

自院のデータを OMOP化しよう

Data Conversion / ETL

2020.11.20

第40回医療情報学連合大会

国際医療福祉大学 平松達雄



OMOP-CDMに変換する利点

医療データを分析用の共通データ形式(OMOP-CDM)に変換すると何がよいのか。

- (1) コミュニティにより専門家のアドバイスが得られ、先人の成果が利用できる。
- (2) 複数のソースデータに対して、同じqueryが使える。
異なる医療施設、異なる種類のデータ
(レセプト,DPC,検査値,レジストリ)
- (3) 無料のODHSIツールが使える。
- (4) OHDSI network study に参加できる。
国内だけでなく国際的に。
真に国際的なCDMはOMOP/OHDSIのみ。

OMOP-CDMへ変換するアプローチ

(1) 既存のツールを使う

元形式・用語が既存(日本標準)のとき。レセプト等。
⇒ OHDSI Japanで整備中。

(2) OHDSIのETL設計支援ツールを使う

形式変換支援：White Rabbit、Rabbit-In-a-Hat
用語変換支援：Usagi

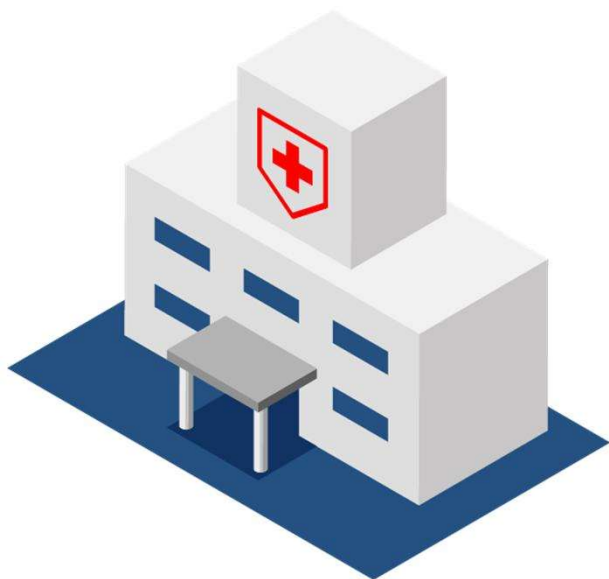
⇒ 設計ができて、変換プログラム実装は必要。

(3) カスタム

設計からプログラムから自分で頑張る。

(4) やってほしい⇒ご相談ください。

元データの準備：どこにあるのか



医事課

レセプト、DPCデータ

既存のファイル

情報システム部門

DWH、SS-MIX2

電カル直接+FHIR

ある程度整備
+今後の取り組み

各部門

部門システム

今後の取組

各診療科

疾患レジストリ

整備あり/なし

元データの準備：誰に利用許可を得るのか

病院業務利用

元データの管理者/部門、診療情報管理室/委員会
加えて、事務長/院長に報告。

規制：個人情報保護法

「医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス」

学術研究利用

元データの管理者/部門。および研究倫理審査。

規制

「人を対象とする医学系研究に関する倫理指針」

どうやって元データを取り出すのか

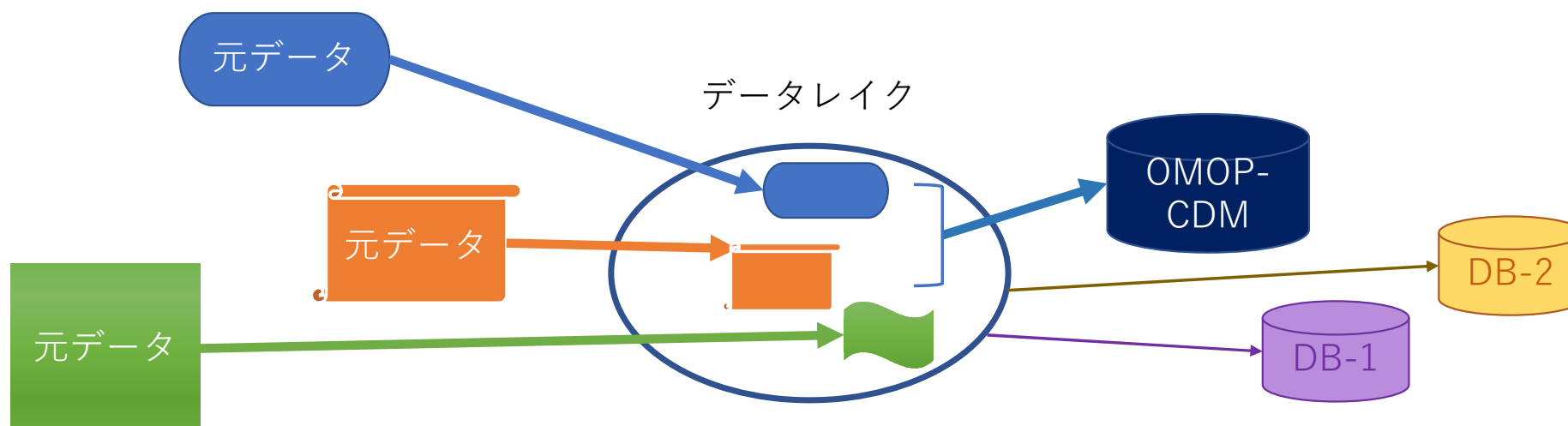
レセプト、 DPCデータ	もともとファイルとして存在。それを元データとする。 DPCデータはID匿名化前のもの（or匿名を戻す）。
DWH	ダウンロード機能でcsv等として取り出す。
SS-MIX2	ディスクをマウントして領域全ファイルを対象とする。 もしくは、トランザクションデータファイル。
電カル直接	電カルの複写テーブルにデータベース接続する。
部門システム	ダウンロード機能でcsv、もしくはDB接続機能。
院内レジストリ	ダウンロード機能でcsv等として。

データレイクをまず作ろう

- 院内にも様々な種類のデータがあり、いきなりすべては無理。
CDMしなくても使えるものも多い。
- 元データは何度も必要になる。

形式を問わず、目的も問わず、**元データをまるごと蓄積する**。
最も基本的な仕組みは「ファイルサーバー」

- ファイル形式になっているものは何でも蓄積できる。
- 整理が人手なので柔軟。だが面倒で不整合も起こりやすい。



最低限必要な機材は？

- **安全な** ネットワーク環境
- データレイク用
ファイルサーバ(NAS): 1人ならUSB外付diskでも
- CDM用
RDBMS(PostgreSQL推奨)
- 分析用 OHDSI-Tool
Windows or Linuxマシン
Tomcat(java)、R-Studio
- 分析用 カスタム
お好きなツールで。

有料なものは、
ハードウェアと
Windows OS

利用の準備

- 相談するコミュニティ
e-Phenotyping課題研究会(Phenotyping関連)
ODHSI Japan (研究だけでなく業務もOK)
- OHDSIでの研究に参加
ODHSI Japan ⇒ Global OHDSI
- 自分で立ち上げる
 - 単独研究 (院内研究)
 - 研究仲間での連携研究
 - 国内の連携研究
 - 世界での連携研究

www.ohdsi-japan.org
「お問い合わせ」から