
一般口演

一般口演14

電子カルテ開発・EHR・PHR・連携

2017年11月21日(火) 16:15 ～ 17:45 J会場 (3F イベントホールA)

[2-J-3-OP14-6] 多施設間の安全なデータ収集を目的とした SS-MIX2標準化ストレージ開発の試み田中 勝弥¹, 山本 隆一² (1.東京大学大学院医学系研究科, 2.医療情報システム開発センター)

[はじめに] 現在多くの施設で導入されている SS-MIX2標準化ストレージは、ファイルシステムとしての制約から、患者横断的な診療データの検索に時間がかかる欠点があり、多施設間横断的な診療データの収集に用いるためには、別途データベースを構築し双方を同期させて運用するなどの追加コストが発生する。また、大規模なデータ収集時は解析に必要なデータ項目が含まれる HL7ファイル群の収集が必要となり、データ項目ごとの収集制御が困難である。本稿では、SS-MIX2標準化ストレージ自体を単一のリレーショナルデータベース（RDB）により実装する手法を提案する。また、患者横断的なデータ収集における検証についても報告する。

[方法] Oracle DBFS(Database File System)を使用し、SS-MIX2標準化ストレージを構築する。ファイルシステムとしての機能を RDB上に実現し、各 HL7v2形式ファイルは BLOBとしてメタ情報とともに同データベースに格納する。ファイルシステムとしての更新をデータベース上のトリガーにより追跡し、ファイルのメタな更新情報をリアルタイムに保持させる。HL7セグメントごとのデータ項目は、更新情報をもとに定期的にパースし、同一データベース内の別テーブルへ HL7セグメントごとに格納し、外部から検索可能とする。

[結果] 提案した方法により、処方オーダ情報（OMP-01）を対象にパースモジュールを PLSQLを用いて作成することが可能であった。ファイル更新情報をデータベースとして監視することにより、拡張インデックスデータベースの作成も可能であることが確認できた。

[おわりに] 単一の RDBを使用して SS-MIX2標準化ストレージを実現することにより、患者横断検索のためのリポジトリの自動構築が可能となった。処方以外のデータファイルへの本提案方式の適用、大規模データ格納時の性能評価、多施設間での安全なデータ収集基盤としての運用可能性、について今後検討を進める予定である。

多施設間の安全なデータ収集を目的とした SS-MIX2 標準化ストレージ開発の試み

田中 勝弥^{*1}、山本 隆一^{*2}

*1 東京大学大学院医学系研究科、*2 医療情報システム開発センター

Development of a SS-MIX2 standardized storage for secure data collection between multiple institutions

Katsuya Tanaka^{*1}, Ryuichi Yamamoto^{*2}

*1 Graduate School of Medicine, The University of Tokyo, *2 Medical Information System Development Center

This paper describes a development of a SS-MIX2 standardized storage for secure data collection between multiple medical institutions. We tried to add an RDBMS interface for sharing internal data items of the messages stored in the filesystem using a technology based on FUSE. It enables us to build an index of SS-MIX2 storage simultaneously during the message transaction on the filesystem. We also developed PL/SQL HL7 message parser to store the data items to user-defined database tables. Attached to PSI protocol, this storage will play an important role as an infrastructure for secure collection of target data among multiple medical institutions with SS-MIX2 storages.

Keywords: Information Security, Secure Data Collection, SS-MIX2, HL7

1. はじめに

現在多くの施設で導入されている SS-MIX2 標準化ストレージは、ファイルシステムとしての制約から、患者横断的な診療データの検索に対しては、無数のファイル群を横断的にパースすることに起因して、膨大な時間がかかる欠点があり、多施設間横断的な診療データの収集に用いるためには、別途、目的に応じたインデックスデータベースを構築し双方を同期させて運用するなどの追加コストが発生する。標準規格で定められている拡張インデックスデータベースは、標準規格上はオプションの扱いであり、利用者側からの要求がない限り、標準装備されない機能である¹⁾。とくに、SS-MIX2 標準化ストレージ自体を遠隔地にバックアップするためには、差分同期を行う方法が最善であるが、ストレージ装置の機能を利用した差分同期を行う場合を除いては、差分の抽出そのものに時間を要し、拡張インデックスデータベースの構築なしに効率的なバックアップが実現できない。同様の状況が、次世代医療基盤法における匿名加工認定事業者の運用として想定されるスキームのような、大規模な SS-MIX2 標準化ストレージデータ収集時にも発生する。この場合には、SS-MIX2 標準化ストレージをバックアップする形で匿名加工認定事業者に集積し、中央集約された状態で、後段の匿名加工およびデータ解析を実施する流れとなるが、基本的には全データのバックアップが行われる形態であり、無数の HL7 ファイル群の収集が必要となる。トランザクションログを利用した収集を行う形態が試行されている。

