公募シンポジウム
# 公募シンポジウム3
# セキュアかつフェアなデータ流通プラットホームとセキュリティ基盤技術
2019年11月23日(土) 14:40 ～ 16:40 C会場 (国際会議場 2階国際会議室)

[3-C-2-05] セキュリティ要素技術による診療情報の安全な統合的利活用基盤の開発

○田中 勝弥[1]、山本 隆一[2] （1. 国立がん研究センター、2. 医療情報システム開発センター）

キーワード：Information Security, Privacy Protection, HL7

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年度より開始された 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てきた。医療分野に関するテーマとしては、医療情報システムが保有し、二次利用が期待される診療情報に対して、おもに組織外部への情報の電送や格納、臨床研究目的の収集および解析をターゲットとして、

・データの消失や流出といった、データの安全管理にかかわるセキュリティ対策に寄与する要素技術の開発、

・統計解析などの二次利活用における匿名加工に対して、プライバシーリスク評価可能な方法を提案し、実証的に示すこと、

に主眼を置いている。

本プロジェクトでの医療分野におけるパイロットシステム構築に関する取り組みとして、

・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

の機能開発を実施してきており、現在、諸機能を基盤システムとして統合する段階にある。

本発表では、医療分野向けテストベッドシステムの概要と機能詳細について報告する。

©一般社団法人 日本医療情報学会

セキュリティ要素技術による診療情報の安全な統合的利活用基盤の開発

田中　勝弥[*1]、山本　隆一[*2]、
*1 国立がん研究センター、*2 医療情報システム開発センター

Development of a secure integrated infrastructure for secondary use of clinical information based on security primitives

Katsuya Tanaka[*1] , Ryuichi Yamamoto[*2]
*1 Nacional Cancer Center Japan,
*2 Medical Information Development Center

Under the new law for secondary use of medical information, which activated on May 2018, the future expected secondary use with information anonymization may contribute to research and development in the medical field of integrated medical research and public health. On the other hand, under the revised personal information protection law and the revised ethical guidelines in medical research, privacy protection and patient consent management is a crucial issue for the management of researches. Our JST CREST project, started in March 2014, has issued on development of technological elements and synthesizing the developed methods for real world system for secondary use and privacy protection of big data on cloud infrastructure, including safe management of clinical information, utilization of commercial cloud and evaluation of privacy risk. In this paper, assuming the utilization of SS-MIX2 standardized storage, target issues such as 1) risk evaluation in analysis, 2) consent management, 3) traceability are described.

Keywords: Information Security, Privacy Protection, SS-MIX2, HL7

1. はじめに

これまでに複数の大規模な医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、次世代医療基盤法にみられるように、診療情報の施設横断的な集積による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 年度より開始された 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てきた。医療分野に関するテーマとしては、医療情報システムが保有し、二次利用が期待される診療情報に対して、おもに組織外部への情報の電送や格納、臨床研究目的の収集および解析をターゲットとして、

- データの消失や流出といった、データの安全管理にかかわるセキュリティ対策に寄与する技術要素を実装したパイロットシステムを示すことにより、
- 統計解析などの二次利活用における匿名加工に対して、プライバシーリスク評価可能な方法を提案し、実証的に示すこと、

に主眼を置いて、テストベッド開発を行ってきた。
本稿では、これまでに開発した各機能の概要と設計、検証の要点を総括する。

2. セキュリティ基盤技術を適用したテストベッド

診療データの情報伝送や外部ストレージへの格納に対しては、安全管理上の対策や、二次利用に対するプライバシー保護の課題があり、双方を満たす技術的対策と、実用可能性を実証する必要がある。いくつかのセキュリティ技術を実装した、実証システムによりその実用可能性の評価を進めてきた。個々のセキュリティ対策における基本的なセキュリティ関連ライブラリは、本研究プロジェクトの大阪大学宮地研究室、および KDDI 研究所清本らのチームにより開発されたソフトウェア群を必要に応じて拡張しながら適用している。

