

メタデータマッピングとeSource

UMIN CDISCシンポジウム2017

CDISC標準を用いた臨床試験データマネジメントの効率化

2017/10/12 UMINセンター 岡田 昌史

メタデータを知らなければ データを交換/検証/解析できない

STUDYID	DOMAIN	USUBJID	DASEQ	DATESTCD	DATEST	DAORRES	DAORRESU
"CDISC01"	"DA"	"CDISC01 .100008"	1	"DISPAMT"	"Dispense d Amount"	"40"	"TABLET"
"CDISC01"	"DA"	"CDISC01 .100008"	2	"DISPAMT"	"Dispense d Amount"	"80"	"TABLET"
"CDISC01"	"DA"	"CDISC01 .100008"	3	"RETAMT"	"Returned Amount"	"22"	"TABLET"
"CDISC01"	"DA"	"CDISC01 .100008"	4	"RETAMT"	"Returned Amount"	"35"	"TABLET"
"CDISC01"	"DA"	"CDISC01 .100014"	1	"DISPAMT"	"Dispense d Amount"	"80"	"TABLET"
"CDISC01"	"DA"	"CDISC01 .100014"	2	"DISPAMT"	"Dispense d Amount"	"80"	"TABLET"

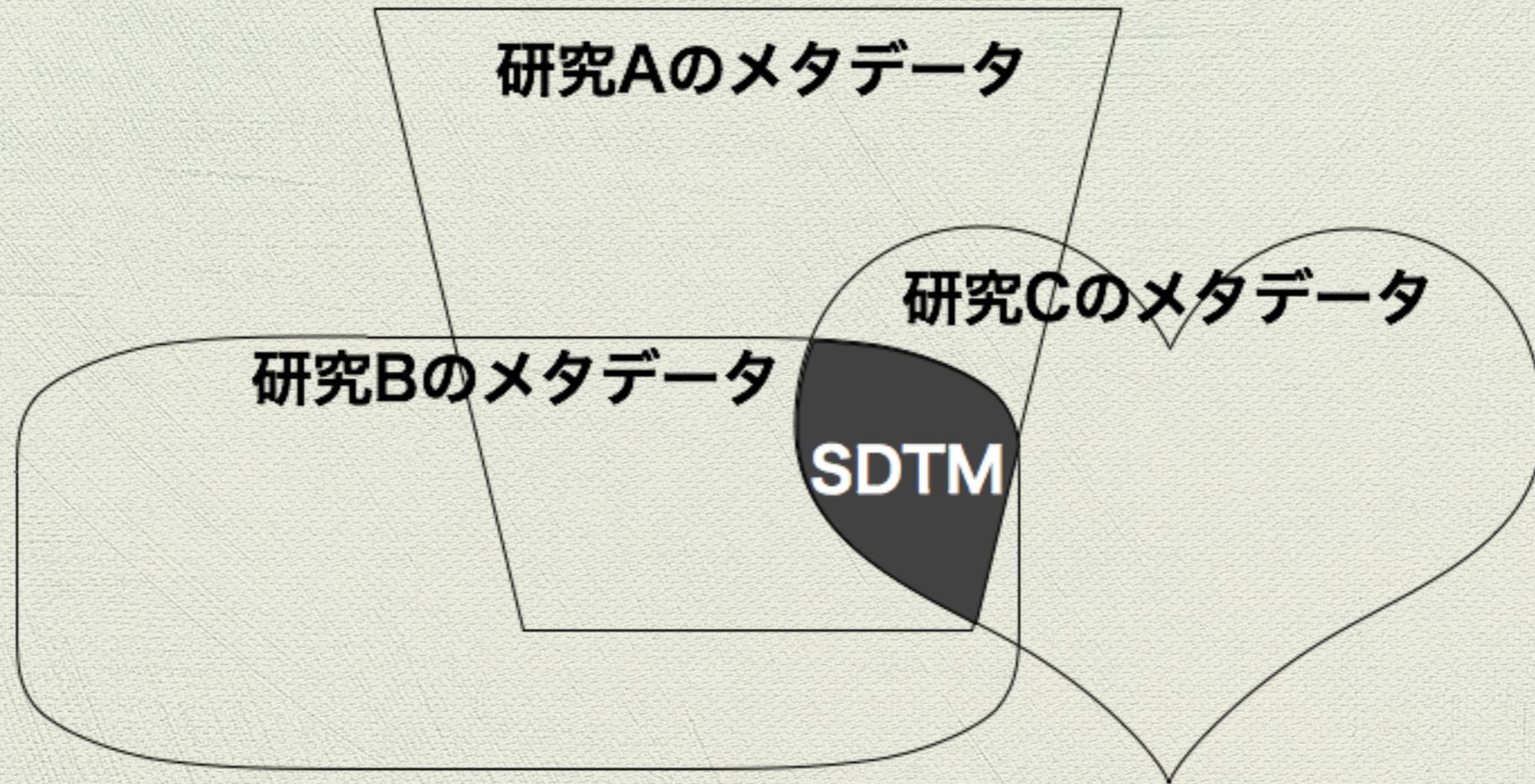
USUBJIDは重複は許
されない

別々の項目が同じ
列に入っている
(CDISC
Normalized Data)

