

公募シンポジウム

公募シンポジウム9

診療データの長期保存における SDMの有効性と課題

2018年11月25日(日) 10:40 ～ 12:10 G会場 (5F 504+505)

[4-G-2-1] 診療データの長期保存における SDMの有効性と課題

○紀ノ定 保臣（岐阜大学）

一般社団法人 SDMコンソーシアムが定義している SDM（Semantic Data Model）は、データウェアハウス（DWH）の標準化を目的としたデータベース（DB）の1データモデルである。SDMによって標準化された DWHは、普遍的な医療情報として長期保存が可能となり、ベンチマーキングやビッグデータ分析への利活用など、多くのメリットを享受することができる。一方、診療データを SDMに保存する際、カルテや部門システムで入力された診療データを、SDMが定義した形式に変換する必要があり、その際、原本性の維持に関する問題もある。

もともと入力され保存された診療データは、それぞれ固有の精度、粒度、頻度、および定義を持ち、SDMが定義しているものと異なるため、変換の際、標本化定理により原データより精度が上がることはない。しかし、補間関係にある複数の原データを用いて変換を行うと、十分な精度が得られる場合もある。例えば、「単位時間あたりの診療行為別売上」という定義に対して、医事会計からは、診療行為と売上の情報が得られ、カルテからは、診療行為と行為のために費やした時間が得られる。これらを組み合わせれば、定義に対して十分な情報を取得できたことになる。

SDMにおいては、データの原本性を保証しつつ、情報の精度を維持するための変換ルールを設けている。例えばテキスト情報に関しては、プレーンテキスト、リッチテキストの両方を記録し、数値情報に関しては、まるめずに原データの有効数字を維持しつつ統一した単位とともに記録する。コード情報はマスター変換結果とともに記録し、日時が分かれている場合は、日時型として記録するなど、変換ルールは、原本のもつ精度と意味を継承することが原則であるが、もし誤変換が行われると、SDMの信頼性が失われることになるため、変換の設計は SDMコンソーシアムで管理している。本論では、具体的な変換例をもとに、SDMの原本性について論ずる。

診療データの長期保存における SDM の有効性と課題

紀ノ定保臣^{*1,*2}、鈴木英夫^{*2}、

*1 岐阜大学大学院医学系研究科医療情報分野、*2 一般社団法人 SDM コンソーシアム

On Effectiveness and Problem of SDM for the Long-Term Archives of Clinical Data

Yasutomi Kinoshita^{*1,2}, Hideo Suzuki^{*2}

*1 Dept. of Biomedical informatics, Gifu University Graduate School of Medicine, *2 SDM Consortium

SDM(Semantic Data Model) has been developed aiming at long-term archiving of the clinical data and information. However, there are some problems in hospitals having SDM/DWH or DB. They are that it is very difficult to transform the clinical data to universal valuable information in hospitals. As results, the number of data which are not transformed into the SDM format have been increased. Therefore, authors have decided to introduce the new definition on concerning transformation rules and policies from the clinical data to our SDM/DWH. In this presentation, we are going to show some results on our rule-changes for better SDM/DWH.

Keywords: Semantic Data Model, Data Warehouse, Long-Term Archive, Medical Information

1. 背景

電子カルテ、部門システムに入力された診療データは、適切な定義とともに普遍的な医療情報として保存することが、医学の発展のために必要不可欠な要件であると言える。演者らは一般社団法人 SDM コンソーシアムを設立し、データウェアハウス(DWH)の標準化モデルである SDM(Semantic Data Model)を開発した。このSDMは接続システムに依存しないため、長期保存することが可能となり、医療機関内での時系列分析、広域なビッグデータ分析等にも応用可能である。一方、統一されていない診療データを普遍的な医療情報へ変換することは容易ではなく、SDMへの変換が困難なデータが多く発生している¹⁾。この問題を解決することが長期保存のために必須であることから、SDMへの変換ルールを定義し、SDMの充足度により評価を行った。(図1)

2. 目的

診療データを一定の定義(スキーマ)に合わせるように変換する手法は、ETL(Extract,Transfer,Load)の一種と考えられる。SDMにおけるETLは、情報統合の意味も含むため、複数の情報源からデータ抽出が行われ、SDMの項目定義に合わせるように取捨選択と加工が行われる。このSDMのETLによりSDMの品質が決まるため、品質を一定にするためのルール策定が極めて重要となる。本報では、このルールを現在稼働している医療情報システムに適用した場合の課題とその解決法について報告する。

3. SDMの概要とETL基本ルール

SDMは、項目名とその項目名から類推される意味が一意となるように定義したスキーマに診療データが格納されているDB(データベース)であり、基本的にはCSV(Comma-Separated Value)ベース、すなわち項目名を列に持ち、履歴情報を行に持つようなデータ・マート形式で記録される。またコードや符号情報はすべてテキストに変換され、計測

値などは単位を統一し、日付と時刻は日時形式で記録するように定義している。また、各テーブルには共通項目が設定され、その共通項目が各テーブル間の関係(リレーション)定義を容易にしている。

3.1 抽出におけるルール

SDMにおいては、医療情報システムにユーザーが入力、あるいは選択し保存したデータを対象とするが、入力時に自動的に記録される日時やログ・データなどの属性データも抽出対象となる。一方、医療情報システムを稼働させるためだけに必要な画面表示に関するデータ、および画像やスキャン・データなどの非構造化データは抽出対象としないため、SDMのデータサイズはコンパクトである。すなわちSDMは医療情報システムのバックアップとして機能するものではない。この方針のもと、以下のルールを設定した。

