

トピック

・「日本の医療 DX の遅れとサイバーセキュリティの遅れは同じです。その 1」

・熊本大学病院におけるデータウェアハウス(DWH)の変遷と今後の展望

第 7 号(2023 年 2 号)では、日本の医療におけるサイバーセキュリティの現状と、熊本大学病院におけるデータウェアハウス (DWH) の 2 つのトピックを掲載させていただきました。個人情報の一次利用はセキュリティが重視され、DWH などの二次利用は、利便性を確保することが重要だとも言われています。本稿のトピックが読者の少しでも皆様のご参考になればと思います。なお、本号最終ページに SDM 研究会の情報も掲載しています。

一般社団法人 SDM コンソーシアム ニュースレター編集部

「日本の医療 DX の遅れとサイバーセキュリティの遅れは同じです。その 1」

Authentication と Authorization わかりますか？

世界の EHR は PHR になりますが、日本の EHR は「EMR の束」です。

近藤 博史

日本遠隔医療学会会長を 2017 年から拝命して 2018 年のオンライン診療以降、医療 DX に関して話す機会を得ています。一方、サイバーセキュリティに関しては 2021 年夏に厚労省からコンタクトがあって医療分野のサイバーセキュリティ関係の調査研究をしています。厚労省からコンタクトがあった当初、私は医療 DX 関係をしたかったのですが、セキュリティから DX の話になればいいでしょうとのことと言われました。実際、2021 年の飯田病院、2022 年の大阪急性期総合医療センターの事件からタイトルのような思いをするようになった。

最近話題の「ランサムウェア」の攻撃手法は 3 つに分かれ、対策はそれぞれ異なります。その前に、なぜ、「ランサムウェア」が出てきたのでしょうか。それ以前のサイバー攻撃はウイルス配布、(ボットネットからの DDoS (Distributed Denial of Service attack/分散型サービス拒否攻撃) 攻撃等でのサーバ停止、リスタート時に入り込み)、秘密情報の取得、裏社会での販売での収益だったと思います。対策は組織ネットワーク入口でのウイルス検知でした。しかし、2010 年頃から垂型の増加、ゼロデイ攻撃で検知されないものが増加し、1 種のウイルス検知ソフトでは見つからなくなりました。これに対して情報流出の通信検知が増えてきました。

そこに仮想通貨が出現し、暗号化する事により情報流出不要になり通信検知されなくなったのです。ウイルスも部品を送り、PC 内で完成するので、PC 内で起動する完成型のソフト (ウイルスか不明のもの多い) を検知する方向になり、EDR (Endpoint Detection and Response) が普及しているのです。EDR には不明のソフトをサーバに集積する機能もありますが、これを活かすには管理体制が必要で委託サービスもあります。この委託サービスを付けた販売が経産省の「お助け隊」サービスです。「ランサムウェア」の 3 タイプを (図 1) にまとめます。

図 1. ランサムウェアの 3 タイプ

- ・ランサム (Ransom: 身代金) ← 仮想通貨により足跡の残らず直接収益 (vs 裏市場販売)
- ・「パラマキ型」: メール添付 (ファイル、暗号化ファイル、オフィスマクロ) URL
 - ・対策: 利用者教育、検知、EDR → ネットワークの階層化と仮想ブラウザ導入
- ・「標的型・侵入型」: FW, VPN の脆弱性を
 - ・対策: 外部接続含むネットワーク全体図+ネットワーク資産台帳作成 & 脆弱性情報の常時監視
- ・「サプライチェーン型」: 接続相手側からの侵入 (VPN, https, 専用回線も)
 - ・対策: FW+DMZ+GW の設置 & 監視
 - ・人と機器の Authentication & Authorization: 認証と権限付与 (管理者、利用者)
 - ・GW 通信の標準化
 - ・ネットワークの階層化

→バックアップがより重要になってきました

大阪城の守りで考えると「パラマキ型」: 内部住人にスパイを作る。→ 住民と天守閣の距離。「標的型」: 抜け道トンネルのドアの鍵がかかっていなかった。→ 鍵の確認。「サプライチェーン型」: 抜け道トンネルから敵が攻めてきた → 抜け道入口に門、枳形虎口、城内に隔壁で区切る、、、