項目によって数値
と判断すべきか、
文字列とすべき
かが異なる

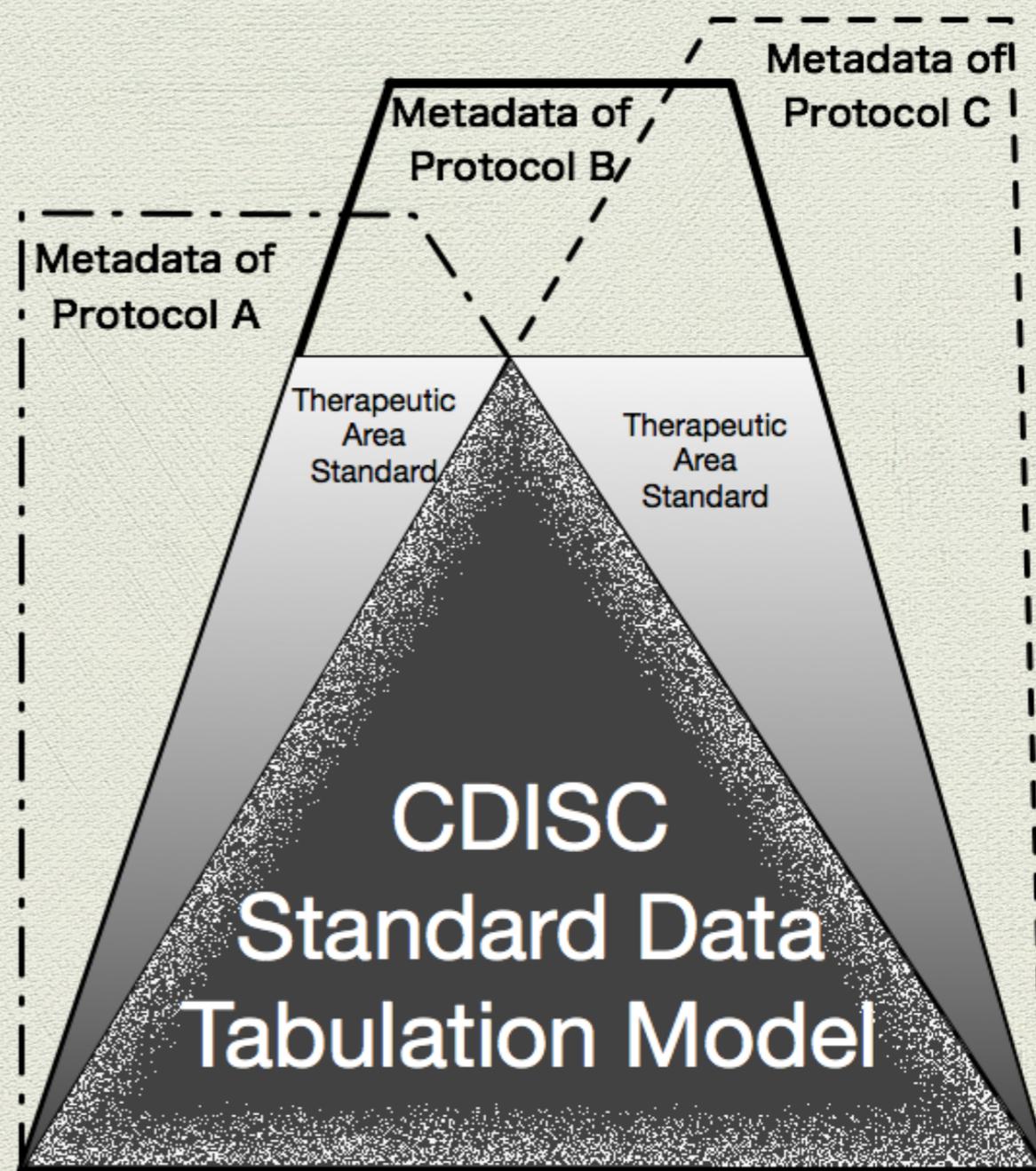
名前は標準用
語に従ってい
る
(CDISC
Terminology)

プロトコルごとにメタデータがある



CDISC標準とは、メタデータの記述方法の標準であり、
同時によく使われる標準的メタデータ自体の集合でもある

- ◆ **メタデータの記述方法**
ODM, Define-XML, SDM-XML, SHARE
- ◆ **よく使われるメタデータの集合**
CDASH, SDTM, Terminology,



eSourceは皆やりたいと思っているがいつも泥臭い話になる

- ◆ EHRからEDCに自動的にデータを取り込みたい！
- ◆ 第1の障壁：APIの違い
- ◆ 第2の障壁：メタデータの違い
- ◆ 第3の障壁：セキュリティ
- ◆ 全部乗り越えようとする、特定のEHRから特定のEDCについては可能だけれども、応用がきかないので、毎回新規開発に。

