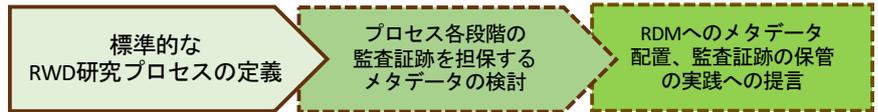


国際的なRWD研究を実現する医療情報分析基盤の検討

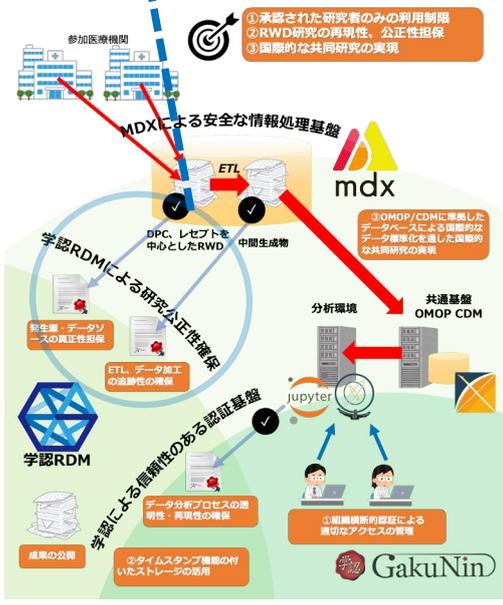
愛媛大学医学部 医療情報学講座 木村映善



Research Question
② RWD研究の再現性、公正性の担保とは？



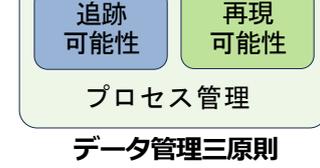
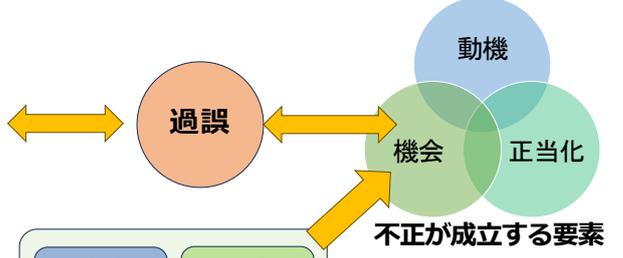
昨年の発表



問題のある研究行為

- ※文献[1]を参考に構成
- 好ましくない研究行為 (QRP: Questionable Research Practice)
- 特定不正行為 (FFP: Fabrication, Falsification, Plagiarism)

[1] 飯室 聡 研究データの公正性を担保するための基本的な考え方. Japan Open Science Summit; 2022.