診療情報を扱うためのテストベッドの全体構想を図1に示す。本プロジェクトでは、診療データの安全な収集を実現するために、診療情報を蓄積させる SS-MIX 標準化ストレージの改良による PSI（Private Set Intersection）ライブラリの適用を行い、多施設間での安全な診療データの突合・収集が可能なプロトタイプを基盤として追加機能を開発してきている。

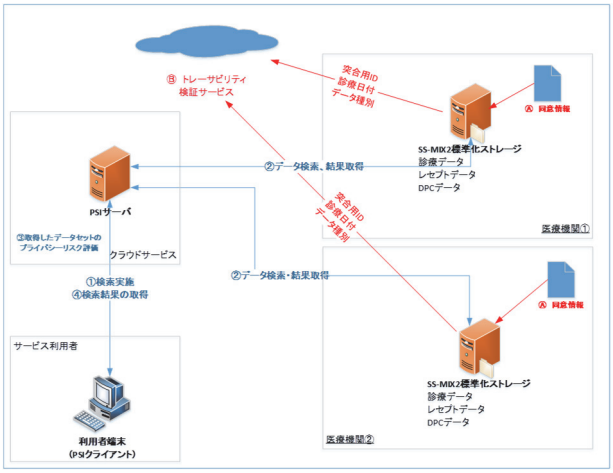

図１　テストベッドの全体概要

最終的には、診療データの二次利用に対して、同意情報によるアクセスコントロール、トレーサビリティ機能による二次利用状態確認、の構築までを目的として、

1) SS-MIX2 標準化ストレージ内診療データの検索性改善
2) PSI（Private Set Intersection）を利用した SS-MIX2 標準化ストレージからの安全な診療データ抽出機能
3) 同意情報の電子的記述と、データ抽出時の同意情報チェック機構
4) 抽出後データセットに対するプライバシーリスク評価機能
5) 患者による検証が可能なトレーサビリティ機能

の各要素機能の実装と検証が完了した。以下に個々の詳細を記述する。

2.1. SS-MIX2 標準化ストレージ実装

SS-MIX2 標準化ストレージは、HL7v2 メッセージを採用した患者番号、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機能できない。

本システムでは、SS-MIX2 標準化ストレージ上のデータを一般的な NFS 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している。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、現時点では、PostgreSQL と pgfuse の組み合わせにより実装を行っている [1]。構成の概要を図 2 に示す。

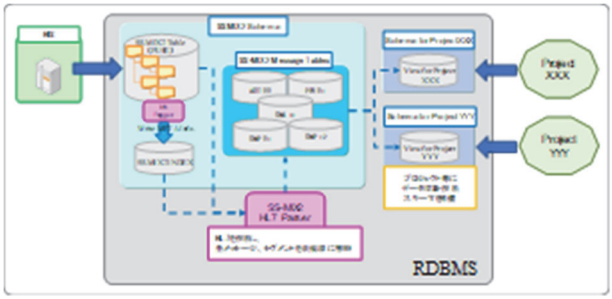

図 2 ストレージシステム構成の概要

2.3. PSI による施設横断的検索

宮地らの PSI[2]は、各施設に分散されたデータセット群を突合し、識別子を突合しながら、異なるデータセット間で必要項目を連結する、あるいはサブセットを抽出するなどの、集合演算が可能な Bloom Filter ベースの突合機能を提供する。本テストベッドでは、多施設間に配置された上述の SS-MIX2 標準化ストレージ上のデータを PSI に適合させる構成とした。

分散する標準化ストレージに対し、PSI ライブラリを適用し、Web サービスとして、ストレージ内のデータ検索、抽出が可能なクラウドサービスを開発した。サービスの動作フローを図 3 に示す。

