

我が国の医療データの 二次利用における 具体的な課題

東京大学大学院医学系研究科

医療情報学分野

教授 大江和彦

kazuhiko.ohe@gmail.com

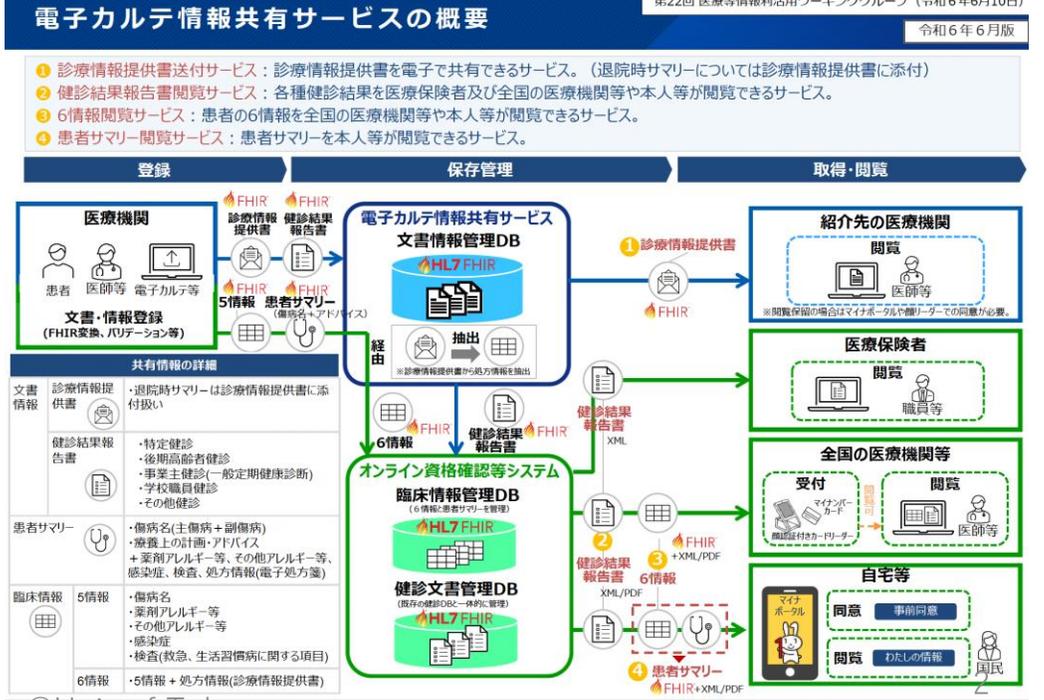
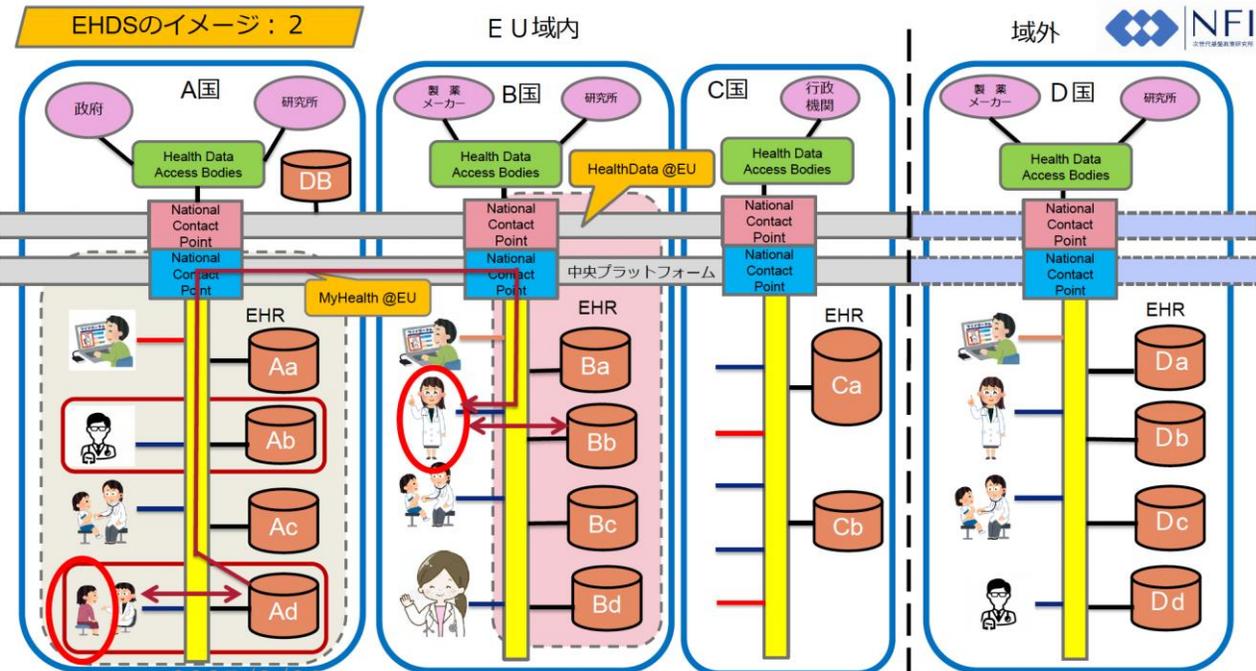
医療データの二次利用におけるEHDSと日本との比較

