

一般口演

一般口演13

データ分析システム

2017年11月21日(火) 16:15 ～ 18:00 G会場 (10F 会議室1006-1007)

[2-G-3-OP13-1] DWHの標準モデルを指向する医療用 Semantic Data Modelの BIに関する有用性の検討

島川 龍載¹, 鈴木 英夫², 津久間 秀彦³, 田中 武志³ (1.広島赤十字・原爆病院, 2.株式会社MoDeL, 3.広島大学病院 医療情報部)

1.背景

これまで HISのデータを二次利用するために構築された DWHには、内部構造の標準規格がなかったため、データ抽出や利活用時に多くの問題が発生していた。これらの状況を受けて、一般社団法人 SDMコンソーシアムが中心となって、DWHの標準化と効果的な二次利用を目指した医療用 SDM（Semantic Data Model）が開発された。今後、データ分析の質向上を図るための新たな標準モデルとして期待されているが、SDMの実業務への適用は端緒に就いたところであり、医療機関での導入効果はまだ十分に検証されていない。

2. 目的

過去の抽出・分析依頼に、現状の SDMがどの程度対応可能か（以下、適合率）を調査する。更に、DWHのスキーマとリレーションと項目を統合的に定義して、一連の処理の自動化を推進したことの BI業務への影響を検討する。

3. 方法

広島赤十字・原爆病院で2014年4月から2017年3月までに医療情報部門に依頼があった抽出・分析依頼を「経営改善／臨床研究／業務改善／医療の質向上」の4つのカテゴリに分類し、SDMの適合率や依頼件数の推移などを調査した。

4. 結果

適合率は、「臨床研究」と「医療の質」が約70%、「業務改善」が約40%、「経営改善」が約30%であった。また、依頼件数は SDM導入後に約30%増加した。

5.考察

適合率は、高くても70%程度に留まったため全体的に更なる改善が必要だが、業務改善や経営改善では特に意識した取り組みが必要なのことがわかった。一方、依頼を処理する分析者は導入前後で2名と変更はないが、分析手順の自動化による医療情報部門の対応力アップを利用者が認識して、「定期的にこのようなデータも出してほしい」といったこれまでなかった要望が増えたことが依頼件数の増加の一因である。

今後、他の SDM導入病院と連携して更に検証の精度を上げると共に、BIへの有用性を高めるためにアウトプットの指標を意識した SDMの構造化の検討を進めることが必要である。

DWH の標準モデルを指向する医療用 Semantic Data Model の BI に関する有用性の検討

島川 龍載^{*1}、鈴木 英夫^{*2}、津久間 秀彦^{*3}、田中 武志^{*3}

^{*1} 広島赤十字・原爆病院、^{*2} 株式会社 MoDeL、^{*3} 広島大学病院 医療情報部

Examining the utility of BI of medical application Semantic Data Model which points on the reference model of DWH.

Shimakawa Tatsunori^{*1}, Suzuki Hideo^{*2}, Tsukuma Hidehiko^{*3}, Tanaka Takeshi^{*3}

^{*1} Hiroshima Red Cross Hospital & Atomic-bomb Survivors Hospital, ^{*2} MoDeL Co.,Ltd.,

^{*3} Department of Medical Informatics, Hiroshima University Hospital

[Background] Because DWH, which was constructed for secondary use of HIS data, did not have internal structure standards, many problems occurred during data extraction and utilization. In response to these problems, focusing on SDM Consortium Foundation, we have developed a medical SDM (Semantic Data Model) in order to standardize DWH and to utilize the secondary data effectively. However, a practical utilization of SDM has just begun, and the effects of introducing the system in medical institutions have not been fully verified yet.

[Purpose] In the data analysis environment that includes SDM, which was introduced at the Hiroshima Red Cross Hospital & Atomic-bomb Survivors Hospital, the relevance ratio of medical SDM, data extraction, and the number of requests for analysis were evaluated. By doing this, we aimed to understand the current response status of medical SDM at the practical level and to investigate for future improvements.

