads/2012/07/0126.pdf>

[35] 総務省、データ流通環境等に関する消費者の意識に関する調査研究(2020)。

https://www.soumu.go.jp/johotsusintokei/linkdata/r02_04_houkoku.pdf

[36] NISTEP、DISCUSSION PAPER 205(2022)。

<https://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-DP205-FullJ.pdf>

[37] 吉田 真弓、田中 勝弥、山本 隆一、第39回医療情報学連合大会 一般口演(2019)、2-H-1-02.

https://confit.atlas.jp/guide/event-img/jcmi2019/2-H-1-02/public/pdf_archive?type=in

[38] 科学技術振興機構研究開発戦略センター、「俯瞰セミナー&ワークショップ報告書：トラスト研究の潮

流 ～人文・社会科学から人工知能、医療まで～」、CRDS-FY2021-WR-05（2022年2月）.
<https://www.jst.go.jp/crds/report/CRDS-FY2021-WR-05.html>

なお、URLは2024年1月23日時点のものである。

作成メンバー

総括責任者	谷口 維紹	上席フェロー	ライフサイエンス・臨床医学ユニット
	永井 良三	上席フェロー	ライフサイエンス・臨床医学ユニット
リーダー	宮薊 侑也	フェロー	ライフサイエンス・臨床医学ユニット
リーダー (2022.4~)	戸田 智美	フェロー	ライフサイエンス・臨床医学ユニット
メンバー	小泉 聡司	フェロー	ライフサイエンス・臨床医学ユニット
	辻 真博	フェロー	ライフサイエンス・臨床医学ユニット
	茂木 強	フェロー	システム・情報科学技術ユニット
	中村 輝郎 (~2022.3)	フェロー	ライフサイエンス・臨床医学ユニット
	荒岡 礼 (~2022.3)	フェロー	ナノテクノロジー・材料ユニット
	坂本 明史 (~2022.3)		戦略研究推進部
	相馬 りか (~2022.3)		戦略研究推進部

戦略プロポーザル

CRDS-FY2023-SP-04

健康・医療リアルワールドデータ利活用基盤の構築と生成AIへの展開

STRATEGIC PROPOSAL

Health & medical real world data infrastructure

- Catalyst of generative AI development in Japan -

令和 6 年 3 月 March 2024

ISBN 978-4-88890-895-5

国立研究開発法人科学技術振興機構 研究開発戦略センター

Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町7 K's 五番町

電話 03-5214-7481

E-mail crds@jst.go.jp

<https://www.jst.go.jp/crds/>

本書は著作権法等によって著作権が保護された著作物です。

著作権法で認められた場合を除き、本書の全部又は一部を許可無く複写・複製することを禁じます。

引用を行う際は、必ず出典を記述願います。

なお、本報告書の参考文献としてインターネット上の情報が掲載されている場合には、本報告書の発行日の1ヶ月前の日付で入手しているものです。上記日付以降の情報の更新は行わないものとします。

This publication is protected by copyright law and international treaties.

No part of this publication may be copied or reproduced in any form or by any means without permission of JST, except to the extent permitted by applicable law.

Any quotations must be appropriately acknowledged.

If you wish to copy, reproduce, display or otherwise use this publication, please contact crds@jst.go.jp.

Please note that all web references in this report were last checked one month prior to publication.

CRDS is not responsible for any changes in content after this date.

FOR THE FUTURE OF
SCIENCE AND
SOCIETY



CRDS

<https://www.jst.go.jp/crds/>

