

公募ワークショップ

## 公募ワークショップ12

### SDM（標準化された DWH）における品質確保に関する検討

2019年11月24日(日) 13:40 ~ 15:10 D会場 (国際会議場 2階中会議室201)

#### [4-D-3] SDM（標準化された DWH）における品質確保に関する検討

島川 龍載<sup>1</sup> (1. 県立広島大学)

キーワード：SDM (Standardized Data Warehouse), Quality of Clinical Data, ER diagrams

筆者らは2017年に一般社団法人 SDMコンソーシアムを設立し、DWHの標準化を目指した SDM (Semantic Data Model) を公開し、医療情報学会等でその運用に関する課題の検討と導入効果について報告を行ってきた。一方、SDMユーザ数の増加とともに、新たなデータの品質に関する課題が顕在化してきた。具体的には、SDMに格納されるデータの意味や定義に関する品質が、SDMの導入医療機関あるいは納品業者によって左右されることがあるという事実である。このような現象は他の標準規格の活用時にも起こり得る現象であるが、広域に収集した医療データを用いたビッグデータ分析を行う際には確保されるべき必須要件である。本企画では、SDMの品質確保に関する検討結果を報告し、具体的な施策を提案する。

SDMは、データの定義や意味を統一して保存する規格である。その利点として複数の医療機関におけるデータの統合、あるいは分析手法の互換性が挙げられる。また医療情報学会においても、異なる電子カルテシステムから作成された SDMの充足度と共通項目に関する報告などを行ってきた。しかし、ベンダー共通の項目であっても、定義や意味が合致していなければ、分析対象の項目にはならない。

筆者らは、今回 SDMの品質確保に関する検討を行った結果、SDMにおいては、電子カルテからのデータ抽出・変換・加工 (ETL) 処理精度の違いにより、データの精度、頻度、粒度に差異が生ずることを明らかにした。また SDMの定義に関する解釈の違いや誤解により、SDMの品質が低下する可能性も示唆された。

本企画では、SDMの品質低下をもたらす要因の解析結果を報告し、施策としての ETL検証方法、および SDMのサンプルデータの有効性に関して提言する。また、品質確保に関する SDMコンソーシアムの今後の取り組みについても言及する。

## SDM(標準化された DWH)における品質確保に関する検討

島川龍範<sup>\*1</sup>、鈴木英夫<sup>\*2</sup>、木村映善<sup>\*3</sup>、近藤博史<sup>\*4</sup>、久島昌弘<sup>\*5</sup>

\*1 県立広島大学、\*2 SDM コンソーシアム、\*3 国立保健医療科学院、\*4 鳥取大学、\*5 沖縄県立中部病院

### The Use of Quality Assurance of SDMs (Standardized DWH) to Increase the Value of Clinical Data

Tatsunori Shimakawa <sup>\*1</sup>, Hideo Suzuki <sup>\*2</sup>, Eizen Kimura <sup>\*3</sup>, Hiroshi Kondo <sup>\*4</sup>, Masahiro Kyusima <sup>\*5</sup>

\*1 Prefectural University of Hiroshima, \*2 SDM Consortium, \*3 National Institute of Public Health,

\*4 Tottori University, \*5 Okinawa Chubu Hospital

An SDM is a type of standardized clinical DWH based on the “Semantic Data Model”, which was proposed by our SDM Consortium on September 26, 2014. The main goal of our project is to establish a clinical DWH, which is open to SDM users especially regarding the definition of entity relationship (ER) diagrams. On the other hand, maintaining quality, is associated with some management difficulties, especially in cases involving vast differences in the understanding of the original definition of ER diagrams and the interpretive activity of engineers. In our workshop, we will discuss a broad range of issues regarding the quality of SDMs.

**Keywords:** SDM (Standardized Data Warehouse), Quality of Clinical Data, ER diagrams

#### 1. 背景

筆者らは 2014 年 9 月 26 日に一般社団法人SDMコンソーシアムを設立し、DWHの標準化を目指したSDM (Semantic Data Model)を公開し、医療情報学会等での運用に関する課題の検討と導入効果について報告を行ってきた。

一方、SDMユーザ数の増加とともに、新たなデータの品質に関する課題が顕在化してきた。具体的には、SDMに格納されるデータの意味や定義に関する品質が、SDMを導入しようとする医療機関、あるいは納品業者によって左右されることがあるという事実である。このような事象は他の標準規格の活用時にも起こり得ることであるが、広域に亘り収集した診療データを用いたビッグデータ分析を行う際には確保されていなければならない必須要件となるため、高い関心を持っている。