- EHDSでも日本でも、医療機関のEHR（いわゆる電子カルテ）データをデータ源として、それを標準化、流通可能化した上で、一次利用、二次利用に供することを狙っている。

多施設間一次利用の場合

- EHDS：ひとつの国内のすべての医療機関のEHRに National Contact Point (NCP) からアクセスできる環境実現を想定している（と見受けられる）
- ユースケースごとの必要データPULL型

- 日本：電子カルテ情報共有サービスに医療機関が、最低限情報（3文書6情報）をPUSHし、サービス主体経由で参照可能とする。
- プレフィックスデータPUSH型



第22回 医療等情報利活用ワーキンググループ（令和6年6月10日）
令和6年6月版

2025/2/26

Copyright(C) 2024 Next Generation Fundamental Policy Research Institute (NFI) All Rights Reserved.

Copyright(C)2025 Kazuhiko Ohe@Univ.of Tokyo

医療データの二次利用におけるEHDSと日本との比較

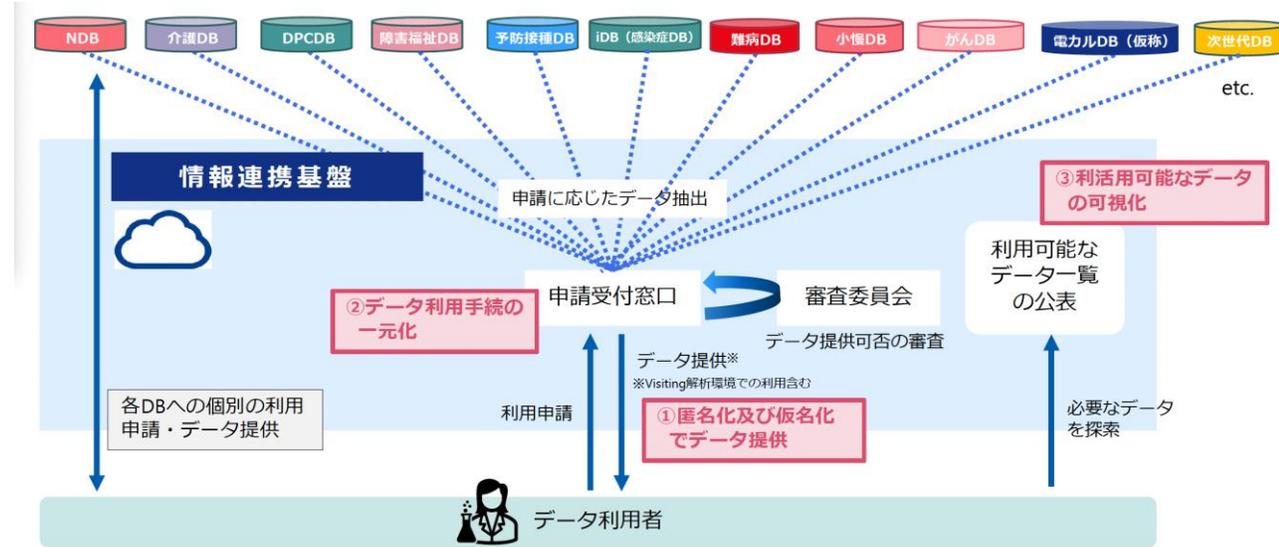
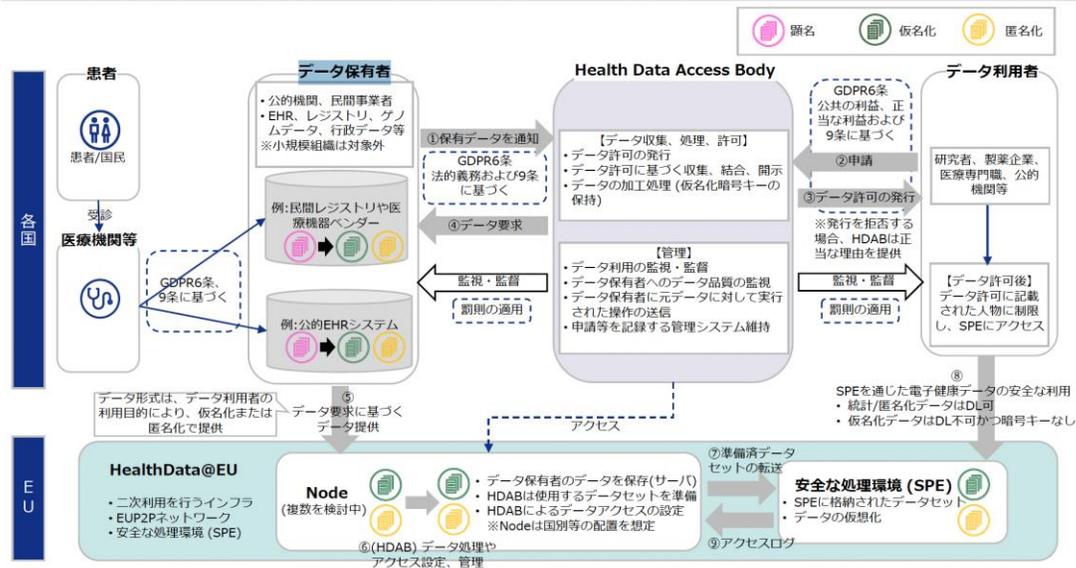
・ 二次利用の場合

◇EHDS: Health Data Access Bodies (HDAB)が「複数のデータ保有者」のデータへのアクセス妥当性のチェックと解析環境サービス提供を想定している。

- 日本: ポータル窓口機関が、データ利用申請に応じて必要なデータセットを抽出・連結して匿名化or仮名化して提供
- 解析環境サービスの提供は明確でない

EHDS法案における二次利用 (想定)

