

データを機関外にコピーせずに 分析するための Trusted Research Environment 第9回バイオバンクオープンフォーラム

2025.8.1

東京大学企画情報運営部 医療リアルワールドデータ活用人材育成事業 特任講師
岡田 昌史

バイオバンク・病院内データの利用形態

- ・ バイオバンク内データ・病院で発生するリアルワールドデータを利用するための方法
 1. 匿名加工情報等に加工作した上で、データセット自体を利用者にコピーして渡す
 2. 研究プロトコルを倫理審査に通して、許可された項目・範囲のみをEDCに入力する
 3. バイオバンク・病院内部の研究者が自らもしくは委託を受け、必要に応じて研究プロトコルを倫理審査に通して、許可された項目・範囲のデータを切り出して利用する
- ・ 自機関内では柔軟にデータ取得・解析が可能だが、他の研究機関がデータを利用しようとすると敷居が上がる

なぜ「敷居が上がる」のか？

- データ収集主体機関にいる研究者がデータを研究利用することは、多くの場合同意が得られている
- 「外の機関」へのデータの第三者提供も適切な手続きを踏めば可能ではあるが、一度外にコピーされた情報は制御が効きにくくなる
 - ▶ 対策1) 個人特定ができない程度に情報を落とし、コピーが出回っても問題が起こらないようにする
→ **解析に耐えなくなるおそれがある**
 - ▶ 対策2) 提供記録・保管条件等を厳密に定め、コピーされたものもオリジナルと同等に管理されるようにする
→ **コピー自体が欲しいわけではないのに、管理が大きな負担に**

機関外の研究者が機関内の研究者と同様に解析をできるようにするには？

- 以下のような条件が満たされるなら、コピーをしないことを条件に、機関外の研究者にも機関内と同程度のデータアクセスを許してもいいのでは？ という考え方が出てきた
 - 研究者は研究倫理・研究手法について十分な知識をもち、データの収集目的外の利用をしない
 - 研究プロトコルはデータの収集目的からみて適切である
 - データが機関外にコピーされないよう、技術的に安全が担保される
 - 研究に明らかに不要な個人情報はいあらかじめ削除する
 - 解析結果から個人が特定できるようなことがないように確認がされる

Five Safes

- **Safe People** – データを利用する研究者が適切なトレーニングを受けていて、規範にのっとって行動すること。
- **Safe Projects** – 研究計画自体が法的・倫理的に適切であること。
- **Safe Settings** – データへのアクセス環境が物理的・技術的に安全であること。
- **Safe Data** – データ自体が仮名化・匿名化等により個人識別リスクを低減させていること。
- **Safe Outputs** – 解析結果に個人識別できる情報が含まれないこと。
- 機微なデータへのアクセス可否を考える上でのフレームワーク
Desai, T., Ritchie, F. & Welpton, R. Five Safes: Designing data access for research. at <https://uwe-repository.worktribe.com/output/914745>.
- これらの評価軸は独立して評価され、すべての評価軸を総合して、安全かどうかを判断する

Five Seven Safes

- **Safe Computing** – Safe settingsはこれまで、オンプレミス環境やプライベートクラウド環境であることが求められてきたが、パブリッククラウド上であっても十分に注意深い設計: Safe computing を実装すればSafe settingsとみなせる。
- **Safe Return** – 解析結果を個別の患者の治療に活かすことを目的として匿名化が行われない場合がある。この場合に誤りなく患者と解析結果を結びつけること。