一方で、目的が特化された多施設間での小規模な診療データの収集・解析には、次世代医療基盤法で示される全データの収集は本来不要であり、必要なデータ項目のみが患者横断的に収集可能であればよい。JST CREST「ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化」では、宮地らが多施設間のセキュアなクロス集計やデータ収集を目的とした PSI (Private Set Intersection) ライブラリを開発してきており²⁾、このような用途についての実現が期待される。

2. 目的

本稿は、SS-MIX2 標準化ストレージに改良を加え、多施設間の安全なデータ収集を PSI ライブラリの適用により実施可能な SS-MIX2 標準化ストレージを構築することを最終目標と

して、そのプロトタイプ開発を行うものである。

3. 方法

SS-MIX2 標準化ストレージ上のデータを一般的な NFS や CIFS 等のファイルシステムとして利用可能なインターフェイスは確保しつつ、同時に RDBMS (Relational Database Management System) 内のデータとして検索可能なインターフェイスを実現する。2つのインターフェイスは、FUSE (Filesystem in Userspace) を採用することで可能であり、今回は、FUSE 応用製品である、Oracle DBFS (Database File System) を使用した³⁾。構成の概要を図1に示す。

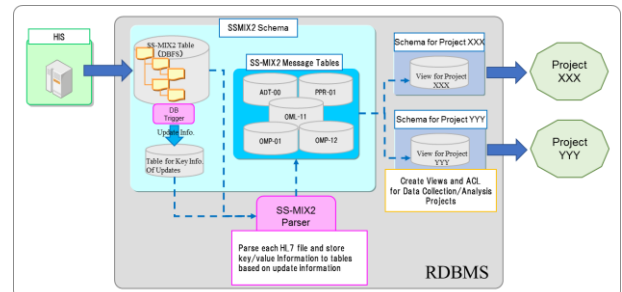


図1 システム構成の概要

ファイルシステムとしての機能を RDB 上に実現し、各 HL7 ver.2 形式ファイルは BLOB としてメタ情報とともに同データベースに格納することが可能である。ファイルシステムへのファイルの作成、削除などの更新をデータベース上のトリガーにより追跡し、ファイルのメタな更新情報をリアルタイムに保持させる。

ファイルシステムに格納されるメッセージファイル内のデータは、トリガーにより生成される更新情報をもとに定期的にストアプロシージャによりパースし、同一 DBMS 内の別テーブルへ HL7 セグメントごとに格納し、外部から検索可能とする。プロトタイプ環境に関する諸元は表1のとおりである。また、処方オーダーメッセージ (OMP-01) を対象としたテーブル定義の概要を図2に示す。データスキーマは、HL7 ver.2 記述形式におけるセグメント、フィールドに従うこととした。

表 1 プロトタイプ環境諸元

CPU	Intel Xeon CPU E5-2640 v4 @ 2.40GHz (20Cx2)
Memory	256GB
OS	CentOS release 6.9
DBMS	Oracle Database 12c Standard Edition

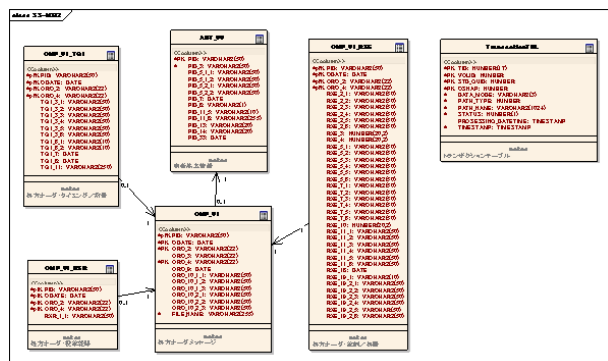


図 2 テーブル構成の概要(OMP-01)

4. 結果

提案した方法により、処方オーダー情報(OMP-01)を対象にメッセージファイルに対するパースモジュールを PL/SQL により作成することが可能であった。HL7 ver.2 のセグメント、フィールド定義に依存しているが、SS-MIX2 標準化ストレージに格納されたメッセージファイル内の各データ項目を SQL ベースで検索・集計することが可能となった。また、ファイル更新情報をデータベースとして監視することにより、SS-MIX2 標準化ストレージ使用で定められている拡張インデックスデータベースの作成もトリガー機構により実現可能であることが確認できた。検証用の OMP-1 メッセージファイル群(ファイル数: 22004、総容量:199MB)に対する性能測定結果は、表 2 のとおりであった。コピー/削除は各々 cp/rm コマンドで行った。

