公募シンポジウム
シンポジウム4
情報の安全な利活用とセキュリティ基盤技術
2021年11月20日(土) 09:10 〜 11:10 C会場 (2号館1階211)

[3-C-1-05] 診療情報の安全な二次利活用基盤における機能要件
Essential Requirements of Secure Integrated Infrastructure for Secondary Use of Clinical Information

*田中 勝弥[1]（1. 国立がん研究センター）
*Katsuya Tanaka[1]（1. National Cancer Center）

キーワード：Electronic Medical Records, Standardization, HL7, Secondary Use, Privacy Protection

疾患別リポジトリをはじめとする大規模な診療情報データベースが構築・運用されてきており、医療機関の外部と情報を共有・収集し解析する多施設間の研究利用も盛んに行われている。また、次世代医療基盤法の成立にみられるように、集積された診療情報の研究開発や公益目的への二次利用は今後さらに促進されることが期待される。一方で、改正個人情報保護法の施行や医学系研究に関する倫理指針の改正に見られるプライバシー保護や患者同意の管理への対応は必須の課題でもある。平成26年度より実施された JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究プロジェクトでは大規模データの二次利用とプライバシー保護への課題に対し、クラウド上に電送・保管される診療情報の安全管理、分析・二次利用に対するプライバシーリスク評価を中心にその要素技術とシステムイメージについて検討を重ね、パイロットシステムを実証した。
本プロジェクトの成果として、
・分散環境下における安全な診療情報収集機能
・抽出データのプライバシーリスク評価機能
・同意情報の電子化と情報連携機能
・二次利用監査に対するトレーサビリティ機能
などの機能開発を行ったが、現在これらの機能を応用した研究における二次利用を目的とした安全な診療情報利活用基盤の開発を行っている。
本発表では、国立がん研究センター中央病院が参画するいくつかのプロジェクトを例にあげ、セキュリティ要素技術の活用場面と機能詳細、今後の展望について報告する。

情報の安全な利活用とセキュリティ基盤技術

田中　勝弥[*1]
[*1] 国立がん研究センター

Secure Utilization of Information using Security Infrastructure and Underlying Technology

Katsuya Tanaka[*1]
[*1] National Cancer Center

Abstract: The collection and utilization of large-scale medical information for the purpose of developing artificial intelligence engines are being actively carried out. In Japan, various collection systems have been built, which mainly operate to collect data for medical image analysis and disease repositories. The enforcement of the Next Generation Medical Infrastructure Law since May 2018 is expected to accelerate the collection of medical information. In the experimental project for the Next Generation Medical Infrastructure Law, a centrally integrated basic system was developed, and standardized electronic medical record (EMR) storage data that was distributed to each hospital was transferred directly in one data center and imported to a database for secondary use. The law requires a mechanism for maintaining a list of notified or opted-out patients by notifying patients to opt in or out. In a previous project, we developed a gateway system using these functions, which was deployed to medical institutions participating in the demonstration project. When considering operating these systems in medical institutions, safe and efficient secondary use of collected information is essential, not only to abide by next-generation medical infrastructure law, but also for large-scale data collection projects, such as multifacility clinical research. In this paper, we consider in detail the requirements for providing access to medical care information to such data collection projects and propose additional requirements for our developed gateway system.

Keywords: Electronic Medical Records, Standardization, HL7, Secondary Use, Privacy Protection

1. はじめに

人工知能エンジンの開発を目的とした大規模な診療情報の収集と利活用が盛んに行われている。我が国でも、医用画像データの収集や疾患リポジトリを中心に収集システムが構築され運用されている。また、次世代基盤法の施行によって、今後診療情報の収集が加速的に進むと考えられる。

技術的な視点からすれば、大規模な診療情報収集システムとしての要点は、収集対象となる電子カルテデータが標準化されており、収集後即時に利活用可能なことである。我が国では、電子的な診療情報交換規格として、SS-MIX2 標準化ストレージ[1]が国の標準として定められており、各種の収集事業において共通の記述規格として採用される場合が多い。なお、同規格では、患者番号、診療日をキーとして、30 数種類の HL7 v2 メッセージファイルを単一のストレージに保管する構造を有している。