- ① 原データの意味を変えないことを条件に複数のデータを組み合わせることは可
- ② 数値の精度を変えないことを条件に単位の変更は可
- ③ フラグやマスターなどコードデータは変換ロジックを記録することを条件にテキストデータへの変換は可

3.2 変換およびSDM登録におけるルール

SDMを導入したユーザーにおいては、抽出元に保存されている原データよりむしろETLにより変換され、医療情報として登録されたSDMを利用するようになる²⁾。すなわちSDMには原データと粒度、頻度、および精度が継承されていなければならない。そのために以下の変換ルールを設定した。

- ① SDMの定義に合うデータが存在しない場合は、欠損データとし、マッピングされていないことを明確にする
- ② 二値フラグの場合、可能な限り+:あり、-:なし、に統一する。
- ③ 原データの粒度、頻度、精度に対し、補間処理などの加工を行ってはいけな。ただし、複数の原データを組み合わせることによる精度向上は可。

3.3 SDMの有効性の評価

SDMは、利用価値のある情報を一元管理するためのデータモデルである。従って、各医療情報システムのデータを取捨選択し、定義された情報に変換したもののみを保存している。すなわちSDMの有効性を評価するには、定義された項目そのものに不足が無いか、定義された項目がETLによって充足されているかの二つの観点で評価されるべきと考える。前者を定量的に評価することは難しいが、後者に関しては、項目数に対しての充足度という方法で評価可能となる。本報では、実際のETL設計図をもとに、充足度を算出し、課題と解決方法を検討した。

4. 結果

SDMを導入している病院において、実際のETLで表面化した問題点とその対応状況、および入院、処方などの充足度の検討について述べる。

まず、抽出時の問題として、医療情報システム側の定義があいまいであること。特にSDMの定義における、日時の定義に該当するデータが取得困難であるケースが見られた。例えば予約に関して、SDMでは待ち時間分析を行うために、予約した時間枠と、来院時刻、診察開始時刻、診察終了時刻を定義しているが、対象とした医療情報システム(A社)には該当する項目が存在しないため、システムのイベントログを対象に、もっとも近いと思われる時刻を取得する方法を取らざるを得なかった。このシステムにはそもそも予約に対する実績が存在せず、予約の対象となった実施テーブルから取得しようにも、予約との連携キーが存在していなかった。この例以外にもSDMの定義に該当するデータが存在していない、あるいは連携していない場合が多々あり、その場合にはSDM側は、N/A(Not Applicable)、NULL、あるいはデフォルト値となる。このN/Aが少ないシステムの方が、SDMの充足度が高くなり、二次利用にとって価値が高いと考えられる。

実際のETLにおける充足度を、入院、予約テーブルを用いて三社で比較したところ、入院の共通項目で65%から83%、固有項目で42%から76%、予約の共通項目で66%から80%、固有項目で42%から85%と大きく差が出た。(表1)

5. 考察

SDMの項目定義は、一般的な診療行為において発生頻度が高いものを中心に形成されている。また、広域にデータ収集可能なように、すべてのテーブルに共通な項目を設置している。これはビッグデータの分析を意識しているためである。また、患者を対象とした分析以外にも、医療従事者を対象とした分析、例えば業務量分析などのため、共通項目には、著作者(行為を行った責任者)が記録される。従って著作者の役割と職種はテーブル毎、あるいはレコード単位で変わる。例えば、オーダであれば、オーダを発行した医師、実施であれば、実施した技師などが著作者となる。従ってSDMの共通項目を満たすためには、部門システムから取得したデータが欠かせないことになる。SDMとして部門システムのテーブルを追加するか、ETL時に部門システムも対象とするかのどちらかで対応する必要がある。

6. まとめ

SDMは医療情報の長期保存を目標とし、医療情報システムに依存しない普遍的な項目定義をもつDBとして考案され開発されてきた。しかし、導入された施設間において、ETLプロセスにより移行されたSDM項目の充足度が大幅に変わることが問題となり、その原因が接続先システムの保存データの

不足にあることが判明した。この当面の課題を解決するために、部門システムも対象とする必要性を提言した。

参考文献

- 1) 久島昌弘:SDMを利用した診療情報データベース, 医療情報の利活用と関連諸問題に関する沖縄シンポジウム抄録集,SDMセッション (S-3)②.
- 2) 島川 龍哉,小園 菜美,鈴木 英夫,紀ノ定 保臣:Semantic Data Model の導入による診療データ分析プロセスの改善, 第22回日本医療情報学会春季学術大会, 112-113,2018. 2011-Nov-11].

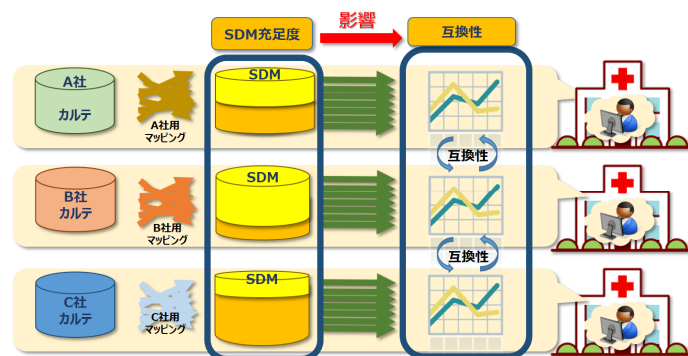


図1. SDMにおける充足度と互換性の関係

共通項目	入院	予約
A社	65%	70%
B社	74%	66%
C社	83%	80%

固有項目	入院	予約
A社	42%	42%
B社	76%	85%
C社	53%	53%

表1. SDM 入院、予約テーブルにおける充足度の各社比較