HealthData@EUは、各国にHealth Data Access Bodyを設置し二次利用におけるデータ収集やデータ許可を発行する仕組みである



<https://www.mhlw.go.jp/content/12401000/001327545.pdf>

第185回社会保障審議会医療保険 部会 資料2

データベースを使用した臨床研究での実際例と課題

1. RSウイルス感染症罹患児の5歳未満における喘息発症リスクの分析
 1. DB: 民間企業提供レセプトデータ、対象：病名=RSウイルス感染症、年齢=5歳未満
 2. 変数：喘息の有無、月齢、性別
 3. 調整した入手困難な変数：母親の喘息の既往の有無、喫煙歴
2. 緩徐進行1型糖尿病に対するSGLT2阻害薬による進行抑制
 1. 糖尿病患者のうち「最初はインスリンに依存していないが、次第にインスリン依存していて実は1型糖尿病だった」と、のちに判明するタイプ（緩徐進行1型糖尿病）では、正確に病名が入力されていないことが多い集団を解析対象とする。
3. 子宮頸がんに対する放射線治療による晩期有害事象の検討
 1. DB: 民間企業提供レセプトデータ10年分、対象：病名=子宮頸がん、治療=放射線治療あり
 2. 変数：年齢、喫煙、BMI、炎症性腸疾患既往、ステロイド薬内服の長期使用、糖尿病既往、骨盤内手術歴、腸閉塞既往
 3. 子宮頸がんの詳細なステージ分類情報がなく、合併症などの代替情報の利用せざるを得ない。
 1. 水腎症が病名についている（推定StageIIIB）症例を除くなど。
 2. 膀胱浸潤を疑う症例（StageIVA）を除くなど。
 4. 学会データベースなどから推定した症例数の1/3程度の症例数だった。調整変数の喫煙、BMIは欠損レコードが多い。既往（骨盤内手術既往、ステロイド薬使用、炎症性腸疾患（IBD）が推定症例数よりかなり少なかった。

医療データの二次利用における潜在的な諸課題

- 電子カルテデータを発生源とすることによる、データ種類とデータ正確性の制約
 1. 記載されていても、決められた欄に記入されていない重要な基本的患者データ(基本診療患者プロフィール)が取得できない。たとえば、身長、体重(BMI)、血圧、体温、喫煙歴、飲酒歴など。
→ 基本診療患者プロフィールに関する電子カルテ入力画面様式を国として取り決め、一連の診療に最低1回の入力と6情報での送信を。
 2. 3文書6情報では、検査結果43項目、感染症5項目と少ない。
→ たとえばPMDA MID-NETでは275項目余が標準コード化され使用できる。
 3. 診断病名(診療対象病名)が明確でない。
→ 外来1診療ごとに診療対象病名を6情報での選択指定と6情報に含めて送信を。入院では、DPC調査票様式1データの3文書6情報に追加する。
 4. 画像検査の診断報告書のデータが(電子カルテには存在するが)、血液検査結果などと同様な方法では取得できないことが多い。
→ データ形式を標準化して6情報での送信を(画像は除く)。画像はnetPDI(by IHE協会 <https://www.ihe-j.org/netPDI/>)の仕組みを参考に、診療情報提供書とともに送信、または患者にネット経由で提供する。
 5. 家庭での血圧等、自覚症状を主徴とする疾患のためのデータがない。
→ PHRデータからの事前固定PUSH収集
 6. 臨床研究に必要な疾患の臨床分類情報(重症度のグレード情報や細分化した病気のタイプ情報)が足りない。
→ 主要な診療対象疾患ごとに、記録すべき疾患分類情報(病型、臨床ステージ、病理ステージ)を取り決め、一定のカテゴリの医療機関ではその記録と病名情報との抱き合わせ送信を。

その他の課題の例

- 現状では、臨床的アウトカムで最も重要な死亡情報、特に原死因情報と、生前の臨床情報との連結解析が困難。
- 現状では、母親と子供とのレセプトにおける連結解析が困難。
- 異なる公的データベースの横断的探索的臨床データベース研究の実施が困難。
 - EHDSでいう「公的メタカタログ」が必要。また、自由にアクセスして研究計画を立案できる「公的データベースのサンプル(ダミー)データベース」の提供が望まれる。
- Visiting解析環境に、最初から潤沢な解析ツールを用意することが重要。
- 主要な医療機関、医療研究機関からVisiting解析環境に、安全に接続でき、職場環境から解析できる「(カプセル型)アクセス解析ブース」を充実させることも検討いただきたい。

AI構築に必要なマルチモーダルデータ

- コード化・数値データ
 - 画像
- } ML、DNN-DL
- Semi-Structured Data
 - Narrative Text Data
- } LLM-based AI