[Methods] The data extraction and analysis requests that were submitted by the medical information department from April 2014 to March 2017, were classified into four categories: "business support, management support, research support, and clinical support," and we analyzed the trial version and official version of medical SDM.

[Result] The conformance rate was 58.0% and 87.1% for the trial version and the official version, respectively, and research support (94.4%) was the highest, and management support (75.4%) was the lowest in the category-based conformance of the official version. The number of submitted request before and after the use of SDM was 822 and 969 cases, respectively, and there has been an increasing trend in the number of submitted requests in every fiscal year.

Keywords: Semantic Data Model, Hospital Information System, Data Warehouse, Business Intelligence

1. 背景

1.1 医療用 SDM の誕生の経緯と現状

病院情報システムの普及に伴い、データを経営・業務改善や医療の質の評価、臨床研究・教育などの目的で二次利用することが期待され、1990年代後半よりデータウェアハウス（以下、DWHとする）の構築に関する研究や、データ活用における仕組みづくり¹⁾の研究が盛んに行われてきた。その結果、実際に多くの医療機関でDWHの導入やデータの利活用の検討が進められてきた。しかし、DWHにおけるデータ保持のための内部構造の標準規格については、ほとんど言及されてこなかったため、統一されないDWHデータ構造に起因する、データ抽出や利活用の促進を妨げるような問題点²⁾が数多く指摘された。具体的な問題点を以下に記す。

1, データ抽出の負荷が高い

DWHが接続するシステムが独自構造のデータベースを有しているため、データ抽出、変換、統合に関する開発のコスト、およびシステム負荷が高い

2, データ間の関係性が不明確

項目名と定義、構造が統一されていない上、テーブル間の関係に関する定義の精度が低いいため、抽出したデータが関係づけられない。例えば関係性のあるテーブル間で項目名が違う例として、同じ更新日に関しUPDATE_DATEやLAST_DATEなど違った表記であるため、結合ロジックの真正性が保証されないことなどが挙げられる。

3, 経営指標や臨床指標などの算出ロジックの根拠が不明

指標など決められた計算式において、頻度や粒度が違うデータを用いて計算した結果の信頼性が保証できないにも関わらず出力されてしまう場合がある。

4, 検索の柔軟性が低い

本来は、DWH上のすべての項目が検索対象となり、非定型な検索が可能であるにも関わらず、特定の検索条件（限られた項目や条件）のみが提供されているDWHの場合、定型業務にしか対応できない。

5, 冗長性の確保が困難

DWHの副次的な利用として、システム停止時の情報検索があるが、医療情報システムと密に連動しているDWHの場合、独立したデータベースとサーバーを持っていないため抽出元のシステムが停止すると検索ができない。

6, 非構造化データ(テキストなど)の検索性が悪い

非構造化データとしてカルテやレポートなどのテキストがあるが、DWH 上に実体がない場合や、独自定義のタグが付与されている場合など、直接テキストから全文検索することが困難となる。従ってテキストを対象とした分析を行う場合には、ユーザー自身が多くの処理を行う必要がある。

7, 二次利用に適した意味のある構造になっていない

セマンティックスが定義されていない DWH は、項目名が同じであっても、定義が間違っていて記録されていることがある。また同じ項目名が複数存在していた場合、どの項目を選択すべきか、あるいは加工する必要があるのかを、ユーザーが考慮する必要がある。

これらの問題を解決するために、一般社団法人 SDM コンソーシアムが中心となって DWH の標準化と効果的な二次利用を目指した医療用 Semantic Data Model (以下、医療用 SDM とする。)が開発³⁾された。SDM は、二次利用を容易にするために検索に適した構造を持ち、ベンダー依存性を排除した DB スキーマとリレーション、すなわちセマンティックスを定義した DWH⁴⁾である。標準化における分類としては①名称コードの標準化、②医療情報通信における標準規格、③医療情報システムの相互接続性のための標準規格に次ぐ第 4 の標準化⁵⁾として、保存規格を定義した DWH であり、データ抽出・分析の質向上や組織横断的なプロセスやアウトカム評価の加速に向けた契機となることが期待されている。