本企画では、SDMの品質確保に関するこれまでの検討結果を報告するとともに、新たに提起された品質維持に関する課題について具体的な施策を提案する。また、学会参加者からの意見も取り込みながら、より良い標準規格の構築を目指したい。

#### 2. SDM の概要、SDM と他の標準規格との関係、標準化 DWH の求める品質の要件について

SDMの構築は、SDM定義を熟知した者が、ユーザ情報システムからのデータベース定義及び診療行為とデータとの関係性を解釈し、SDMの定義にマッピングするという作業工程を繰り返す地道な ETL (抽出、変換、登録)開発工程により、その品質を保ってきた。一方、SDMの普及を考えた場合、ETL開発の効率化を考える必要がある。筆者らは、過去に標準規格を介してSDMに効率よくデータを入力する手法を試みた経験もあるが、カルテベンダーが自社の定義の中で、最低限の項目だけを選択して出力しているなど、データの一部しか出力されていないなどの課題があったため、SDMの充足度を満たすことができなかった苦い経験を有している。

一方、昨今話題となっている FHIR に関しては、一定程度のデータ品質が保証されたデータ API が提供されていることから、SDMとしてはFHIRの標準化項目やProfileとも整合性を取ることで、一定の品質水準を保った環境下でETLを効率的に提供することが可能になると期待している。

本稿のワークショップでは、検査の結果やアレルギーの有無等に関するデータの取扱を題材にSDMへの取り込み・格納について検討し、今後のSDMとETL設計における方針の方向性を提言したい。

#### 3. 品質の問題が二次利用に及ぼす影響について

診療データを二次利用するためには、対象とするデータベースに必要なデータが存在し、かつデータの意味的な定義が明確でなければならない。しかし、現状のベンダーが提供するDBあるいはDWHは、独自の構造を持ち、同じ項目名であっても医療機関ごとに意味が異なるなど、二次利用に供するためには多くの事前処理が必要となる。

一方、標準化DWHを目指すSDMは、項目の意味的な定義を厳格に指定したDWHであるため、利用者にとっては前処理が不要になるという利点がある。しかし、同じ情報ソースの診療データであっても、SDMに変換するETL(抽出、変換、登録)が、ベンダーにより異なっている場合、異なる結果がSDMに格納されることになるため、SDMの定義を前提とした分析の信頼性に悪影響を及ぼすこととなる。特に複数施設の診療データを広域に収集する場合は、ベンダー毎のSDMへのマッピング充足度が、結果として共通利用可能なデータの種別を制限してしまう可能性が高くなる。特に、SDMを利用する際の分析軸となる共通項目(いつ、誰が、どの立場で、誰のために、何を、どうした)の粒度、頻度、精度が異なる場合、集計することは極めて困難になる。このように、データの品質問題は、二次利用に大きく影響する重要な課題であることを認識することが大切である。

#### 4. 品質の問題が地域医療に及ぼす影響について

地域医療の観点から必要な診療情報は、客観的な結果データと医師、看護師等の解釈、重要度、方針などの記録である。SDM では、診療データを広く効果的に二次利用する為に、その定義の統一を推進してきた。この結果、地域の医療機関においては診療上必要なデータが不足する可能性がある。また、現在のカルテに記載されていない情報、例えば、診療方針を追加・変更した理由などを共有できれば、地域包括的に患者をケアすることが容易になる。

SDM においては、上記の観点から、可能な限り診療データをすべて保存し、共有できることを期待している。一方、現時点での地域医療においては、情報を伝文により交換することが一般化しており、伝文で定義されていない診療データを得ることは出来ないという課題もある。このような場合、SDM のような標準化 DWH を活用することが出来れば、アプリケーションを経由せず、情報検索により必要な情報を取得することができる可能性もあり、多様な SDM の活用に期待している。

#### 5. 品質の問題が診療支援や長期記録に及ぼす影響について

SDM は診療行為を記録することを目的として設計されたデータベースである。また、長期に亘り運用してきた診療支援 DB とその一次利用である診療支援アプリに適応すべく、カルテのデータをリアルタイムに取得するような ETL を設計して SDM を構築している。同時に、蓄積された診療データを分析などの二次利用に応用でき、さらに次期システム調達においても保存された診療データを移行することなく長期間保存を可能とするベンダーフリーの SDM を念頭に置いている。すなわち、短期利用、中期利用、長期利用に耐え得る診療用 DB を SDM という一つのデータモデルで表現する挑戦的な試みを実施してきたことになる。