次世代医療基盤法における実証事業においても、中央集積型の基盤システムを構築し、各病院に分散した標準化ストレージデータを整形なしに、直接暗号化して 1 か所のデータセンターへ収集後、データベースにインポートして二次利用する方式が採用されている。また、同法が求める、通知によるオプトイン、オプトアウトによる同意撤回、を実現するために、通知済み患者、オプトアウトした患者のリストを保持する機構が開発された。また、これらの機能を有するゲートウェイシステムを開発し、実証事業に参加する医療機関へ配備した。

医療機関規模で考えたとき、次世代医療基盤法事業だけでなく、多施設臨床研究など、複数の大規模なデータ収集プロジェクトへ、安全かつ効率的なデータ提供が求められるケースが多いと考えられる。著者の所属する国立がん研究センター中央病院でも、AMED による医療技術実用化総合促進事業「Real World Evidence 創出のための取組み」など、大規模なデータ収集を目的とした研究プロジェクトに参画しているが、いずれの事業においても各医療機関に存在する SS-MIX2 標準化ストレージが二次利用対象となるデータソースの中心として扱われている。とくに、標準化ストレージ内データの品質向上が注目されているが、

1. 部門システムでの標準コード変更に対する部門側での人的コスト
2. 治験等、第三者に暴露してはいけない情報への配慮
3. 収集事業ごとに構築される収集システムの維持、管理への負担

といった課題を解決すべく、データ収集に用いるゲートウェイシステムは、研究を目的とした情報保護だけでなく、利活用の観点から多様な機能が要求される。

本稿では、このようなデータ収集事業への診療情報提供時に考慮すべき要件を詳細に検討し、これまでに開発したゲートウェイシステムに追加すべき機能を提案する。

2. 方法

2.1 設計の要点

これまでに 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」（代表：宮地充子）において開発した診療情報収集用ゲートウェイに対して、本稿で新規に開発する診療情報収集用ゲートウェイの基本機能要件を以下に列記する。

2.2 技術的要件

2.2.1 検索が可能な標準化ストレージ

SS-MIX2 標準化ストレージは、記述規格として、HL7 v2 メッセージを採用し、患者番号、診療日付、イベント種別によりソートされた一連のファイルシステムであり、患者横断検索や複数のストレージ横断検索にはその性質上不向きな構造を有している。また、標準規格として規定されるインデックス DB はファイルの検索を可能とするものの、ファイル内コンテンツに対する検索には対応していない。本稿のシステムでは、SS-MIX2 標準化ストレージ上のデータを一般的な NFS（Network File System）や CIFS(Common Internet File System)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設ける方式を採用する。

図1に実装の概要を示す。電子カルテシステムから出力される SS-MIX2 メッセージファイルは、ファイルシステムとして本ゲートウェイストレージ内の仮想ファイルシステム（Filesystem in Userspace, 以下 FUSE と呼ぶ）に複製し、複製されたメッセージをシステム内で解析し、HL7v2 のセグメント、フィールドの定義に従って、RDBMS 内のデータ格納用テーブルに保持する構造としている [2]。

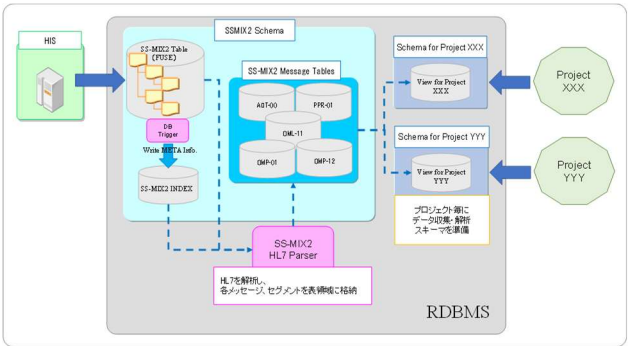

図 1 本システムの標準化ストレージ実装概要

2.2.2 標準化

メッセージファイル内の主要な記述コードとして、病名、検体検査、処方薬、などがあげられるが、コードマスタとしては標準化ストレージの規定にしたがって、わが国の厚生労働省標準を採用する。実際には、オーダエントリシステムが標準コードに対応できておらず、標準化ストレージにローカルコードで出力される場合も散見されるため、ローカルコードから標準コードへのマッピングにおける対応表を別途作成し、保持させる必要がある。後調査が可能なように、標準コードの付記はローカルコードを維持するために、HL7 v2 における ER7 形式で追記する。

