

公募シンポジウム2

安全安心なデータ流通プラットフォームとセキュリティ基盤技術

2020年11月19日(木) 14:00 ~ 16:00 F会場 (イベントホール・特設会場1)

[2-F-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

*田中 勝弥¹、山本 隆一² (1. 国立がん研究センター, 2. 一般財団法人医療情報システム開発センター)

*Katsuya Tanaka¹, Ryuichi Yamamoto² (1. 国立がん研究センター, 2. 一般財団法人医療情報システム開発センター)

キーワード : Information Security, Privacy Protection, SS-MIX2, 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、本プロジェクトでは、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た。

また、多機関分散環境下におけるデータ収集においては、SS-MIX2標準化ストレージを対象として、FHIRにおけるユーザインターフェイスの標準化やコードマッピングの自動化などの手法も合わせて検討を継続している。

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、今後の展望について報告する。

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

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

Secure Integrated Infrastructure for Secondary Use of Clinical Information Based on Security Primitives

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

*1 National Cancer Center Japan, *2 Medical Information System 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, some primitives developed, current status of development applying the developed methods in National Cancer Center, and future issues 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(以下、CREST)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る中でいくつかの要素技術が開発された。

また、医療 AI の研究に必要なデータ収集は、リアルワールドデータを対象として、医療機関内あるいは医療機関横断的に集積され、利活用されることが基本形となる。国立がん研究センター(以下、NCC)でも官民研究開発投資拡大プログラム(Public/Private R&D Investment Strategic Expansion PrograM:「新薬創出を加速する人工知能の開発」)(以下、PRISM プロジェクト)を中心として、データ利活用基盤の検討が進められている。本稿では、これまでに開発した各機能の要点、および、コンセプトを応用した、NCC における取り組みについて紹介する。

2. 医療テストベッドの概要

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。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、CREST 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。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医療テストベッドとして開発した。

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
- といった要素機能の実装と検証を終えた。以下に個々の機能概要を記す。

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、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。本テストベッドでは、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た¹⁾。

2.2. PSI による施設横断的検索機能

宮地らの PSI 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している。本テストベッドでは、多施設間に配置された上述の SS-MIX2 標準化ストレージ内のデータベースに PSI を適用可能な構成とした。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、抽出が可能なサービスをこれまでに開発した。

2.3. 同意情報チェック機構

現時点で医療者、患者の署名を必要とする同意書情報は、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、組織横断的なデータ収集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 サービスから参照する構成を提案した。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ている²⁾。

2.4. プライバシーリスク評価機能

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。

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サービス化して構成し、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³⁾。

2.5. トレーサビリティ機能

診療データを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⁴⁾。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。

3. NCC PRISM プロジェクトの概要

3.1. 医療情報統合プラットフォーム構想

官民研究開発投資拡大プログラム (Public/Private R&D Investment Strategic Expansion Program: PRISM (プリズム) プロジェクト)「新薬創出を加速する人工知能の開発」では、国立がん研究センターの研究者や関連企業が参加して、肺がんの症例 1500 例以上を収集することを目標とした世界でも最大規模の肺がん症例データベースの構築を進めている。

一方で、NCC 中央病院は、2021 年 5 月に次期病院情報システムが稼働する予定にあり、「統合データベース」と称する部門システムを含めた横断的なデータ収集機能を構築する計画にある。AI 研究をはじめとするデータ利活用研究のために、キャンパス内にリアルワールドデータの提供基盤としての PRISM プロジェクトシステムとデータ連携し、研究者に二次利用可能な基盤を提供可能とする構想である。

3.2. 医療情報統合プラットフォームの要素技術

PRISM プロジェクトは、1 プロジェクトでありながら、NCC のキャンパス内データ利活用研究を担保すべく、実際の観察研究やデータ二次利用研究に適用し、複数の研究チーム、研究テーマに対するインフラ基盤として位置付ける想定にある。

前述の医療テストベッドで構築した機能群から、単施設向け機能としては、同意情報チェック機能、プライバシーリスク評価機能、トレーサビリティ機能、を現時点では実装のスコープとして含め、検討を続けている。

同意情報チェック機能は、オプトインあるいはオプトアウトで実施される臨床研究に対して、診療情報を二次利用として抽出、研究利用してよいかどうかを患者の同意情報に基づいて制御するための機能となる。基本的には書面を通して得られる IC 情報を、プロジェクト事務局としてスキャンし、本プラットフォームへ取り込む想定にある。

プライバシーリスク評価機能は、医療テストベッドと同様に、研究者へデータセットを提供する際に、データセット内の個人情報情報を除去し、個人特定性を判別しておく機能にあたる。AI システムへの利用など、他の個別システムへ抽出したデータセットが転送され、再利用されるため、取り出されたデータセットの機微性を判別し、記録する機能を持たせる。

3.3 今後の展望

医療情報統合プラットフォームの構築は 2019 年度後半から本格的な構築をスタートさせ、2020 年度中の試験稼働開始を目標としている。「診療用統合データベース」を有する次期病院システムの導入により、テキスト情報、医療文書、医用画像を含む多種多様な情報が本プラットフォームに連携される予定である。

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、開発要素の PRISM プロジェクトへの展開と、今後の展望について記載した。

なお、本研究の一部は、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) 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.
- 2) 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.
- 3) 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ーリスク評価の試み. 医療情報学連合大会論文集. 2018;38 回:360-2.
- 4) 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.