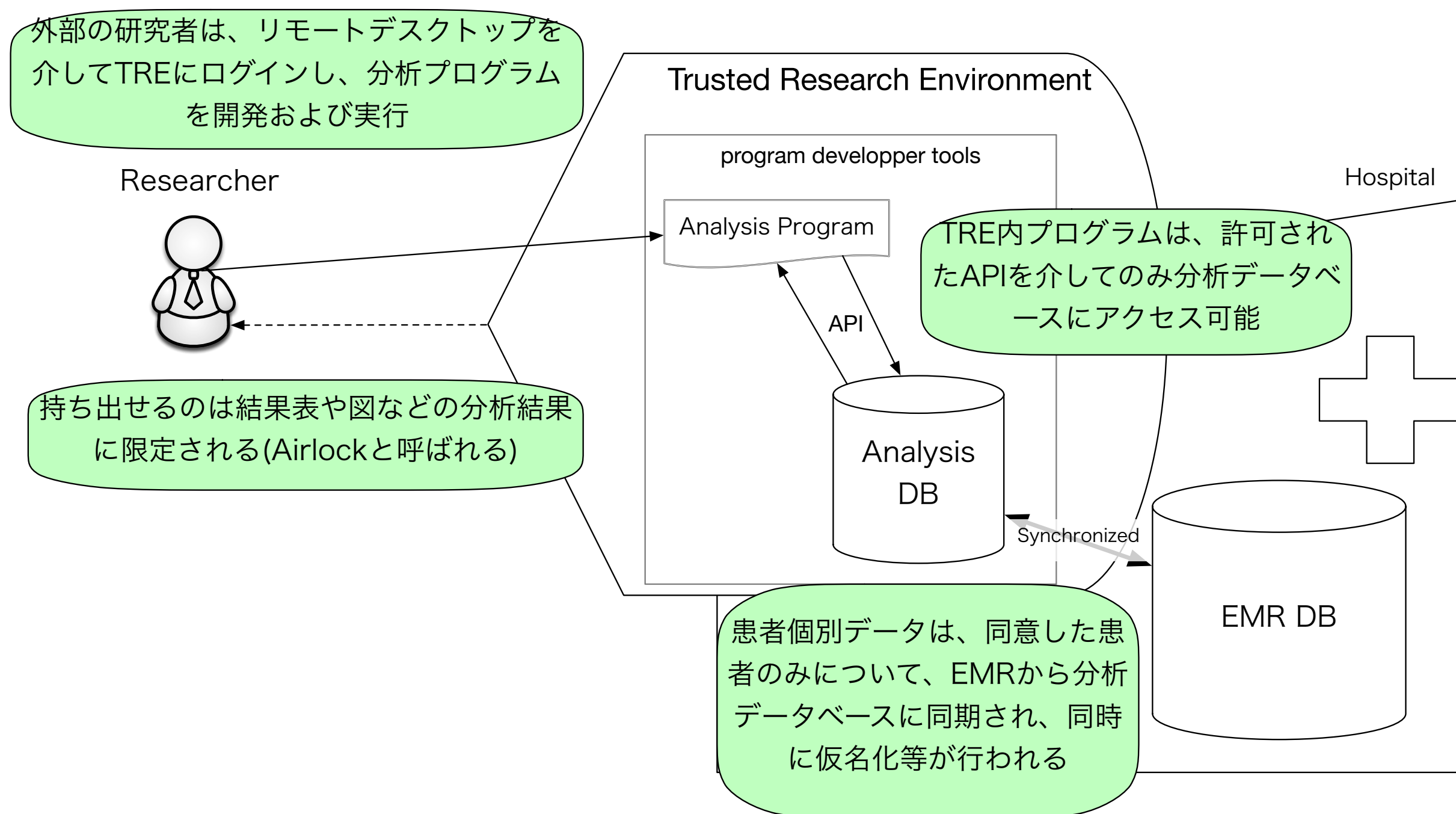
Trusted Research Environment (TRE)

- Five Safesのうち、Safe Settings, (Safe Computing), Safe Dataを実現するための技術的手段として、Trusted Research Environment (TRE)が提案され、英国を中心に普及した。

UK Health Data Research Alliance. Trusted Research Environments (TRE) Green Paper: A strategy to build public trust and meet changing health data science needs. https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf (2020).

- 研究者がバイオバンク・医療機関等からデータをコピーすることなく、ほぼ加工されていないデータを分析することを可能にする
- 少し前はオンサイトセンターとして物理的にSafe Settingsを実現していたが、パンデミック後ごろからリモートアクセスが主流になってきた
- TREを利用し、**かつ** Safe People+Safe Projects+Safe Outputs (もしくはSafe Return) が担保されるならば、機関外の研究者にも解析させてもよろしかろう、という**考え方**
- 考え方なので、導入する地域のコンセンサスを得て、法令に合わせて適用していくことが必要

典型的なTrusted Research Environment の構造とデータフロー



日本におけるTRE

- NDBはじめ、匿名化された大規模なデータベースでは、以前よりオンサイト環境という形でアクセス制御された解析環境を用意しており、最近ではTRE的なクラウド環境（医療介護データ等解析基盤）が導入されつつある。
- しかし、たとえTRE上で操作をしても他機関研究者がデータを分析するためには従来通りの適切な手続きを行う必要があると思われ、導入・維持費用もかかるので、個別のデータ提供機関への導入は未だ大きくは進んでいない
- データをコピーを作らなくても解析できるようにすることから、個人情報の安全性を上げる技術と考えられるため、5(7) Safesが担保される条件のもとでは、機関外研究者による解析の敷居を下げる方向での議論を期待したい。