また、本項では、病名、検体検査、処方薬の3つのマスタについて、当院電子カルテシステム内のマスタと標準コードとのマッピングを行い、突合可能性を検証した。

2.2.3 同意情報

臨床研究における大規模な診療情報収集は、患者同意に基づいて行われる。同意の方式として、オプトイン、オプトアウトに対応可能な機能が求められる。我が国では現時点で、研究参加および拒否への意思表示は紙面への署名によって運用されるため、署名後の同意書を電子化し、データ抽出機能と連携させる。電子的な表現形式は、HL7 CDA により実現し、電子カルテシステム本体およびスキャン文書管理システムから拡張ストレージへ掃き出し、保持する。

本研究では、同意情報の記述様式として、HL7 CDAR R2 Implementation Guide: Privacy Consent Directives, Release 1[3]で規定される記述規格による同意文書情報を XML として、上述のデータ検索が可能な標準化ストレージへ格納し、データベース上で検索可能とする方式を採用する。

2.2.4 多目的利用

標準化ストレージをデータソースとする臨床情報収集研究が複数実施される状況にあり、また、それぞれの研究プロジェクトごとにデータの加工方法（採用するコードマスタ、データの加工精度など）が異なることが生じうる。本研究では、オリジナルの標準化ストレージを医療機関に1つだけ配備する前提とし、データ収集プロジェクトごとの抽出条件、データ変換規則を定義可能としておく。複数の研究に対し異なるルールでの整形出力が可能となるよう整形ルールの保持機構、整形後の標準化ストレージデータをアーカイブする領域を収集事業ごとに配備する方針としている。

2.3 セキュリティ要件

2.3.1 情報保護

さまざまな目的のデータ収集研究が対象として考えられるが、電子カルテシステム内に存在するデータのうち、一定の症例については第三者への提供を制限する必要がある。具体的には、治験実施患者など契約に基づいて規定される場合が考慮される必要がある。患者プロファイルや診療実績に応じて、第三者提供が可能かどうかを判別し、収集プロジェクトへ提供するかどうかを制御する機構が必要となる。

上述のように、観察研究におけるオプトアウト同意のように、紙面による署名をもって、研究参加を拒否できる場合には、同意書面のスキャンデータを個別のデータとして、検索可能なストレージ領域へ保持することにより、研究不参加となった患者の情報については保護可能と考えられるが、治験参加患者については、同意書面が原本管理として電子化されていない状況があるため、たとえば、治験管理システムや、患者プロファイルなどの電子カルテシステム内の情報を抽出し、二次利用を不可とする患者リストを生成する方式を採用する。

2.4 実データによる検証

上述の諸機能を実装する場合に考慮すべき課題について、国立がん研究センター中央病院の診療データを事例として参照し、実装方法への対応を検討した。

3. 結果

3.1 開発したシステムの概要

設計したゲートウェイシステムの概要を図 2 に示す。電子カルテシステムから、出力される SS-MIX2標準化ストレージデータをメインデータソースとして位置付ける。

電子カルテシステムのオプションとして導入される標準化ストレージおよび拡張ストレージ内のメッセージファイル群を、本システムの FUSE 領域へそのまま複製し、逐次パースすることによって、メッセージファイル内に記載されたデータ項目を検索可能となる構造とした。

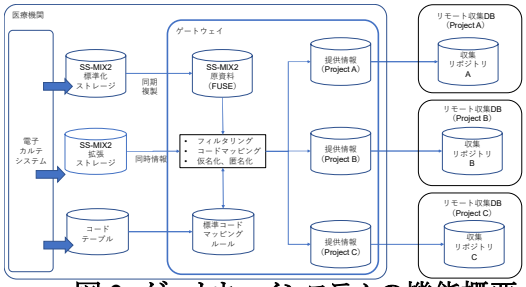

図 2 ゲートウェイシステムの機能概要

3.2 フィルタリング

臨床研究の多くは、受信日や病名、検査の実施有無、特定薬剤の投薬有無により、観察対象とする患者を限定してデータ収集を行うため、これらの一次スクリーニングに対して標準化ストレージを患者同意に基づく除外データや、治験患者情報、特に配慮するべき患者等、研究プロジェクトごとのポリシーにより、包含ないしは除外される患者が規定される。本システムでは、電子カルテシステムから対象患者リストや除外患者リストを目的ごとに出力し、ゲートウェイの抽出処理プロセスから参照させることにしている。