また、情報の利活用のために構築された DWH は、本来は単なる組織内の情報活用のためだけでなく、新たな付加価値を生むための新情報系システム⁶⁾としても期待されており、近年では、DWH、データマート、BI の役割の関係性において、戦略的な経営情報システムとしての期待も高まりつつある。しかしながら SDM の実務への適用は端緒に就いたところであり、医療機関での導入効果はまだ十分に検証されていない。

1.2 広島赤十字・原爆病院のデータ分析環境

広島赤十字・原爆病院(以下、当院)では、2010 年 9 月に電子カルテベンダーが提供する DWH を導入したが、データ抽出、分析の依頼に対する 2 次利用できるデータが DWH には存在しないことが数多くあったため、実テーブルの複雑なデータ構造を解読しながら、実データと DWH を組み合わせで対応するなど、分析者のスキルへの依存度が高く、結果を出すまでに多くの時間と手間を有していた。また、スキルに依存することで正確性や再現性にも課題が生じていた。そこで前述の状況を踏まえて 2015 年 10 月に DWH を含めたデータ分析環境を整備し、主に医療用 SDM を中心としたデータ構造の標準化とビジネスインテリジェンス(以下、BI とする)の環境を設計して、データ分析の正確性・再現性の向上、及び自動化・視覚化の促進を目的とした意思決定支援環境を構築⁷⁾⁸⁾している。

導入前、導入後の分析環境の違いは図1に示すとおり、① DWH がメーカー側の準備した DB から、メーカー依存しない

独立した DB に変更したこと、②データ分析にかかる自動化及び視覚化のツールが導入され、SDM と連動したことが大きな違いとして挙げられる。なお、当院で採用している医療用 SDM は試行版である。

データ抽出、分析の依頼は、Microsoft Sharepoint Server のワークフロー機能で行われ、依頼目的、依頼条件、依頼項目を目視で確認し、分析者が IBM SPSS Modeler を利用して、DWH にアクセスしてデータ抽出、分析を実施する。業務の流れ、分析者がデータ抽出、分析に利用するツールは、医療用 SDM の導入前後で変更はない。

運用体制は、分析者として 2009 年度、2011 年度から担当している 2 名で、導入前、導入後で人数、担当者ともに変更はない。

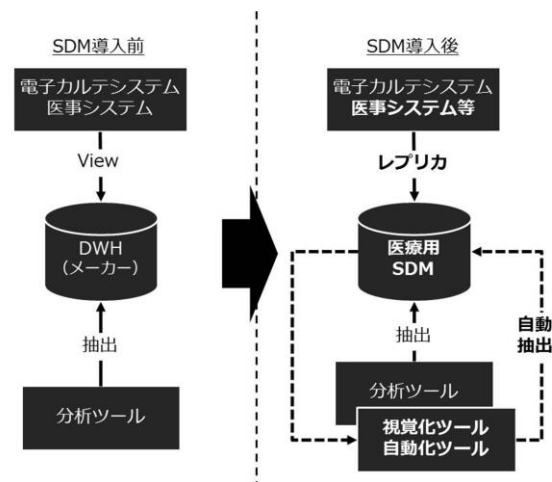


図1 導入前、導入後の分析環境

2. 目的

当院の電子カルテシステムと分析環境を利用して以下の2点を評価することで、1.1 節の 7 つの問題のうち、2)、3)、7)の解決にどれだけ寄与できているかを検討する。

なお、今回の検討では、電子カルテシステム等の実装・運用に依存する蓄積データの質の要因⁹⁾には言及せず、現在蓄積されているデータに対する医療用 SDM の効果に焦点を絞って検討する。それにより、医療用 SDM の実務レベルでの現状の対応状況と今後に向けた改善点の検討に資することを目的とする。

