

## がん登録データを活用するための リモート解析の実現可能性と課題

梅沢 淳

国立がん研究センターがん対策研究所

### 1. はじめに

リモートアクセスを介したデータ解析（以下、リモート解析）は、すでに多くの医学・疫学研究でも成果を挙げており、様々な法律やルール、あるいは空間的な制約を受けるがん登録データのさらなる利活用においても期待が高まっている。本稿ではリモート解析の概要、国内外での先行事例、国立がん研究センターでの新たな取り組みなどを紹介しつつ、がん登録における実現可能性と課題について整理したい。

### 2. リモート解析とは

これまでリモート解析について一般化された定義はなく、便宜上「VPN（Virtual Private Network）、リモートデスクトップといったリモートアクセス技術により、遠隔地にあるデータや解析資源をあたかも手元にあるかのように利用し、解析すること」と定義した。特に目新しい技術ではなく、近年ではコロナ禍におけるリモートワークの導入により普及が進み、同じく研究での利用も促進されているように見受けられる。

国内で運用中のリモート解析用の基盤には、NDBのVisiting環境にあたるHIC（Healthcare Intelligence Cloud）や、ACC（Asia Cohort Consortium）のリモート解析基盤等がある。データはそれぞれ、クラウド上に構築された安全なプラットフォーム上と、国立がん研究センター内で厳重に管理されているオンプレミスサーバーから外部へ持ち出されることなく、ユーザーは

様々な場所からPC一台でアクセスし、より柔軟で探索的な解析が行える。

一方、連合学習（Federated Learning）はスマホ等の予測変換や医療AI等での利用が拡大している技術である。データを所在地から動かさず、解析プログラムと結果・パラメータのみをやり取りすることができるリモート解析手法の一つであり、がん登録の分野でもRARECARE net Asiaでの取り組みが進んでいる。探索的な解析には不向きではあるが、十分に標準化されたデータに対して国境をまたいで網羅的に解析することには長けている手法と言える。

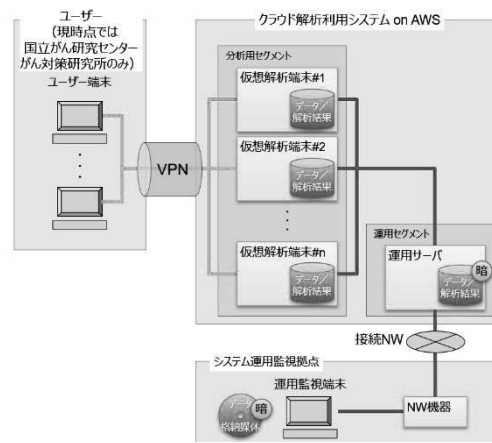


図1 試験用リモート解析システム

### 3. リモート解析環境構築の試み

2023年10月の「全国がん登録及び院内がん登録に係る課題と対応方針 中間とりまとめ」にて、CD-ROM等の媒体の郵送に変わる、安全かつ利便性の高い匿名化されたがん情報の提供方法として、リモートアクセス等を活用した情報提供体制の整備について調査研究事業において検討を進めるべきと記載されたことを受け、当センターではその試験環境の構築等を含め検討を進めている。ユーザーの手元にあるPCよりクラウド

ド上の閉鎖NWへのVPN接続と、仮想解析用端末へのリモートデスクトップ接続を行い、配置されたデータおよび統計ソフト等を用いて自由な解析を行える (Fig. 1)。

現行は試験用の簡素な構成であり、がん登録の安全管理基準を満たせるものではないものの、これまでのサンプルデータを用いた少人数による模擬解析において特に大きな課題は見つかっておらず、技術的には十分実用に耐えうるものとの見通しである。

#### 4. 課題と展望

実際に多くのユーザーに利用いただくためには、第一にルール of 整理、第二に限られたリソースの中でどのくらいの規模や自由度を目指すかという指針作りが最優先課題である。加えて、運用事務的な負荷や導入・運用コスト、また年々被害が増えているサイバー攻撃に対する対策も十分に検討し、備える必要もある。さらに、登録実務やデータ利用申請とその承認に係るフローとも足並みを揃えて、多次元的な検討 (Fig. 2) で最適解を見極めつつ、制度全体でのボトルネックの解消や最適化を図るべきである。

その第一歩として、まずはがん登録情報のアクセスポリシー (Table. 1) の整理と、制度やシステム設計への落とし込みが可能な人材の確保やWGの設立を提案したい。システム内部や外部連携部分をシンプルに構成すること、個人や組織の認証技術と適切な権限管理を行い証跡も残すこと、連合学習など別の技術との使い分けること等も有用であると考えている。今後でもできること、できないことを見極めながら、攻守兼ね備えた持続可能なデータ利活用基盤の構築、運用を通じて、様々ながん対策が実現することにエンジニアとして貢献したい。

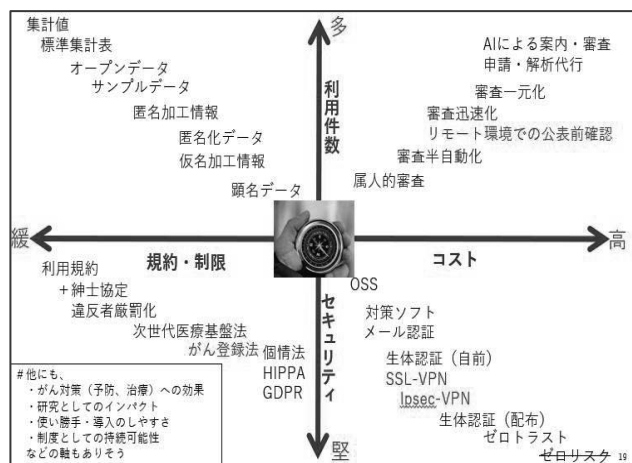


図2 多次元的な検討が必要であることを表すイメージ図

表1 がん登録情報アクセスポリシーのイメージ

データ種別	ユーザー (所属)	身元確認IAL	本人認証AAL	必要な技術
オープンデータ	共通	低	低	ウェブサイト
デモデータ	共通	低	低	ウェブサイト Github
(サンプルコード)	共通	低	低	ウェブサイト Github
匿名化データ	国内研究・医療機関	低 (~中?)	低 (~中?)	認証・解析基盤
	国内民間企業	中 (公的身分証)?	中 (2要素) ~?	認証・解析基盤
	...			
	大規模な国際共同研究			認証・解析基盤
	海外研究・医療機関	?	?	認証・解析基盤
	海外民間企業	?	?	認証・解析基盤
顔名データ	国内	中 (公的身分証) ~ 高 (対面含む確認)	中 (2要素) ~ 高 (2要素以上)	???
	海外			

# 個人認証のレベルを定め、必要な技術を整理する。届出、整理とのバランスか、統合か。既存認証基盤 (GakuNin...オン資など) との連携も検討の余地あり。