目次

「日本の医療 DX の遅れとサイバーセキュリティの遅れは同じです。その 1」

近藤 博史 1

熊本大学病院におけるデータウェアハウス(DWH)の変遷と今後の展望

山ノ内 祥訓 4

SDM Tips 7 鈴木英夫 6

事務局より (SDM 研究会のお知らせ) 7

①パラマキ型：これは 2017 年英国医療システムで大きな被害を与えた WannaCry や最近では Emotet になります。メール添付のオフィスファイルのマクロ、文書中の URL、圧縮ファイルの中、その文書の URL などから PC にインストールされ、ネットワーク内にばら撒かれ、ウイルスは Windows ActiveDirectory, Backup 先の PC, サーバのファイルを暗号化します。暗号化された機器にウイルスが見つからないこともあるので調査は大変です。②脆弱性からの侵入型（標的型）：VPN-FW(Virtual Private Network-Firewall)機器の脆弱性について、ハッカーが侵入し、何日もかけてネットワーク内を調査して、重要ファイルを暗号化します。ハッカーが操作するので準備周到で要求金額も大きくなります。対策は最近言われている通り、外部接続を含む全体のネットワーク図の作成とネットワーク機器、サーバの資産管理台帳の作成です。台帳には機器の正確な機種情報、OS、ソフトのバージョン等を記入し、最新のものにしておくこと、それに機器の最新の脆弱情報を収集し管理台帳に該当機器がないことを確認し続けることです。小規模施設では難しいことであり、保守委託業者のサービスに入れるべきでしょう。③サプライチェーン経由の侵入：安全に機能している外部接続通信の相手側からの侵入です。管理外のところでハッキングされていない保証はないので、対策しておくことは当然でした。しかし、多くの医療機関は専門家だからと信用していました。過去にはソフトウェア販売会社がハッキングされ市販ソフトにウイルスが混入していた例もありますので、IT では「繋がってれば、危険は必ず存在する。」のが常識です。パスワードで成立している仕組み故、たまたま一致する確率はゼロにはならないのです。ですから対策は「現時点で安全性が高いものを選択する。」ことです。それよりも逆に相手側の入口に FW、DMZ(DeMilitarized Zone)、GW(Gateway)が無い。直接内部に繋がるとは非常識です。日本のお城にある「枳形虎口」図2は DMZ と監視のための GW です。データが標準化されていると GW での排除が容易です。現状でも放射線部門、検査部門の機器の保守、遠隔画像診断サービスなど外部接続の多くは業者任せです。

もっと非常識に思うのは、ネットワーク、サーバの ID、パスワードを医療機関が管理しているでしょうか。経験からですが、業者は導入時の利便性で「共通の ID、パスワード」を使用します。病院は引き継ぎで知らされ、保守を頼むわけですが、保守の利便性から「共通の ID、パスワード」の利用のままではないでしょうか。それも「管理者権限」です。また日本では NIST(National Institute of Standards and Technology)の Zero Trust Architecture の端末と人の Authentication と Authorization がわかる方は少ないのではないのでしょうか。具体的に言うと Authentication 認証は ID パスワード、多要素認証で理解されていますが、Authorization 権限付与が十分ではないと思います。管理者権限と利用者権限です。利用者権限であればソフトの利用に限られますが、管理者権限であれば、ソフトの導入も容易になってしまいます。

地域医療連携で SS-MIX2 のファイル転送では利用者権限で十分です。サーバ設定で通信プロトコル、相手アドレスも限定すると安全性は向上します。逆に言うと外部接続で相手を限定しない設定を要求された場合には交渉が必要です。保守接続であれば、接続時間の制限、ソフトウェア更新時には管理者権限が必要になりますが、常時権限利用可能にしない、あるいは現地保守にするなどの対策が必要です。逆に、保守用 PC を専用 1 台にしてセキュリティ対策を十分させるのもよいかも知れません。

政府の資料では、VPN-FW 接続の場合のみに DMZ 等記載している状況ですが、外部接続すべてに言えることです。https 接続、専用回線接続で危険な状況が広がっています。これまで保守の専用回線に利用してきた ISDN のサービス終了から安価な携帯デジタル通信のサービス (LTE) が普及しています。ネットワーク機器や PC に USB で接続するだけです。https 接続では持ち込み PC 一台が外部の Web サーバにアクセスを許可するだけで、DICOM(Digital Imaging and Communication in Medicine)画像を取得しレポートを返すサービス、保守サービスを提供しています。病院管理者からはわかりにくい状況ですし、各部門担当が理解していなければ、携帯電話類似機器を USB 接続するだけなのでわからないと思います。(図3)