2.4. 同意情報の電子的記述

現時点で医療者、患者の署名を必要とする同意書情報は、

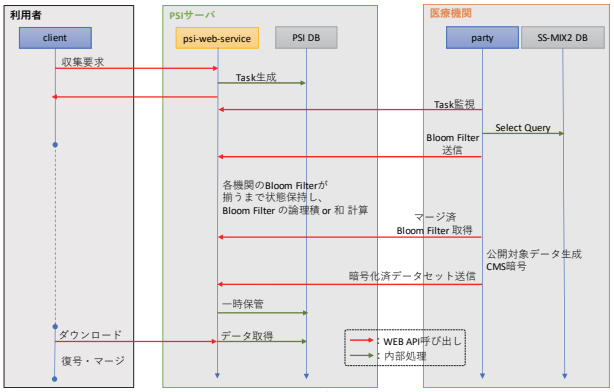

図 3 PSI サービスの動作フロー概要

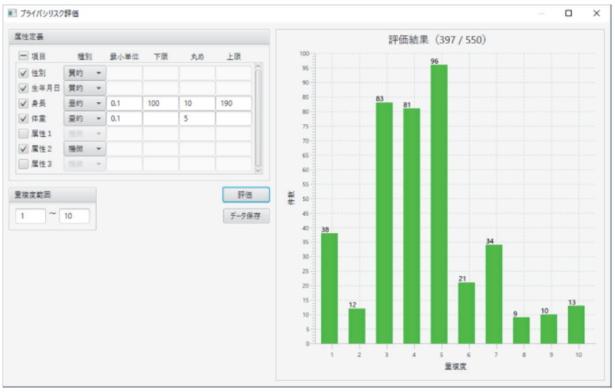

図4 プライバシーリスク評価ツールの外観

紙媒体として一次運用され、電子カルテシステム上ではスキャンされた電子ファイルとして管理される場合が多く、PSI サービスへの適用を考えた場合、標準的な文書記述規格により、SS-MIX2 ストレージなどに構造化された形式で格納され、PSI サービスから参照できる構成を提案してきた [3]。本システムでは、HL7 CDA R2 Implementation Guide: Privacy Consent Directives, Release 1 を電子的記述規格として採用し、スキャンされた同意あるいは同意撤回情報を SS-MIX2 の拡張ストレージに格納した状態で、PSI サービスと連携させることとした。CDA におけるヘッダ情報に相当する文書のメタ情報は、書面様式によりある程度規定が可能と考えられる。

2.4. プライバシーリスク評価

PSI サービスにより、SS-MIX2 ストレージから検索、抽出されたデータセットにより、一定程度匿名加工されたデータのセットを作成することが可能であるが、個人識別可能性の低減を考慮した場合、k-匿名化手法では、素データの精度を低減させ、解析結果に影響を及ぼす可能性がある。

また、収集、加工後に抽出されたデータセットは、個別提供ないしはオープンデータとして他へ流通される想定であり、プライバシー侵害が起きないように抽出後のデータセットに対する安全管理措置にも留意する必要があるが、それぞれの利用目的に応じてデータセットへの個人特定性を低減するための匿名加工方法は異なることが想定される。

本システムでは、PSI サービスにより抽出されたデータセット（匿名化サブセット）に対して、清本らが開発したプライバシーリスク評価ライブラリをクラウドサービスとして構築し、抽出したデータセットに対して、データ項目レベルの重複度（k-匿名化の k 値に相当する指標）をダイナミックに確認できるインタ

ーフェイスを試作、装備した [4]。開発したソフトウェアの外観を図 4 に示す。

2.5. トレーサビリティ機能

診療データの利用した多施設共同研究は多くの場合、各所属の倫理審査委員会の審査を経て、患者に対する公開または通知を経て実施される。ほとんどの場合、患者は自身の情報がどの研究に利用されているかを確認するためには、自らが受診した医療機関のホームページ等からの情報を得る以外の手段がなく、実際の利活用実態を確認するのは困難である。本システムでは、二次利用時のトレーサビリティを担保すべく、自身の診療データの流通実態を検証可能な機能を試作した。実装としては、Block Chain Network（BCN）への接続機能を、各 SS-MIX2 標準化ストレージに持たせ、データ抽出のログを集約し、BCN へ格納する方式とした。患者は BCN へ参加することで、本システム内のデータ抽出記録を検索、参照することが可能である。