図 3 は、本稿のゲートウェイシステムへの連携を想定した患者由来の同意文書の運用概念図である。一般的に、患者本人の署名を必要とする侵襲的検査や手術への同意文書作成のための文書作成システムには、印刷紙面に 1 次元ないしは 2 次元バーコードを印字する機能が装備されており、署名済み文書のスキャン時に患者や文書種別の識別に用いられている。臨床試験や研究参加への同意や撤回に同様の機構を応用することで、患者により署名され提出された研究への同意や同意撤回をスキャン時に識別し、研究への同意情報データベースとして登録しておく。本システムでは、登録された同意情報を、データ抽出時にチェックすることにより、研究事業単位でデータそのもののへのアクセスを制御することを可能とした。

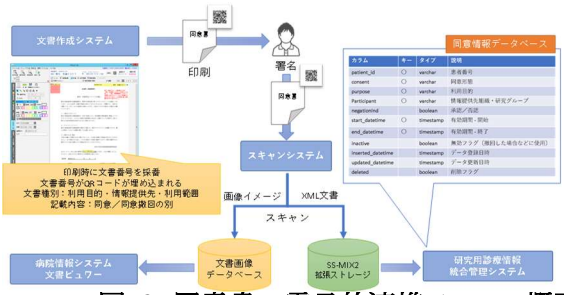

図 3 同意書の電子的連携フローの概要

3.3 コードマッピング

標準化ストレージは、各診療情報項目について標準コードで出力されることを規定しているが、現時点で実在するすべての電子カルテシステムにおいて完全に標準コードのみが適用されているわけではない。本院の事例からも、標準化ストレージに出力される一部のメッセージにはローカルコードが実際に出力されるケースが散見された。具体的には標準コードマスタに含まれない治験薬剤や、会計が特別に考慮されるべき検査、薬剤、医療材料などは、運用上の理由から特殊なローカルコードが採番されており、標準コードが採番されないケースが確認された。

臨床研究においては、病名、検査、薬剤などのデータ項目はコードによる検索が必須であり、本研究では、ローカルコードから標準コードへのマッピングに対する検証を行った。具体的な、標準コードへのマッピング方法は、図 4 に示す、いくつかの突合ステップを含むロジカルなコードマッピング手法により実施した。本院に存在するあらゆるマスタと医事コードマスタ、点数マスタを総合的に連結し、必要に応じて、医事会計実績と患者のオーダ歴、疾患情報を照合することにより、マッピングの統計的経験的自動化を目指すものである。現時点では、まだ開発の途中であるが、病名・医薬品・検体検査の3つのマスタにおける標準マスタへの突合成績として、有効な登録数・標準マスタへの突合不可数・マッピング可能率、は表 1 のような成績を得ており、さらなる自動マッピング成功率向上を目指している。

図 4 標準コードマッピングフローの概要

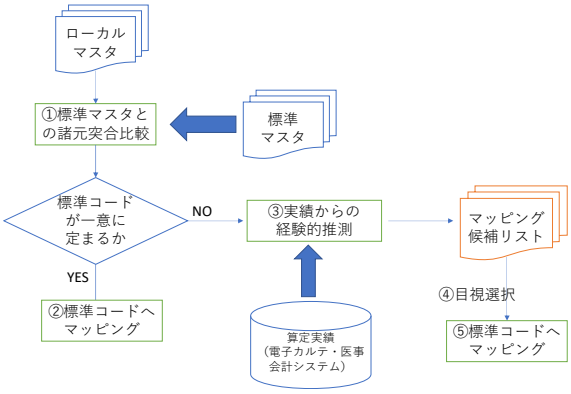

表 1 標準マスタへのマッピング率

マスタ種別	有効登録数	照合可能数	マッピング可能率
病名	30146	30124	99.9 %
医薬品	2354	2235	94.9 %
検体検査	1205	950	78.8 %

3.4 メッセージ変換機構

匿名加工は HL7v2 フィールド単位の操作とし、匿名加工ガイドラインに示されている、1)項目削除、2)一般化、3)トップ(ボトム)コーディング、に対応可能とした。ただし、ハッシュ値などの変換メソッドによって生成される文字列は、患者基本情報として、文字列規定長を越える場合が確認されており、HL7v2 規約に従って実装された受信側取り込みシステムでエラーとなる可能性がある。個々のデータ項目に関する加工方法への制約は、引き続き課題として検討する。