しかしながら、データの取得先をカルテに集約したため、部門からカルテへの情報伝達におけるロスの課題や、取得のタイミングの課題など、品質に関する新たな課題も見えてきた。例えば、重症系のデータを対象とした診療支援アプリにおいては、患者の状態変化に合わせて診療行為を変えていく必要があり、重症系のデータがカルテに反映されるタイミングとカルテから取得する ETL のタイミングのずれにより、最新のデータが結果としてカルテに反映されないケースがある。

さらに、トリアージを含む患者受付をカルテアプリで入力し、その結果を重症系部門システムが取得したのちに、部門システム側のデータ入力が始まるため、同一患者のデータが異なるシステムに記録されるというデータの不連続性など、カルテ導入以前には存在しなかった新たな課題が散見されるようになってきた。

これら実際に起きた課題を検討した結果、SDM の一次利用に関しては、カルテからの情報だけでは粒度が不十分であり、変化するデータ取得とその記録頻度も不十分であることが新たな課題として顕在化してきた。従って、SDM を診療支援に用いる場合は、診療支援など“一次利用目的”のリアルタイム DB と、“二次利用や長期保存用”のアーカイブ DB の二種類に分けた構築をすることが必要であることを提言したい。

#### 6. 品質改善・維持のための具体的な施策及び SDM コンソーシアムのロードマップについて

SDM は、ベンダーフリーでアプリケーションに依存せず様々なアプリケーションで利用可能な汎用 DWH を念頭に置いて開発してきた。一方、アプリケーションに特化した専用 DWH の粒度、頻度、精度を 100% とすると、SDM の充足度は 80% 達成できれば良いと考え、アプリケーション固有のテーブルや項目を出来るだけ少なくすることを考えてきた。その結果、二次利用や長期保存に関しては、ほぼ満足できる粒度、頻度、精度を持っていると評価されている。

一方、診療支援などの一次利用に関しては、ETL の課題も含めて粒度、頻度が充分とは言えない状況であることが明らかになってきた。また SDM の定義に関して、セマンティクスを意識しすぎた結果、ETL 開発にかかる労力が増加し、品質管理の困難さ等、普及するためには明らかな阻害要因となることが指摘され始めた。

上記の新たな課題を解決すべく、SDM コンソーシアムとして、以下に示す改善案の検討を開始することとした。

- ① 一次利用 SDM と二次利用 SDM を物理的に区別し、一次利用 SDM から ETL で二次利用 SDM を構築する
- ② 一次利用として不足項目、テーブルを追加する
- ③ 標準規約とのマッピングを行い、ETL 開発負荷を軽減する
- ④ SDM に診療行為のワークフロー設計を加えることにより、設計のゆれを最小化する
- ⑤ ETL 開発ベンダーへの教育を実施する
- ⑥ ETL 設計書の検証を実施する
- ⑦ SDM サンプル DB を ETL 設計ベンダーに公開する

上記の施策に加えて、SDM コンソーシアムの活動計画に関する具体的なロードマップを提示する。

#### おわりに。

医療機関におけるすべての診療行為は、手動入力、自動入力、スキャン入力、通信等により電子媒体へ記録されている。一方、そのデータは入力アプリケーション毎に管理され、他アプリケーションへは、伝文という方式で情報が部分的に伝達されているため、診療行為を主体とした分析を行うには、複数のアプリケーション内に記録されたデータ収集作業と、粒度、頻度、精度が異なるデータを統一するクレンジング作業が必要となる。

一方 SDM は、診療行為そのものを分析対象とすることに最適化したデータベース・モデルであり、診療データを多面的に分析するための分類軸が一元管理されており、診療行為の分析に適した構造となっている。しかし、この SDM にデータを充足させるためには、各々の診療行為に関連するアプリケーションから直接データを取得し、粒度、頻度、精度を SDM 定義に合わせるための変換を行い、データ間の関係(リレーション)を維持するためのキーを決めなければならない。これらの方法を定義しているのが ETL であり、ETL の設計が SDM の品質、すなわち価値を左右すると言っても過言ではない。

筆者らは、SDM の品質改善、および品質維持に関して、標準規格との比較、二次利用、地域医療、診療支援、長期保存等、それぞれの視点において考察し、その必要性と期待、課題と改善案を具体的に提示した。