

厚生労働科学研究費補助金（がん対策推進総合研究事業）
分担研究報告書

Federated Learning によるデータシェアリング

研究代表者 松田智大 国立がん研究センター社会と健康研究センター国際連携研究部 部長
研究分担者 杉山裕美 (財)放射線影響研究所(広島) 主任研究員
研究分担者 雑賀公美子 国立がん研究センター社会と健康研究センター国際連携研究部 外来研究員

研究要旨 国内、国際共同研究での個別データを外部に一切持ち出さずに、データ集約手法を導入する。既に台湾がん登録で導入しているシステム **Federated Learning** を、大阪国際がんセンター、佐久医療センター、放射線影響研究所と国立がん研究センターとオランダ **IKNL** をつないで試験導入した。第2段階として、東京、長野、広島と大阪のがん登録データに基づいて、それぞれのがん登録の年齢罹患率の算出、4登録合計の年齢調整罹患率の算出に成功し、国がんでサーバーを立ち上げる事にも成功した。

RARECAREnet Asia でも本手法を採用予定、日台韓での接続試験済み、生存率のリモート分析検証のため、各国のデータで、**STATA** でのプログラムを実施し、「解答」の作成もすんでいる。今後は、国立がん研究センター中央病院が進める **ATLAS** 計画においても、本システムでのデータシェアについて検討する予定である。

A. 研究目的

GDPR を初めとする情報保護に関する法律が多く、多くの国で適用され、匿名であろうと、個別情報の国外提供が困難となってきた。オランダ国立総合がんセンター(**IKNL**)は、**Vantage6** というシステムを開発し、この技術は、がん登録情報の分散分析化を可能にしている。このプロジェクトの目的は、日本のがん登録情報の、集計データの計算と共有のためのシステム利用の技術的実現性を実証することである。

分析を行うために、**Vantage6** インフラを利用可能にする (図 1)

- **Vantage6 Trolltunga** サーバーへのアクセス
- 日本の担当者がシステムのインストールおよび使用に関するトレーニングを実施
- 日本の **CR** から指定された分析(水平分割データの年齢調整罹患率アルゴリズムなど)の開発
- 新しいアルゴリズムや解析の開発についての指導

B. 研究方法

本プロジェクトにおいて、**IKNL** は以下のものを提供した。

- 日本のがん登録室ががんデータの分散

C. 研究結果

本プロジェクトでは、**NCC** および日本のパートナーに対して、連携技術の利用方法、

応用方法、可能性について幅広い知識を提供することができた。

D. 考察

Vantage6 は将来的にその解決策の一つとなりうるが、現時点では、パッケージとして完成していないため、導入に一定以上の IT の知識が必要であること、各サイトの PC 環境、ネット環境、セキュリティ環境が大きく異なる中で、データ交換においてエラーが生じた場合の、問題解決が非常に難しい。また、基本的な解析のアルゴリズムは開発済みであるが、複雑な解析のアルゴリズム解析には非常に長い時間がかかり、検証も大きな負担がかかる。

本技術が全世界的に一般化されるには、こうした技術的なハードルを下げる必要がある。

E. 結論

今回のパートナーシップは、日本のがん登録関係者全員にとって刺激的で生産的なものであった。欧州、日本だけでなく、今後アジア諸国においても、情報の国外提供が困難になると考え、こうした技術の適用が想定されている。



図 3 新旧のデータ交換の方法の比較