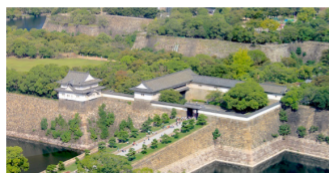


著者 近藤 博史
遠隔医療学会会長

経歴

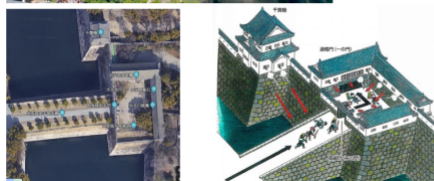
- ・ 1981 年 大阪大学卒。大阪大学医学部附属病院、第 3 内科、放射線科で研修。
- ・ 1984 年 放射線科スタッフとなり、臨床研究開始。
- ・ 1993 年 大阪大学病院の移転に伴いインテリジェントホスピタル計画で一人の放射線部門のデジタル化担当として、アナログ画像とデジタル画像、画像圧縮の診断能研究、放射線部門のペーパーレス化、依頼から会計レポートの電子化と研究基盤 DB 構築する。
- ・ 1997 年 医療情報部副部長となる。
- ・ 1998 年 経産省の四国 4 県電子カルテネットワーク事業で 5 検査会社の検査結果を HL7 で診療所にオンライン配信する。
 - ・ MINCS 誘致、第 1 級陸上特殊無線技士免許取得、このこともあり NASDA 衛星利用促進委員会に参加。
 - ・ 徳島大学病院では日本初の ICU システム（現在のフィリッポスのシステム）と NEC システムの HL7 接続対応。
 - ・ 注射オーダ導入にあたり、NEC 標準システムの説明を止め、NEC 担当部長に病院運用を調査して頂き、類似運用の病院（東京医療センター）を病院 WG 主査を派遣し、導入仕様を決定。
- ・ 2001 年鳥取大学医学部附属病院医療情報部長、教授となる。
 - ・ 総合メディア基盤センター-米子サブセンター長としてキャンパスネットワークの管理し、数年単位のセキュリティ対策を進める。
 - ・ NASDA と共同研究で衛星利用在宅医療実験 農家庭にパラボラアンテナ設置し、TV 会議+ファイル転送機能を使用する。
 - ・ 2003 年鳥取大学医学部附属病院で電子カルテ 100%稼働 国立大学病院初。

図2. 枳形虎口



虎口：敵の侵入経路を限定する。
外部接続 (VPN、https、専用回線) をできれば、まとめる。

枳形：入ったところを囲み、確認、攻撃する。
DMZ領域にGWを配置して、確認対応する。



(<https://shirobito.jp/article/794>)

図3. 携帯デジタル通信機器

・携帯デジタル通信機器が、安価な専用接続として広がっている。



③の攻撃が容易な現状は、どうして起こっているのでしょうか。①医療機関側は専門家に保守委託しているが、管理責任者としても技術に疎い状況は理解できるでしょう。②ベンダーの専門家はどのように常識が教育されていないのでしょうか。簡単です。彼らは経験がないのです。日本の医療分野のネットワークで2つ以上の別組織のシステムの接続はほとんどされていないのです。日本の医療システムではデータ統合したデータベースが動いている事例が少ないと思います。

私はこれについて日本の医療サービスが診療所と中小病院は半分以上を占めていることに起因すると考えています。1980年代後半から欧米の病院見学をする機会を得ていたのですが、医療機関にサーバ、ネットワーク技術者が直接雇用されていました。彼らはベンダーにも異動する機会がありました。当初のコロナ禍で日本の医療機関のサービスが欧米よりも維持できたのは小規模のためにクラスターの発生の影響が医療サービス全体に対して少なかったからだだと思います。さて、技術者を直接雇用しない日本では、まとめて一社に導入、保守契約することになり、専門家への委託も増えます。業務維持に困り込み傾向のあるベンダーは他社との連携、統合はしない、苦手になって来たと考えます。技術を知らない医療機関は専門家ベンダーの説明に従う構図になってきた。

例えば、日本の EHR ですが、F 社も N 社も作っていますが、2 社のシステムは 1 社の系列あるいは 2 社並列で使っていますが、どちらの場合も 1 つのデータベースに統合されていません。実際の画面を考えても、患者選択し、病院選択してそこでの患者の時系列データを見るだけです。少なくとも病院間でデータ連携したら、2つの組織のデータが1つのデータベースになって時系列で出て欲しいものですが、そうありません。結局、その病院のその患者のデータが「見える」だけです。先日、日本政策銀行に資料(図4)を提出し、海外の EHR はそのまま PHR サービスができていのに、日本ではできないことがおかしかったです。海外の EHR を統合導入している鳥取のおしどりネットでは1つの統合 DB 化しているので複数医療機関のデータを時系列に出せます。アクセスを患者に許可すれば PHR になります。コロナ禍で政府が作った感染者 DB、ワクチン接種 DB、重症患者 DB それぞれ担当者が入力して Web 表示されますが、統合されていません。医療 DX の1つである mobile Health のデータもベンチャーがデータをクラウドサーバに入れています。病院の EMR(Electronic Health Record)には連携されていません。患者はベンチャーと主治医に参照権限をあたえる同意をして、主治医はインターネットのホームページで参照しています。

本来、これらの医療システムは患者毎にデータが時系列に表示され、初めて薬の効果やアレルギーがわかるようになるので、EHR と言っても PHR(Personal Health Record)としての表示ができることを医療関係者、患者は求めているはずですが。その目的が達成されていないのです。EHR はその上にある薬の投薬患者の検査結果、感染症の発生、重症化率など研究的、政策的情報をもたらすことが求められているはずですが。

コロナ禍では個人情報保護法の壁とも言われましたが、最初のデータ登録で患者から最終的な統合の許可を得ていれば、可能だったと思います。個人情報保護法はシステムの管理責任を中心に考えますが、政府のシステムとして都道府県のシステム融合を最初から計画して同意を得ていれば可能になったと思います。最終目的の想定ができていなかったと言えます。