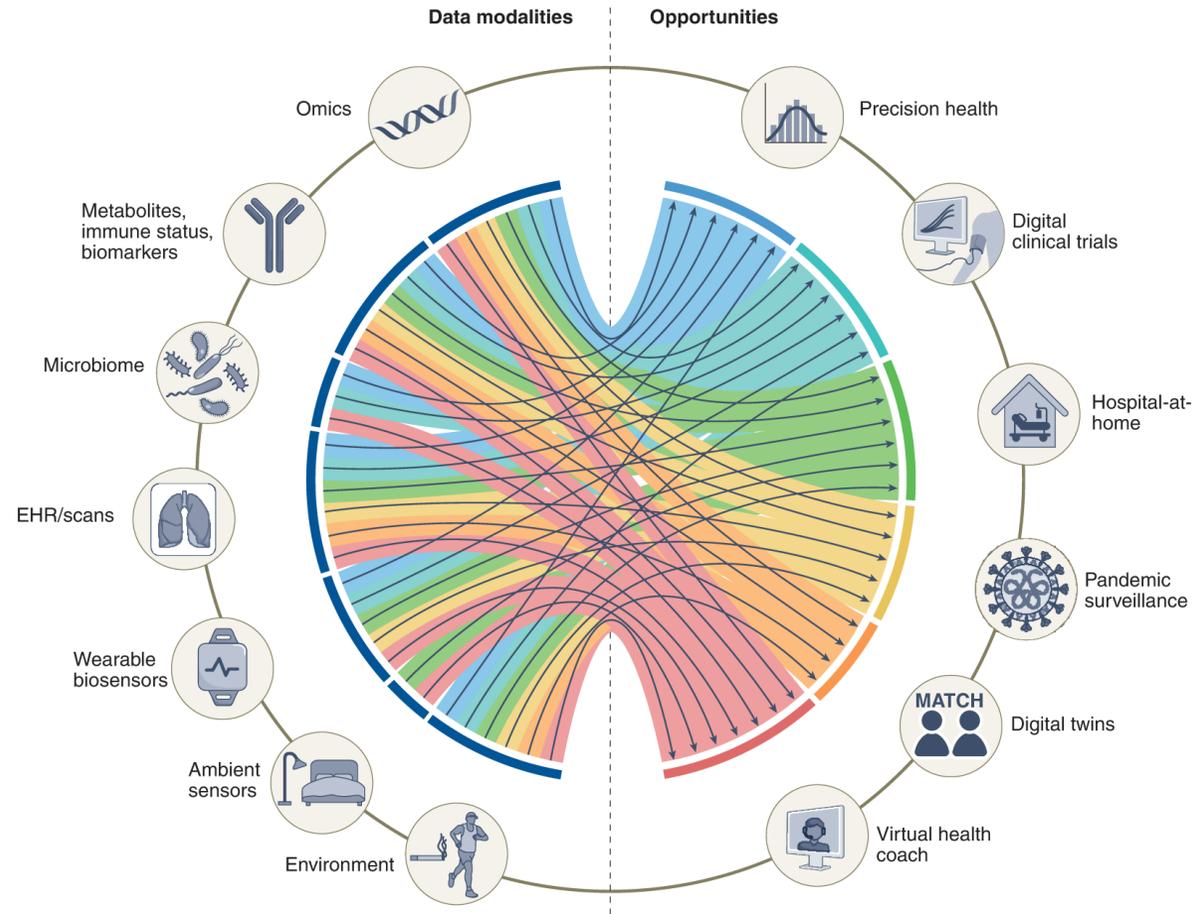


MMM-based AI
(Multi-Modal Model)



AGI

(Artificial General Intelligence)



Created with [BioRender.com](https://www.biorender.com/).

Fig. 1: Data modalities and opportunities for multimodal biomedical AI.

From: [Multimodal biomedical AI](#)

[Nature Medicine](#) volume 28, pages1773–1784 (2022)