表 2 性能測定結果

コピー:HDD→HDD	102 sec (Ave: 15.6 Mbps)
コピー:HDD→DBFS	1902 sec (Ave: 837 Mbps)
コピー:DBFS→HDD	147 sec (Ave: 10.8 Mbps)
削除:HDD	48 sec (Ave: 458 files/sec)
削除:DBFS	858 sec (Ave: 25.6 files/sec)
メッセージパース	101 sec (Ave: 218 files/sec)

5. 考察

本稿で提案する手法により、SS-MIX2 標準化ストレージに対し、ファイルシステムとしての要件を確保しながらも、同時に汎用検索向けのデータベースとしてのインターフェイスが構築可能であることが示された。本手法では、SS-MIX2 標準化ストレージへのファイル更新がリアルタイムに捕捉される一方で、HL7セグメント・フィールドごとのデータ項目はオフライン処理による定期的な更新であり、リアルタイムには捕捉されない。しかしながら、検索可能なデータ項目の公開がおよそ日単位の更新で充分であると考えられるため、実用的には十分と考える。今回、検証対象としては処方オーダーメッセージのみであるが他のデータ種別メッセージへも同様の手法によるRDBMSへの取り込みは制約がないため、引き続き各種イベントタイプに対するパースモジュールの作成を進める。

拡張インデックスデータベース仕様で規定されており患者横断検索に検索可能なデータ項目は、「診療日」「データ種

別」「診療科」程度であるが、本稿で提案する仕組みにより、HL7フィールド単位での横断的検索が可能となった。実質的には、たとえば、対象となる検査や薬剤に対して、コードや名称が限定される検索要件に対して、多数のデータファイルを精査することなく対象とするセグメントデータ、あるいは、メッセージファイルを特定可能となる。これにより、標準化ストレージ全データの収集や、集計に不要なメッセージファイルの収集を回避できる。

また、SS-MIX2 標準化ストレージを単一のRDBMSとして構築できることで、DBMS付属のレプリケーション機能によるバックアップ方式の採用が可能となる。これは、無数の HL7 ファイル群をリアルタイムや短時間に同期するための要件を充足させやすい機構と考える。

宮地らの PSI は、多施設に分散配置されたデータセット群の間でデータ項目ごとの秘匿収集が可能な技術であるが、今後、本稿で提案した標準化ストレージ機構へPSIを適用するための技術的検討を進める。多施設に配置された SS-MIX2 標準化ストレージから必要なデータ項目を安全に収集するための基盤としてPSIとRDBMSを連携させる。分散モデルとしては、中央にクラウドサービスとして Trusted な収集サーバを用意し、クエリおよびアクセス制御はクラウドサービスとして中央管理が可能な方式を想定する。実装および検証は今後の課題とした。

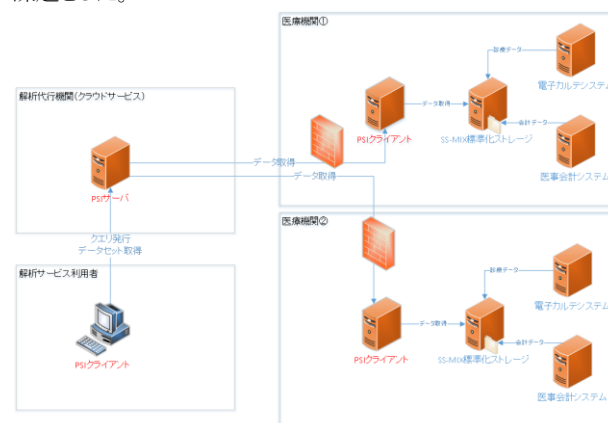


図 3 多施設間の安全なデータ収集システム構成案

6. おわりに

本稿では、SS-MIX2 標準化ストレージに蓄積されるデータを多施設間で安全に収集、検索するための基盤としてPSIが適用可能な標準化ストレージの実装手法を提案した。提案した方法により、処方オーダー情報(OMP-01)を対象にパースモジュールを PL/SQL を用いて作成することが可能であった。対象メッセージの増加、PSIの適用について今後開発を進める。なお、本研究は、JST CREST グラント番号 JPMJCR1404「ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化」により実施した。

参考文献

- 1) 日本医療情報学会(JAMI):医療情報の標準化に関する情報・資料など。 Available: <https://www.jami.jp/jamistd/ssmix2.html>
- 2) A. Miyaji, K. Nakasho, and S. Nishida, "Privacy-Preserving Integration of Medical Data," Journal of Medical Systems, journal article vol. 41, no. 3, p. 37.
- 3) データベース・ファイルシステムの概要。 Available: https://docs.oracle.com/cd/E49329_01/appdev.121/b71284/adlob_fs.htm