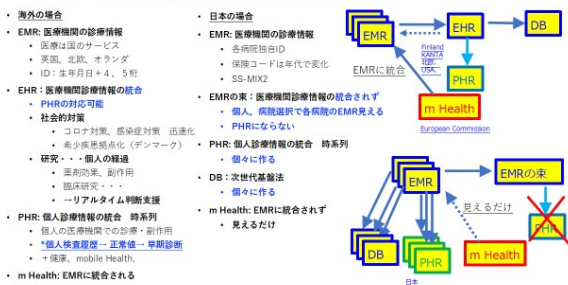
最後にサイバーセキュリティに対抗する技術を述べます。私は鳥取大学医学部附属病院で最初に 100% 電子カルテ稼働を 2004 年に達成しました。すぐに、手術室の休憩室からホームページからのウイルスダウンロード事件を経験しました。対策ソフトで被害はなかったのですが、検知逃れを危惧し、2008 年のシステム更新時に日本で初めて電子カルテ基盤にシンクライアントを導入しました。内部ネットワークの階層化、院外バックアップもしてきました。2020 年にはコスト削減、省エネ、高速化を目的に統合仮想サーバ上に PACS, 電子カルテ部門など載せました(図5)。このことはセキュリティ上も重要でネットワークとサーバ監視は地元ベンダー保守に統合されています。これらのオンプレミスクラウド技術導入で思ったのですが、これらの技術はクラウドサーバで実現していることです。セキュリティ向上にクラウド化が叫ばれていますが、オンプレミスで同様のことをしてもセキュリティレベルは同じだと思います。それよりも現状のクラウドサービスが複数の病院対応がされず、スケールベネフィットが十分でないこと、クラウドサーバとの通信の問題、ガイドラインで医療現場に通信切断時対応のため、診療に支障のないデータ保存を現場に求められている現状では、我々の仕組みが推奨されると考えています。

経歴 (続き)

- ・ 2005 年から 2007 年科学技術振興調整費研究「衛星利用在宅医療と災害医療支援実験」鳥取、沖縄の患者宅にパラボラアンテナ設置在宅医療支援開始する。また、モバイル機器にして鳥取県の防災訓練で利用する。
- ・ 2008 年電子カルテ基盤にシンクライアント導入。翌年、このことから日本ガートナーで初の医療系の招待講演を行なう。
- ・ 2010 年地域医療再生基金で鳥取県地域医療連携「おしどりネット」をシンクライアント基盤で相互参照として開始。
- ・ 2012 年おしどりネットでは、病院からの SS-MIX2 データを交換し、IHE ~XDS/XDS-I で管理し (IBM がカナダで開発した XDS, GE がベルギーで稼働させた XDS-I を結合統合した。)、ソフト vpn とシンクライアント基盤で参照する仕組みで運用する。)
- ・ 2016 年おしどりネットサーバに大容量フラッシュメモリー導入し高速化を図る。
- ・ 2017 年日本遠隔医療学会会長となる。NICT の委託研究の専門員となる。
- ・ 2019 年-2020 年国立大学病院医療情報部長会長となる。
- ・ 2020 年病院システム更新に、電子カルテと PACS のサーバ・ハードウェアを仮想サーバに統合し非医療系ベンダーに導入保守をさせる契約を初めて行う。これにより電子カルテの更新を無期限に延長できるようにした。また、オールフラッシュサーバにより表示時間6割の高速化、省エネ化、省スペース実現する。また、セキュリティ上の向上も図った。
- ・ 2020 年厚労省の指定研究 CS と ISAC のための調査研究代表に指名される。NPO 法人おしどりネット開始、副理事長となる。
- ・ 2021 年 AMED 評価委員就任、ICON (ICU collaboration network) 理事就任

- ・ 2022 年 3 月鳥取大学を定年退職し名誉教授となり、協立温泉病院 院長就任。
- ・ 2022 年 4 月神戸情報大学院大学 特命教授、T-ICU 顧問、PureStorage 顧問となる。ICU 関係の DB 構築と AI のための AMED 研究分担者となる。おしどりネットを iPhone, iPad で参照可能にする。診療所医師からもテキスト情報と画像情報の入力できるおしどりノート運用開始。SS-MIX2 のバックアップ、おしどりネットを介した参照サービス開始。

図4. EMRからSEHR, PHRと、DB構築



病院情報基盤の仮想統合化はサイバーセキュリティを向上する



図5 病院情報基盤の仮想統合化



著者 山ノ内 祥訓

熊本大学病院 総合臨床研究部 研究データ管理センター 特任助教

一般社団法人 SDM コンソーシアム 理事

略歴
2002 年東海大学工学部卒。

同年熊本県内のソフトウェア会社で主に医療介護分野のシステムエンジニアとして勤務。

2015 年熊本大学病院に
入職し現在に至る。

主な業務は臨床研究におけるデータ管理及び IT システム全般。

2022 年熊本大学大学院医学教育部博士課程修了。

博士(医学)(医療情報医学専攻)
医療情報技師

熊本大学病院(以下当院)の診療データウェアハウス(DWH)は病院情報システムの一部として導入されている。初回の導入は電子カルテシステムが稼働した 2011 年である。その後、2017 年に現行の病院情報システムの更改に合わせて DWH も更改が行われており現在に至っている。また、2024 年に次期病院情報システムの稼働を控えている。このような状況の中で過去の DWH がどのような方針に基づき構築されたかという経緯と共に運用上発生した課題を振り返ることと、今後稼働する次期病院情報システムにおける DWH を含めたデータ利活用環境構築への取り組みを紹介する。