2.1 医療用 SDM の適合率

過去のデータ抽出や分析依頼に対して、医療用 SDM のテーブル構造がどの程度対応できているか、適合率(医療用 SDM からデータ抽出可能な件数/全体依頼件数)を調査することで、医療用 SDM の有用性を検討する。SDM の適合率を調べる意味は以下の2つである;1)実務に合わせたデータ抽出、分析への対応状況の把握につながるため、データの関係性の明確化と指標の算出ロジックの根拠に関連する;2) 医療用 SDM のテーブル設計は、電子カルテシステム、医事シス

テムで入力された実データを対象として医療プロセスをモデル化しており、実務で必要とされるデータとの差異がどこにあるのか確認することにつながる。⁴⁾

2.2 データ抽出、分析の依頼件数の推移

BI システム更新前後の依頼件数を比較することで、副次的ではあるが、医療機関での BI ツールの活用や加工処理の自動化の事例¹⁰⁾¹¹⁾にも報告されている、正確な数値でより短期間にレスポンス良く対応し、時には必要に応じて定期出力(自動化)に対応することで、多様な依頼を迅速に処理し、意思決定に対して適切に役立てることが可能になる。DWH と合わせて BI や自動化ツールが導入されることで、データ活用の視点に立ったアウトプットへの効果がどの程度期待できるのか検討する。

3. 方法

3.1 分析対象の決定

当院で、2014 年 4 月から 2017 年 3 月までに医療情報部門に依頼があったデータ抽出、分析依頼から、画像データやファイルの抽出などを除く、医療用 SDM の対象である構造化データの抽出依頼を今回の分析対象とした。

データ抽出、分析の依頼は、グループウェアのワークフローシステムに依頼目的、依頼条件、依頼項目などのデータが蓄積されており、データを CSV に出力して BI システムの分析ツールにインポートし、集計作業を行った。

なお、BI システムにおける医療用 SDM 導入後(2015 年 10 月以降)は、月次、週次、日次にかかる定期でのデータ抽出のうち、システムで自動抽出(グループウェア等での BI の視覚化を含む)されるようになったものがある。自動でのデータ抽出は 1 つの依頼に対する対応であるため、依頼件数は 1 件としてカウントしている。

3.2 適合率の算出方法

今回の分析には医療用 SDM の試用版と正式版(Ver.1.00)を用いて下記のプロセスをそれぞれで実施した。

対象となる依頼の依頼内容(医療情報部門への依頼者からの依頼目的、依頼条件(期間等)、依頼項目(抽出するデータの中身)と SDM のテーブル定義のフィールド名およびその補足説明を突合し、目視で比較した。実際のデータ検索を想定し、医療用 SDM のテーブルから結果の出力が可能であるか、抽出条件と抽出項目の一致を確認し、出力可能とした場合に SDM に適合するものとした。

更に適合と判断された依頼データを、医療情報学会の課題研究会での成果¹²⁾において示された DWH の主な役割のうち今回の分析対象のデータに該当する院内でのデータの利活用を中心とした①業務支援、②経営支援、③研究支援、④診療支援、の 4 つのカテゴリに分類した。

3.3 医療用 SDM 導入前後の依頼件数の比較

当院の BI システムにおける医療用 SDM 導入前(2014 年 4 月～2015 年 9 月)と、導入後(2015 年 10 月～2017 年 3 月)の件数を集計し、上記①から④のカテゴリに分類した。

4. 結果

4.1 分析対象の抽出

今回の分析対象は 1790 件であった。分析対象の内、依頼目的が不当・曖昧、或いはデータが取得出来ないなどの理由で依頼を却下したものが 51 件(2.8%)あった。しかし、全体件数に対する却下率が低いこと、および実際のニーズ(依頼を行ったという事実)に対する適合性を調査することが実業務に於ける利用につながるため、今回の分析対象は却下したものを含めた全ての依頼データを対象とした。