第1の障壁: APIの違い

- ◆ EHRからデータを取得するための統一インターフェース: FHIRがこれからそうなるかもしれないけど、現状では「ない」と答えるべきか。
- ◆ でも、EHRのバックアップ用アーカイブからデータを取得するための統一インターフェースならある:
SS-MIX2
- ◆ EDCにデータを格納するための統一インターフェース: ODM

第2の障壁：メタデータの違い

- ◆ EHRの中のデータのメタデータのうちよく使われるものの標準: openEHR Archetypes, HL7 RIM, 検査や処方コードたち
- ◆ EDCのメタデータのうちよく使われるものの標準: CDASH, SDTM, Terminology
- ◆ EHRの中のデータのメタデータの記述方法の標準: openEHR, HL7 CE など
- ◆ EDCのメタデータの記述方法の標準: ODM, Define-XML, SDM-XML, SHARE

第3の障壁：セキュリティ

- ◆ EHRの全てのデータがEDCへの入力に必要なわけではない
- ◆ 特定の人の特定の日付の特定のデータだけが見せられる(同意が取得されている)
- ◆ SS-MIX2ではこの抽出は非常に容易

eSourceの障壁をどうすれば超えられるか

- ◆ APIの違い: SS-MIX2からODMとすることで解決に近づけそう
- ◆ セキュリティ: SS-MIX2から抽出することで障壁を下げられそう
- ◆ メタデータの違い: 銀の弾はない

メタデータの違い

- ◆ そもそも「標準的によく使われるメタデータセット」がかなり違う。
 - ◆ 「定義されているもの」自体が違う
 - ◆ 運よく同じものがあったとしても、「用意されている選択肢」も相当違う
- ◆ メタデータの記述方法も違う。
 - ◆ openEHRがまだ十分普及していない
 - ◆ Define-XMLなどは比較的単純な1層構造だが、openEHRはオブジェクト指向の階層構造

ではどうするか？

- ◆ 人力。特に経験豊富な人の作業。
- ◆ でも、その作業過程が不透明であったり、担当者によってあまりにも異なってしまうては困る
- ◆ メタデータマッピングの作業ガイドラインと、マッピングテーブルの標準的な記述方法が必要

SS-MIX形式で標準化された診療情報の CDISC標準への変換に関する研究

- ◆ 計画している成果物1: SS-MIX2ストレージに格納された電子カルテデータから、CDISC標準規格を使用している臨床研究にデータを抽出する作業を支援するソフトウェアに通常求められる要件と仕様例をまとめた文書
- ◆ 計画している成果物2: SS-MIX2ストレージに格納された電子カルテデータから、CDISC標準規格を使用している臨床研究にデータを抽出する作業を、臨床研究データマネジメントの専門家が実施する場合のBest Practiceを含む作業ガイドライン
- ◆ 計画している成果物3: CDISC標準規格を使用している臨床研究にデータを抽出することを前提とした場合に、SS-MIX2拡張ストレージに格納されていることが望ましい情報の規格案

マッピングテーブルの記述方法

- ◆ EHR上のある項目から、何らかの値の変換、用語の対応付け、もともとないメタデータの付加などを行なって、CDISC標準(CDASH+CDISC Terminology)の項目に対応させる
- ◆ 「何らかの」自体にバリエーションがある: 主語, 動詞, 目的語、いずれも拡張可能でなければならない
- ◆ 文法としては、Resource Description Framework, RDF の出番. OWLだと少し自由すぎるのでRDFが適切か。
- ◆ 場所としては、SS-MIX2の拡張ストレージがある

RDFを用いたメタデータマッピングの記述方法 (案)

- ◆ SS-MIX2ストレージに格納されているのはHL7メッセージ
- ◆ HAPI(JavaによるHL7パーサライブラリ)を用いて、HL7メッセージをRDFに変換する試みがなされている
Kawazoe Y, Imai T, Ohe K. A Querying Method over RDF-ized Health Level Seven v2.5 Messages Using Life Science Knowledge Resources. Bamidis P, ed. JMIR Medical Informatics. 2016;4(2):e12. doi:10.2196/medinform.5275.
- ◆ 例えば上記研究の要素名を使わせてもらって記載してみると,
 - ◆ hl7:OBX.5 hcm:copy <<http://rdf.cdisc.org/std/cdash-2-0#DataElement.LB.LBORRES>>
- ◆ のようなRDF Tripleの形のマッピングテーブルがありうる
- ◆ 単なるcopyであればpredicateひとつで書けるが、いくつかの項目の組み合わせや複雑な変換を伴う場合にはSPARQLクエリ自体を用いることで関係を記述できるかもしれない

まとめ

- ◆ eSourceを一般的に実現するための障壁として、APIの違い、メタデータの違い、セキュリティの3つの障壁がある
- ◆ SS-MIX2ストレージから入力し、ODMとして出力することになれば、APIの違いとセキュリティの障壁は低くなる
- ◆ 残ったメタデータの違いは、EHRと研究プロトコルの組み合わせごとに作成するしかないが、マッピングテーブルをSS-MIX2拡張ストレージとRDFを用いて標準化することにより、ソフトウェアによるマッピング作業支援、テーブルの再利用なども可能性が見えてくる