第 1 期 DWH

最初の DWH は 2011 年の導入であるが、それ以前にもオーダリングシステムや部門システムなどで簡易的な DWH が存在しデータの検索や抽出が行われていた。ただ、本記事では院内統一的なデータ利活用環境を整備したということで第 1 期と呼称している。この DWH の構成図を図 1 に示す。院内統一的な DWH を整備するとしたものの対象であるシステムは電子カルテシステムと医事システムの 2 つであり、それ以外の部門システムは蓄積対象外だった。また、DWH はクライアントサーバシステムであり、DWH を使用する部署の電子カルテ端末に DWH クライアントアプリケーションをインストールする必要があった。

当院ではデータ利活用のためにはより多くのデータをリアルタイムに蓄積することが重要である、という考えに基づき構築を進めた。電子カルテシステムは、データベースに格納されている主要なテーブルを DWH への連携対象とした。データ連携はデータベースリンクとマテリアライズドビューを併用した構成であり電子カルテの更新データがほぼリアルタイムに DWH で参照できる構成となった。一方、医事システムは、統計データ以外のデータ公開ができなかったため、日次や月次の統計処理に合わせて DWH にデータをアップロードする構成だった。

データ利用者は、データソースのテーブル構造をそのまま使用する場合、ビューを定義してそれを使用する場合、データマートによる集計処理を実行しその結果を使用する場合、の 3 つを使用できた。作成したビューやデータマートの活用については、他の職員と共有することが可能で医療情報担当など技術的に詳しい人間が作成したものを業務部門が検索日付などを変更して使用する、といった利用が行われた。

第 1 期 DWH の最も大きな課題は、電子カルテとのデータ連携がデータベースリンクを用いたことで実行速度が非常に遅くなった、ということであった。稼働前の段階の予測でも、ある程度実行速度が遅くなることは想定されていたが、実運用に入ると想定以上の速度低下に

より単純な検索でも数分～数時間かかることが多かった。そのため、一部の利用ユーザのみが使用する状況だった。その一方でデータ連携がリアルタイムであることからいわゆる 1 次利用や 1.5 次利用の割合が比較的高かった。

この実行速度の課題を解決するための一つの方法として 2014 年後半に Business Intelligence (BI) ツールの導入も病院情報システムの調達とは別に行われた。BI ツールはデータを取り込みメモリ上に展開するタイプのため、いったん読み込んでしまえば高速に動作することからデータの探索や可視化に有用であり第 2 期 DWH でも継続して運用している。ただし、設定できる人材が少なく限定された用途で利用されている。

もう一つの課題は DWH に蓄積されたデータの範囲が狭かったことである。電子カルテの中でも対象外となったテーブルがあったがそれ以外の部門システムのデータ(特にサマリレポートのデータ)がないことで症例検索に支障をきたすことがあった。

第 2 期 DWH

第 1 期 DWH の課題をふまえ 2017 年稼働の第 2 期 DWH は対象となるシステムの増加と実行速度の改善を目指して構築した。第 2 期 DWH の構成図を図 2 に示す。DWH 本体は現行のリレーショナルデータベースに変えてポストリレーショナルデータベースを採用した。データ連携も電子カルテのデータを直接参照することはやめて、DWH ベンダーによるデータ連携アプリケーションによるリアルタイム連携に変更した。対象となるシステムの範囲も電子カルテシステム、医事システム、サマリシステム、眼科・耳鼻科カルテシステム、生理検査システム、カテレポートシステム、循環器カンファレンスシステム、統合認証システム、物流システム、に拡充された。このうちリアルタイム連携となったのは電子カルテシステム、カテレポートシステム、循環器カンファレンスシステム、の 3 システムである。なお、医事システムは更改されなかったため対象データは統計のみだった。DWH クライアントは Web システムに変更され院内の全電子カルテ端末から利用できるようになった。

一方、データ利用者の利用方法に変更はなく、テーブルの直接検索、事前定義ビューを用いた検索、データマートの結果をもとにした検索の 3 パターンである。ただ、利用権限については整理され通常ユーザは検索のみ、一部ユーザのみデータのエクスポートが可能となった。また、利用可能なテーブル、ビューも選択できるようになっており職務に応じた最小限のデータのみ閲覧可能とすることができるようになった。

第2期 DWH の効果は、まず、対象システムが増えたことでより広範囲の検索が可能になったことである。退院サマリ監査等即時性を有しない業務においては DWH を用いたデータ活用が増えており業務改善に寄与している。また、心電図やエコー、心カテ、といったこれまで部門にのみ格納されてきた検査の数値データも検索できるようになったことで臨床研究での利用が増えている。次に、前回より速度が上がったため利用ユーザーが増えたことである。検索テンプレートを作成するには至らないが、医療情報担当に作ってもらった検索テンプレートを利用して定型的な検索を行っている部署が多くなった。これは検索速度が向上したことでより業務フローに組み込みやすくなったと考えられる。