分析対象のうち、28 件が自動抽出されたものであった。

4.2 SDM の適合率

医療用 SDM の適合件数と適合率を図 2 に示す。適合率は、試用版(58.0%)、正式版(87.1%)と大きな差があった。カテゴリ別の試用版と正式版での適合件数の差では、業務支援(231 件)が最も大きく、適合率の差では、経営支援(57.6%)が最も大きかった。

正式版のカテゴリ別適合件数では、業務支援(653 件)と最も多く、経営支援(199 件)と最も低かった。正式版のカテゴリ別適合率では、研究支援(94.4%)が最も高く、経営支援(75.4%)が最も低かった。

なお、正規版で不適合のデータは、大きく 2 つに分類された。1 つ目が、抽出対象がログデータやマスタデータなど、電子カルテシステム、医事システムに入力された実データではないもので、119 件(全体の不適合件数のうち、51.5%)であった。2 つ目が、抽出対象が部門システムに蓄積されているデータの一部(検査レポートの中身や各科検査や治療等に起因するもの)であるもので、112 件(全体の不適合件数のうち、48.5%)であった。

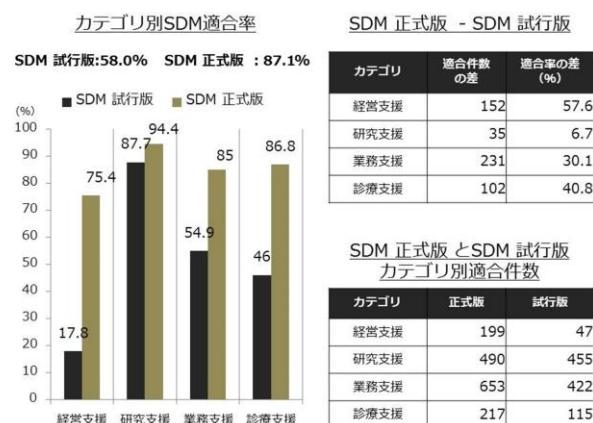


図 2 医療用 SDM の適合件数と適合率

4.3 SDM 導入前後の依頼の比較

データ抽出、分析の依頼件数の推移を図3に示す。全体の依頼件数は、SDM 導入前 822 件、導入後 968 件で、年度別依頼件数の推移も増加傾向である。

カテゴリ別依頼件数では、①業務支援(768 件)、②研究支援(519 件)、③経営支援(264 件)、④診療支援(250 件)の順に件数が多く、SDM 導入前後で経営支援、研究支援、業務支援は増加傾向にあるが、診療支援は減少している。

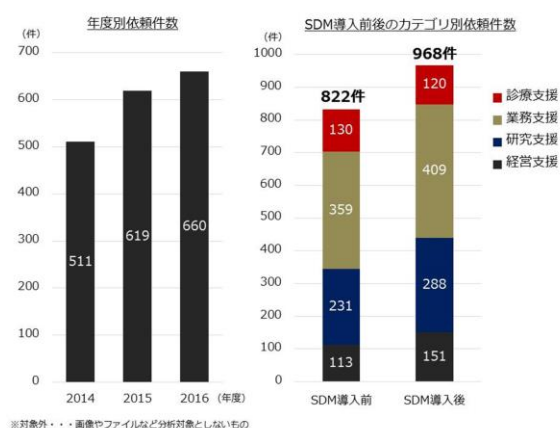


図3 データ抽出、分析の依頼件数の推移

5.考察

試行版の医療用 SDM は電子カルテシステム、医事システムに蓄積されたデータを対象としており、未整備のテーブル定義が多くある。正式版は、部門システム、DPC データなども対象可能となるため、今回の調査で行ったように正式版と試行版を比較することで、段階的に実務的ニーズへの対応を確認できる。