追跡可能性の担保のために必要な要件

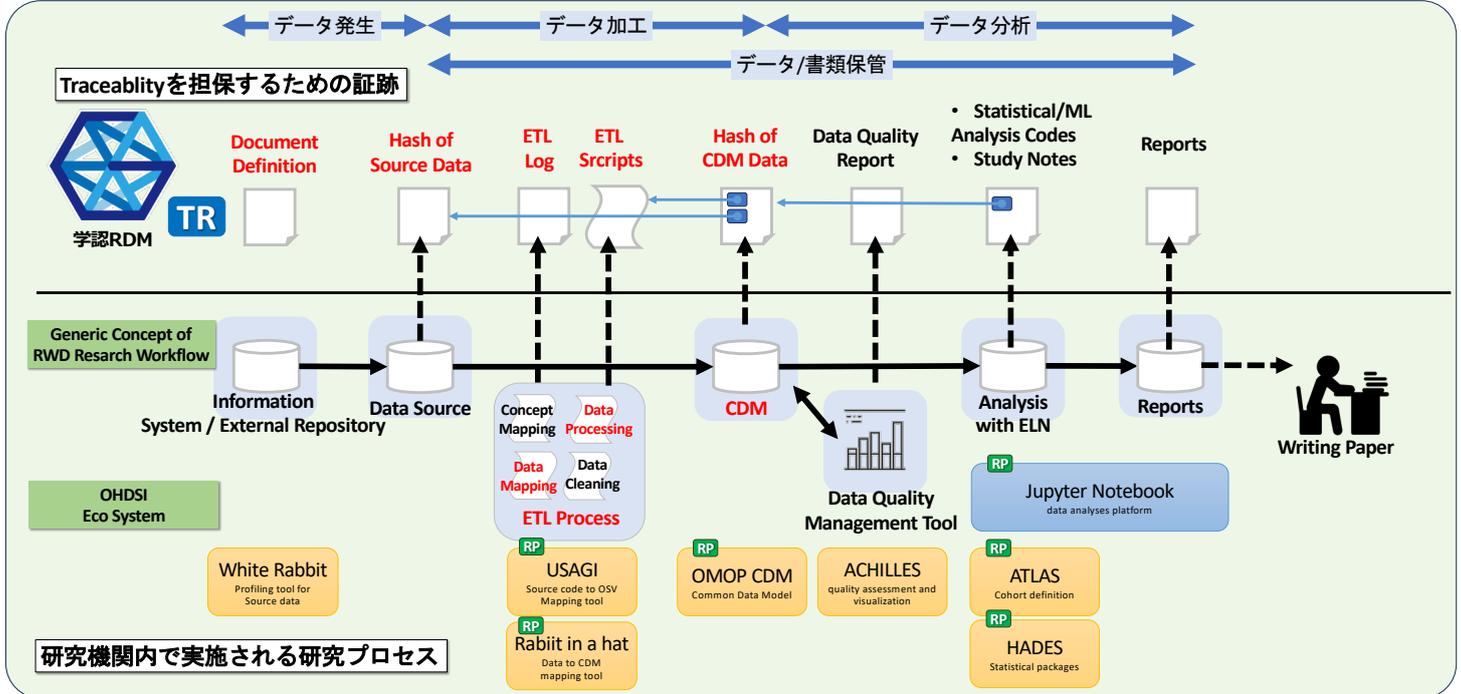
- 追跡可能性を担保できない、あるいは再現可能性からの追跡が困難なプロセスがある場合は、追跡可能性を担保する必要がある。
- 例:
 - ・ソースコードがないプロプライエタリツールで加工処理した。あるいは、その手順が残っていない
 - ・手動で加工した
 - ・外部に委託した etc...



Jupyter Notebook等で、この「ファイルを読み込んだ」というだけではNGで、読み込んだ時の時間と、ハッシュ値もNoteBook上に残す必要がある。



RWD研究の潮流となっているOHDSIのエコシステムを調査・類型化して、メタデータ定義やRDMを利用した研究ベストプラクティスの検討の基礎とする



導出した要件

- TR**
- Electronic Lab Note等から参照されるCDMについて版情報も含めて追跡できること
→ データベースを参照するのは難しいので、データベースに取りこんだテーブル用データのハッシュへ迎れること
 - CDMを構成したSource Dataに迎れるようにすること → テーブル用データから使用したETL ScriptのハッシュとSourceのハッシュへ迎れること

今後の取り組み

- ETL前後の追跡性を担保するためのメタデータの開発
ETLを実施するスクリプト自身のハッシュとソースデータ、そして変換したデータの3つのハッシュを格納したメタデータを設計し、ETLを実施するスクリプトがそのメタデータを生成するように改修する。
- CDMを引用するためのメタデータの開発
ELN等からCDMを利用する研究者が、CDMを引用するためのメタデータを設計し、上記のETL実施に自動的に生成するようにする

- RP**
- RWD研究の再現性を高めるために、OHDSIで提供されているライブラリ、フレームワークを極力活用すること。
 - 各種ツールで生成した定義ファイル、設定情報をRDMに保存すること。

- RWD研究の再現性を担保するための推奨リスト作成
OHDSIの各種ツールを調査して、RDMに保存すべき構成情報等を特定し、RWD研究時に再現性を担保するための推奨行為リストを作成
- 研究プロセスのPoC実施
OMOP/CDMとJupyter Notebookを稼働させ、データソースの取り組みから分析までを縦貫して実施し、今回の検討の裏付けを取っていく。