第2期 DWH の課題は、第1期からの継続課題である実行速度が挙げられる。第1期に比べてデータ量が增大しているにもかかわらず第1期より速いレスポンスなのは確かだが、それでも遅いという点がある。次に、データソースのテーブル構造をそのまま DWH に格納しているため正規化された構造となっており、業務部門が構造を理解しづらいため、検索上の課題となっている。テーブル中の各種区分やフラグの説明も不足しているため検索を行うまでの事前調査に時間がかかっているのが現状である。また、事前定義ビューやデータマート作成の手間が大きくなかなか活用されていない。そのため、検索テンプレートでの検索が主流となっている。最後に、リアルタイム連携となっているが更新頻度の高いテーブルは負荷のかかる日中で1時間程度遅れて反映されることがあり1次利用で運用するのは困難だった。

第3期 DWH に向けて

これまで2世代の DWH を運用したうえで課題となったことを列挙すると以下の通りまとめられる。

- 検索したときの速度が遅い
- 検索できるデータの種類の少ない
- 1次利用を想定したリアルタイムデータ連携になっていない
- エンドユーザにわかりやすいデータ構造になっていない

これらの課題をふまえて現在第3期 DWH の構築を開始しようとしている。第3期 DWH の構成図を図3に示す。データベース本体はリレーショナルデータベース製品に回帰するがカラム型ストレージを使用することでデータの圧縮および検索性能が向上すると見込まれる。また、対象となるシステムも現行に加えてナースコールや放射線、リハビリなど多くの部門システムとの接続を予定している。データ連携のリアルタイム性についても DWH ベンダーのデータ連携ツールだけでなくエンタープライズサービスバス(ESB)というデータ連携専用のシステムを構築する。これにより他システムとのデータ連携を容易にし、業務上のデータ連携にプラスして DWH にデータを蓄積していくことが可能になる。

DWH の構成は、データレイクとマート済 DWH の2つに分かれている。データレイクはこれまで同様に各システムから送られてきたデータをそのまま保存する場所とし、マート済データはデータレイクをもとに用途別に自由にデータマートが行える場所として定義した。このマート済 DWH には SDM を使用することを予定している。SDM を用いることでエンドユーザにわかりやすいデータ構造とすることができる。また、DWH には簡易的な BI ツールも含めるようにしたため1次利用や1.5次利用で頻用される診療ダッシュボードや臨床指標モニタリングにも利用できると考えている。そのため、基本的にはマート済 DWH を利用者は検索することになるが、必要時にはデータレイク側も検索できるようにしてアドホックな検索ニーズに対応できるようにする予定である。また、ESB と連携することで DWH に一方的に蓄積するのではなく、今後接続が見込まれる診療予測や診断支援などの AI 系システムに必要な学習データのデータソースとしての役割も想定している。

一方、第3期で断念した点はクラウドサービスの利用である。クラウドサービスは BI も含めた機能が豊富であることに加え、蓄積されるデータの増加を支えるだけのストレージ容量も安価に準備することができる。また、使用頻度が今後増えるといっても通常の業務システムよりは低い稼働率を考えるとクラウドサービス上に展開するほうが安価である。セキュリティについても現状主要なクラウドベンダーでは豊富なセキュリティ機能が備わっているため技術的な問題は見られなかった。しかしながら、調達及び契約上の兼ね合いでクラウドサービスは原則調達対象外となったことで同時導入することはなくなり、今後別途検討することとなっている。

本記事ではこれまで当院が導入した DWH の構成と課題についてまとめるとともに次期 DWH はどのような方針に基づいて構築したかを紹介した。第3期 DWH の稼働は2024年1月を予定しているため、その結果はまた別の機会で紹介したい。