4. 考察

4.1 本システムの意義

本稿で提案、開発するゲートウェイシステムは、法制度による診療情報収集や、研究ベースのデータ収集プロジェクトなど、多目的なデータ収集事業に対応が可能となるように設計を配慮した。このため、本ゲートウェイシステムは、広範囲の医療機関において、多目的な適用が可能と

考えられる。とくに、同意情報や二次利用不可とした患者リストが機械的に標準化ストレージと連携され、診療情報の外部への情報提供において機械的にチェックされる仕組みは、本ゲートウェイの重要機能に位置付ける。

また、臨床研究における二次利用場面では、受診日、特定の検査の実施有無、検査結果値の範囲、特定の薬剤の投与実績、などを対象患者の抽出条件として指定する場合が多い。このような条件を指定して、対象患者を選別するための仕組みとして、本研究で実現するフィルタリング機能は、標準化ストレージがカバーする HL7v2 で定義されるフィールド情報を検索対象として利用可能となった。上述したコードマッピングと合わせて実施することにより、ローカルコードで出力された標準化ストレージ内のメッセージを、標準コードを用いて即座に検索をかけることが可能と考える。

4.2 コードマッピング

本稿で紹介したマッピング手法は、基本的に機械化、自動化を目指した方針である。マッピングの結果、ローカルコードと標準コードの対応表が電子カルテシステム外部で生成される。この対応表を機械的に取り込み、メッセージ変換ルールとして、本ゲートウェイシステムへ組み込むことにより、本院にみられるローカルコードで表記された標準化ストレージを、二次利用時には標準コードで検索できる状況を短期間に実現することが期待できる。検体検査マスタについては、他の2つのマスタより有効登録数が少ないものの、標準コードへの突合成績が低く、今後の課題とした。

4.3 維持管理コスト

中央集積型のデータ収集システムは、クラウドストレージなどの集積場所に対する維持継続にかかるコストを研究実施期間中確保せねばならず、研究そのものの継続性担保において大きな課題となる。一方で、本稿で提案するゲートウェイシステムは、分散配置された各ストレージに検索機能を有するため、集積場所としての中央設備を最小化し、オンデマンドに症例情報を収集することが可能である。今後、病名、検査値、薬剤などの代表的な検索条件に対応可能なよう検証を進める予定である。とくに、研究計画として、条件指定された患者数がどの程度当該医療機関に存在するか、という患者計数機能は研究の初期段階で非常に重要と考えられるため、早期に実現したい。

4.4 プライバシーリスク評価

中央集積型の事業では、収集後に適用条件による抽出、匿名加工、データセットの作成、を行う流れが妥当と考えられる。収集前にある程度の匿名加工を行うべきかどうかは、対象データの機微性に応じても変化するが、今回提案したゲートウェイシステムは、目的別のリポジトリを医療機関側へ配置する構成としているため、ある程度固定化された加工条件については、収集前のデータ加工が可能と考えられる。代表的な加工条件の選定が多岐に渡るため、今後の検証課題とした。

5. おわりに

本稿では、次世代医療基盤法で開発した大規模な診療情報収集プロジェクト向けのゲートウェイシステムを多目的の臨床研究事業にも応用することを目的として、データ収集用ゲートウェイシステムの追加機能開発について報告した。大規模な診療情報収集目的が公的あるいは閉鎖的である場合の両方に対応可能なゲートウェイとして基本要件を検討し、必要機能の提案を行った。

参考文献

1) Kimura M, Nakayasu K, Ohshima Y, Fujita N, Nakashima N, Jozaki H, et al. SS-MIX: a ministry project to promote standardized healthcare information exchange. Methods of information in medicine. 2011;50(2):131-9.
2) Tanaka K, Yamamoto R, Nakasho K, Miyaji A. Development of a Secure Cross-Institutional Data Collection System Based on Distributed Standardized EMR Storage. Stud Health Technol Inform. 2018;255:35-9.
3) International HLS. HL7 Standards Product Brief – HL7 CDA® R2 Implementation Guide: Privacy Consent Directives, Release 1 2017 [Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=280.