医療用 SDM 導入以前の DWH に試行版 SDM と同等程度のデータが蓄積されていることを考えると、今回正式版の適合率が高かったことは、データ抽出、分析のニーズへの対応能力が改善していると言える。その理由として正規版では試行版と比べて、レプリカ元となるシステムの対象範囲が広がり、データ細目が増えたことがその主な理由と考えられる。実業務に合わせて、SDM コンソーシアムにフィードバックされた要望が、精査されて SDM 標準構造に組み込まれていることは、試行版での他の事例検討においても示されている。¹³⁾

但し、今後はカテゴリ分類毎の適合率にばらつきがあることや適合できなかった依頼への対応も検討する必要がある。適合できなかったデータに対しては、電子カルテ内の実データを扱う医療用 SDM がどこまで依頼者のニーズに答える必要があるか、今後の議論が必要になるであろう。

また、医療用 SDM 導入後での依頼件数は増加傾向であり、少なくとも BI への医療用 SDM の導入が業務効率を大きく阻害してはいないと考えられる。今後は、職種別にそれぞれのニーズに合わせて依頼内容を検討し、BI への活用を促進す

ることで、新しいデータと分析の軸をもって導入効果の検証を実施したい。

6. 結語

今回の調査において、SDM の適合率の状況から、実務に合わせたデータ抽出、分析に概ね対応できていると言える。

今後は、他の医療用 SDM 導入病院と連携した検証やアウトプットの指標を意識した SDM の構造化の検討を進めていくことで、今回示された医療用 SDM の現状から更なる飛躍が期待できるであろう。

7. 参考文献

- 1) 稲岡 則子. データウェアハウスとデータ活用. 医療情報学 2007;27 (3) :261-268.
- 2) 紀ノ定保臣. SDM コンソーシアムの役割とその社会的な背景. 中部 IT 融合セミナー2014. [http://sdm-c.org/].
- 3) 紀ノ定保臣. SDM コンソーシアムの目的と意義. 第 36 回医療情報学連合大会 36th JCMI(Nov.2016):158-159.
- 4) 鈴木英夫. SDM オリエンテーション. 第 36 回医療情報学連合大会 36th JCMI(Nov.2016):166-167.
- 5) 近藤 博史. 第4の標準化としてのデータベースの標準化の必要性. 第 36 回医療情報学連合大会 36th JCMI(Nov.2016) :164-165.
- 6) 田淵哲明. 戦略的経営情報システムとしての DWH の構築. 名古屋経営短期大学紀要 55 (2014):1-11.
- 7) 島川龍載. システムとマネジメントの水平統合における医療データ統合基盤構築と将来展望 (総特集 評判の新技术で進化する最新 HIS 事情) —(医療を先導する HIS の新構築), 月刊新医療 42.11 (2015):37-41.
- 8) 島川龍載. DWH を駆使し病院経営の意思決定への施策と実際 (総特集 折角の DWH を本当に活用できているか) —(DWH 導入がもたらすメリットと課題), 月刊新医療 44.2 (2017):52-56.
- 9) 津久間秀彦ら. 病院情報システムの蓄積データの信憑性のための機能要件. 医療情報学 2010;30:145-148.
- 10) 荒井 康夫. 病院情報システムの DWH に接続した BI ツール活用の試み. 第 33 回医療情報学連合大会 33rd JCMI(Nov.,2013) :140-143.
- 11) 高野 与志哉ら. ETL ツールを中核とした電子カルテデータのリアルタイム加工処理の自動化. 第 34 回医療情報学連合大会 34th JCMI(Nov.,2014):424-427.
- 12) 本多正幸ら. 患者データの 2 次利用システム DWH の役割と課題 - 課題研究会における経過を踏まえて -, 医療情報学 2010;30:1320-1321.
- 13) 久島 昌弘ら. SDM を利用した診療支援データベースの構築. 第 33 回医療情報学連合大会 33rd JCMI(Nov.,2013) :162-163.