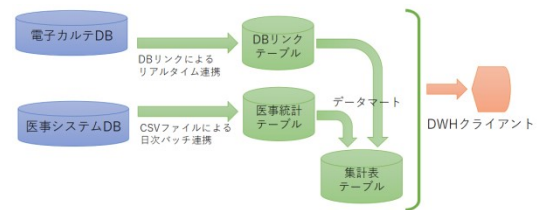


図1. 第1期DWH

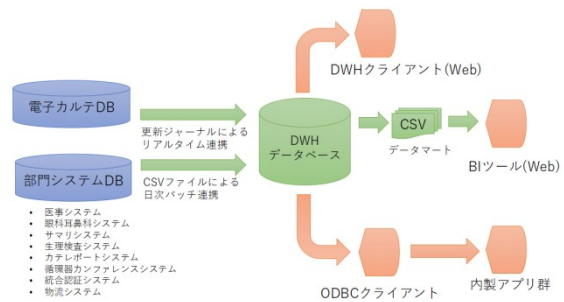


図2. 第2期DWH

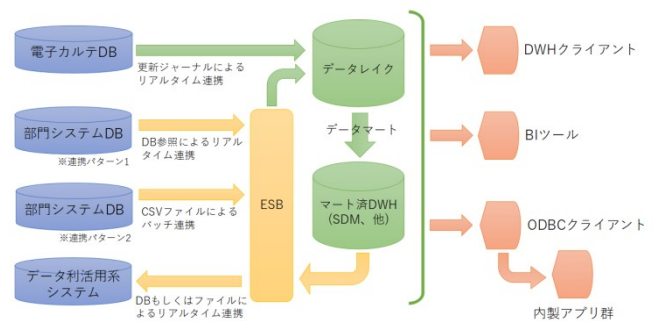


図3. 第3期DWH



前回までで、SDM の基本的な考え方やモデルについて説明してきました。ここからは、SDM の実際の利用方法を説明する予定でしたが、今回は、SDM 構築までのまとめとして、SDM における ETL (抽出、変換、登録) に関して解説することにします。SDM の ETL は、抽出元のデータベースに接続して、データベースの複数のテーブル (表) から、エンティティ (項目) に格納されているデータを抽出し、SDM のテーブルの項目へ格納する一連のプロセスを自動化したものであり、SDM の各テーブル単位で作製されます。以下、その手順を解説します。

(1) 設計 (図)

① テーブルマッピング

テーブルマッピングは、文字通り、テーブル対テーブルの対応付けのことであり、SDM の各テーブルを生成するために必要な、抽出元のテーブルを列挙し、対応表を作製します。

② 抽出条件

SDM テーブルの1レコード (1行) を決定するための抽出元のテーブルに対する抽出条件を決定します。

③ エンティティ・マッピング

SDM テーブルの各項目に必要な抽出元の項目とその抽出条件、および変換ロジックを項目ごとに作製します。

(2) 開発

設計書に従い、ETL を生成しますが、その方法は、市販のソフト (フリーウェアを含む) を利用する方法と、プログラムを開発して行う方法があります。

① 市販ソフトの利用

いわゆる ETL ツールという ETL 専用ソフトの場合は、GUI(Graphical User Interface)で、入力側 (抽出元) と出力側 (登録先) のそれぞれの DB をオンラインで接続し、入力側からの抽出条件の設定や、型変換、数式など変換ロジックを実現するための様々な関数、DB 登録の方法 (挿入、修正、

削除) など、RDB(Relational Database)をコントロールする SQL(Structured Query Language)などの知識が無くても、ETL 設計図があれば、ETL を作製することができるツールです。ただし、関数の不足や処理速度があげられないなど、問題が生じた場合は、ユーザー側での対応が難しいというデメリットもあります。

② プログラム開発

ETL プログラムを開発する場合は、プログラミングのスキル以外にも、環境設定、DB 接続、SQL、およびパフォーマンスチューニングに関するスキルが必要になります。特に、パフォーマンスに関しては、SQL を埋め込む場合と、プログラム言語側でロジックを組む場合で、大きくパフォーマンスが変わることがあります。抽出元の DB 環境、ETL 実行環境、SDM の環境が、すべて違う環境で稼働している場合は、SQL は、それぞれの DB が稼働している環境のエンジン (主に CPU:Central Processing Unit) とメモリを使用していて、ETL 実行環境のエンジンとは別なので、SQL で表現できる部分は、SQL の埋め込みで行う方が多いのですが、DB のデータモデルにも依存するので、そのあたりの修正ができるかが問題となります。

(3) テスト

ETL は、毎回全データを対象とするのではなく、差分データを取得するのが目的なので、抽出元の DB から差分を取得するロジックを含めたテストが必要になります。まずは抽出元の DB 接続と、抽出元にテストデータがある場合に、そのテストデータを対象とした ETL 実行テストを行います。その結果は、設計上で移行されるべきデータが、SDM テーブル上に存在しているかどうか、レコード数が抽出元と抽出条件に合っているかどうか、また移行されたデータの値が移行元および変換ロジックと合っているかどうか

を、各テーブルで確認します。その確認作業は、SQL で行いますが、DB 管理ソフトや BI(Business Intelligent) ツールを用いた方が効率良く確認できます。

(4) データ移行

ETL は、差分データ抽出のために作製しますが、同じ ETL を用いて過去データを SDM に移行することができます。もし、差分データ取得のロジックが DB 登録の日時であった場合は、日時を抽出元のもっとも古い登録日時に設定すれば良いのですが、その場合には移行だけで数か月を要することもあります。それを避けるために、別の条件、例えば患者 ID などに抽出条件を変更して実行する方法もあります。いずれにしても抽出元 DB にインデックスの定義がされている項目を抽出条件にすることが重要です。