なお、これまでの検証結果から、BCN 自身へのトランザクション性能は秒間 100 程度であることが分かっており、すべての抽出ログを BCN へ書き込むのではなく、ストレージごとのログを患者単位に集約したインデックス情報を書き込む方式を提案し、実装を進めている [5]。（図 5）

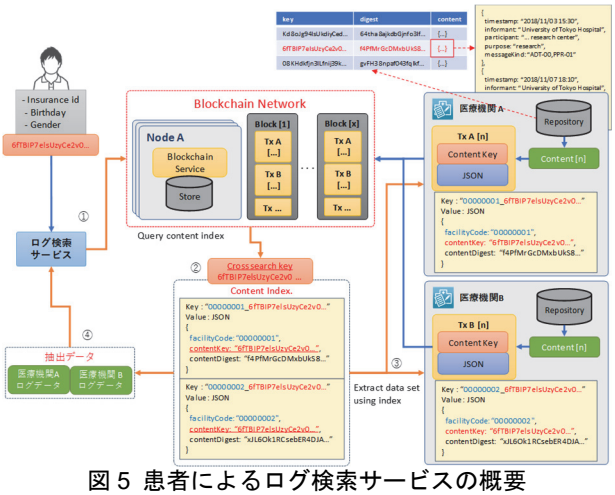

図 5 患者によるログ検索サービスの概要

3. 今後の展望

現在、上述の機能群を統合し、パブリッククラウドへ適用可能な Web サービスとして利用可能なことを念頭に、開発を進めている。また、ユーザビリティの向上のため、検索インターフェイスへの HL7 FHIR の適用を検討中である。（図 6）

4. おわりに

本稿では、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」における技術開発要素に対して、電子化された診療情報の利用場面におけるセキュリティ課題に対するこれまでの医療向けテストベッド構築に向けた各機能要素開発の取り組み、今後の展望について記載した。

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。

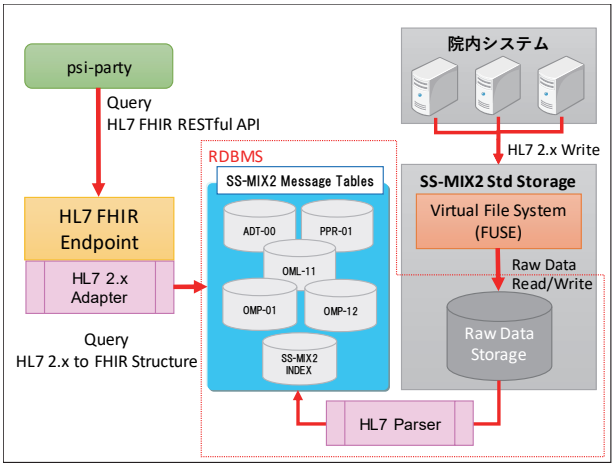

図 6 HL7 FHIR によるインターフェイス統合

参考文献

1) 田中勝弥, 山本隆一. 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. 医療情報学連合大会論文集. 2017;37 回:592-3.
2) Miyaji A, Nakasho K, Nishida S. Privacy-Preserving Integration of Medical Data. Journal of medical systems. 2017;41(3):37.
3) 田中勝弥, 山本隆一. 同意情報の電子的記述に関する検討. 医療情報学連合大会論文集. 2018;38 回:1082-4.
4) 田中勝弥, 山本隆一. 多施設間の安全なデータ収集を目的とした情報基盤におけるプライバシリスク評価を試み. 医療情報学連合大会論文集. 2018;38 回:360-2.
5) Tanaka K, Yamamoto R. Assessment of Traceability Implementation of a Cross-Institutional Secure Data Collection System Based on Distributed Standardized EMR Storage. Stud Health Technol Inform. 2019;264:1373-7.