(5) 運用

ETL の運用に関しては、その実行スケジューリングが必要となります。通常の DWH では、日次の実行で、0 時開始で、分散して実行させるなど、抽出元の負荷を避けるようにすることが多いと思います。また、抽出元の DB がメンテナンスのため稼働できない場合は、その時間帯を避けなければなりません。一方、リアルタイムモニターが必要な場合、例えば、現在の病床空き具合や外来の待ち人数の把握などに利用する場合は、リアルタイムでの対応が必要となり、ETL 実行を数分単位で行うケースもあります。ETL タスクスケジューリングは、ETL ツールでは標準機能ですが、負荷軽減のための最適化機能はそなわっていない場合が多いので、導入時の負荷テストが重要になります。

次号からは、SDM の実際の利用方法について解説します。

キーワード
【SDM】
 Semantic Data Model
【ETL】
 Extraction, Transformation, Load
【SQL】
 Structured Query Language

著者 鈴木英夫

MoDeL 代表取締役
 SDM コンソーシアム理事
 千葉大学客員准教授

出身

神奈川県横浜市

学歴

千葉大学
 千葉大学大学院
 学術博士 (自然科学)

職歴

IBM Japan (1983-2014)
 IBM Corp. (1998-1999)
 Findex Inc. (2017-2020)

教歴

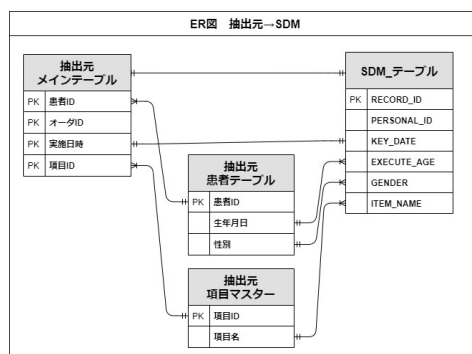
非常勤講師

千葉大学大学院
 東京女子医科大学
 岐阜大学大学院
 鳥取大学医学部

専門

モデル・デザイン
 データ・サイエンス

図 SDM ETL の設計例



一般社団法人
SDM コンソーシアム
郵便番号 223-0066
横浜市港北区高田西 2-4-10

電話
(045)567-3613

電子メール
info@sdm-c.org

ユーザーが育てる DWH

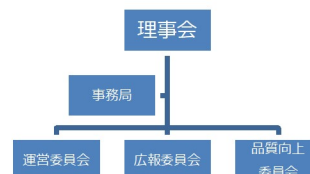
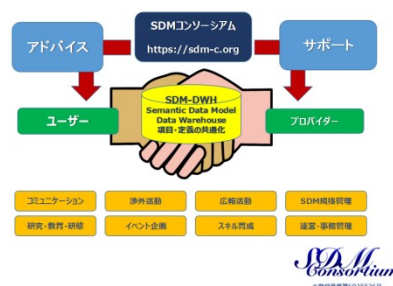
SDM コンソーシアムに関して

SDM コンソーシアムは、

- ・ SDM 設計書の整備・保守
- ・ 分析サンプル作成・公開
- ・ アナリストの教育・スキル育成
- ・ プロジェクトの支援
- ・ セミナー・ミーティング・イベントの企画
- ・ ライブラリの整備
- ・ メンバー・コミュニケーション
- ・ 広報活動

などを通して、SDM 普及を行っております。

医療情報学会春季学術大会 シンポジウム
2023 において、チュートリアル「二次利用
ユーザーのためのデータ抽出、変換、登録
(ETL) 入門」を実施します。



SDM ユーザー研究会のお知らせ (参加費無料)

7月1日(土曜日)午後(医療情報学会春季学術大会 シンポジウム 2023 の最終日)に、SDM ユーザー研究会を実施します。SDM ユーザーに限らずどなたでも参加できます。プロム那覇古島(ゆいレール古島駅前)で開催予定です。現地参加は30名までですが、当日オンライン参加も可能です。詳しくはホームページをご覧ください https://sdm-c.org/download/sdm_workgroup20230701.pdf

一般社団法人 SDM コンソーシアム 2023年4月1日 現在

組織・役員

【代表理事】	紀ノ定保臣	(岐阜大学)	
【監事】	本多正幸	(長崎大学・千葉大学)	* 広報委員長
【理事】			
近藤博史	(鳥取大学・神戸情報大学院大学・協和温泉病院)		* 運営委員長
村垣善浩	(神戸大学・東京女子医科大学・早稲田大学)		
木村映善	(愛媛大学)		
島川龍載	(県立広島大学)		
飯田征昌	(蒲郡市民病院)		* 品質向上委員長
山ノ内祥訓	(熊本大学)		
土井俊祐	(千葉大学)		
鈴木英夫	(株式会社 MoDeL ・ 千葉大学)		
【事務局】	濱田麻里		
【賛助会員】			

Web サイトにてお待ちしております。

Web サイト アドレス:
<https://sdm-c.org>

富士フィルムメディカル IT ソリューションズ株式会社
株式会社医用工学研究所
株式会社ジャストシステム
株式会社ファインデックス
株式会社医療ラボ
IQVIA ソリューションズジャパン株式会社
コスモ開発株式会社
メディカル・データ・ビジョン株式会社
株式会社エムケイエス
データキューブ株式会社
株式会社アドバンスト・メディア
株式会社メディクト