公募シンポジウム

公募シンポジウム4
医療分野の研究開発に資するための匿名加工医療情報に関する法律施行後の状況

2018年11月23日(金) 16:00 ～ 18:00 E会場 (5F 501)

[2-E-3-5] セキュリティ基盤技術による診療情報・同意情報の安全な統合的利活用

○田中 勝弥 （東京大学大学院医学系研究科）

2018年5月に医療分野の研究開発に資するための匿名加工医療情報に関する法律（次世代医療基盤法）が施行され、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促進されることが期待されている。一方で、改正個人情報保護法の施工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、診療情報の安全管理、分析に向けたクラウドの利活用、またそれに伴う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。本稿では、 SS-MIX2標準化ストレージの施設横断的な利活用を想定し、1) 二次利用シーンでの抽出対象データセットに対する匿名加工とリスク評価機能、2) 同意情報の管理を考慮したトレーサビリティ機能、などの医療情報分野への取り組みについて紹介する。

セキュリティ基盤技術による診療情報・同意情報の安全な統合的利活用

田中　勝弥[*1]、山本　隆一[*2]
*1 東京大学医学部附属病院、*2 医療情報システム開発センター

Safe integrated utilization of clinical information and consent information by security infrastructure technology

Katsuya Tanaka[*1] , Ryuichi Yamamoto[*2]
*1 Graduate School of Medicine, The University of Tokyo, *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 on 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。

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いる。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、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する。

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開発してきた。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、の構築を目的とする。

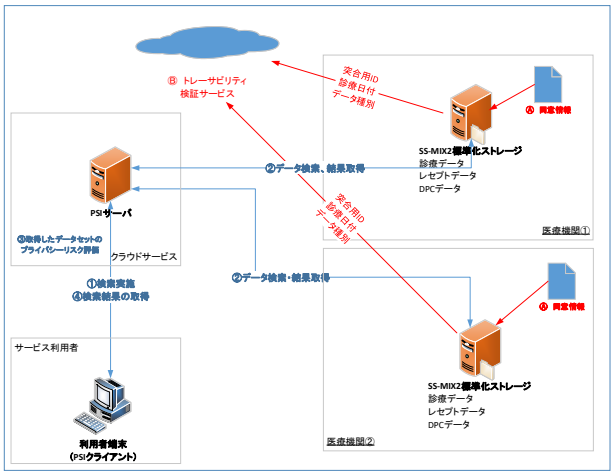

図1　医療用テストベッドの全体構想

本稿では、1)PSI（Private Set Intersection）を利用した SS-MIX2 からの安全な診療データ収集、2) 二次利用シーンでの解析対象データセットに対するプライバシーリスクアセスメント、3)同意情報の電子的記述、のおもなテーマの取り組み

について紹介し、今後の展開についても記述する。

2.1 PSI(Private Set Intersection)

宮地らの PSI[1]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、集合演算が可能な機能を提供する。ここでは、多施設間に配置された SS-MIX2 標準化ストレージ上のデータを PSI に適合させる試みについて記述する。SS-MIX2 標準化ストレージそのものを、仮想ファイルシステムの一つである FUSE（Filesystem in Userspace）と RDBMS（Relational Database Management System）上に実装することで、SS-MIX2 メッセージ内のセグメント、フィールドをデータ項目単位で検索可能とするプロトタイプをこれまでに試作した[2]。

本ストレージに対し、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2に示す。

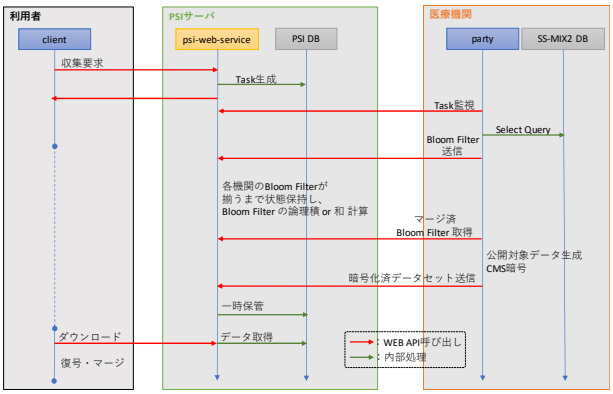

図2　PSI サービスの動作フロー

構築したプロトタイプを用いて、100 万件規模の性能評価を行った結果を表1に示す。レコード数とともに処理時間は単調に増加するが、100 万件規模で 1 分程度のオーダでデータ検索・突合・抽出が可能な性能をえた。

表 1 PSI サービスの処理性能測定結果

レコード数	20,000	40,000	80,000	160,000	320,000	640,000	960,000	1,280,000
検索	0.1	0.2	0.2	0.3	0.6	1.0	1.5	2.0
Bloom Filter処理	0.5	0.5	1.0	2.0	2.9	7.4	13.5	20.7
抽出	3.3	3.0	3.1	3.4	4.0	10.1	15.4	23.5
計	3.8	3.7	4.3	5.7	7.4	18.5	30.4	46.2

2.2 リスク評価機能付き匿名化

現在、多施設から SS-MIX2 標準化ストレージのデータをバックアップし、匿名加工したのちに、統計処理・データマイニングを行う基盤の開発が「SS-MIX2 を基礎とした大規模診療データ収集と利活用に関する研究」（代表者：山本隆一）において進められている（図4）。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優先した場合、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。

本研究では、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した。開発したソフトウェアの外観を図3に示す。

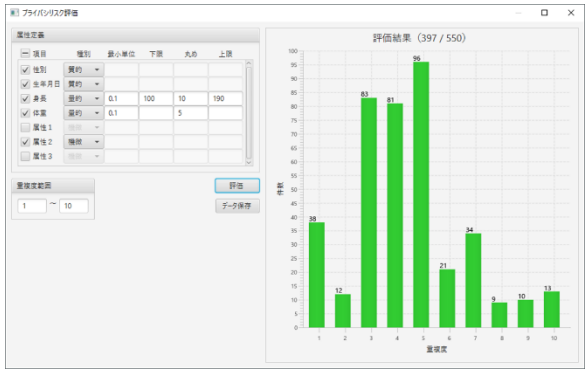

図3　プライバシーリスク評価ツールの外観

2.3 同意情報の電子的記述

侵襲を伴う検査や手術、治療法などのように明示的な同意がとられる場合、また疫学的研究や次世代医療基盤法のようにオプトアウトにより診療データの二次利用を行う場合、に大別される同意情報の取得形態があるが、いずれの場合にも、同意情報は、患者情報、情報提供目的、提供の範囲、情報提供先、同意の可否、を明示した記述が必要となる。これらの情報が明示されれば、上記の PSI サービスのアクセスコントロール源として参照し、患者が望まない不必要な情報の暴露を抑制できる。

一方で、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い。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が望ましい。

このような条件下において、HL7 CDAR R2 Implementation Guide: Privacy Consent Directives, Release 1[3]は同意文書の電子的記述規格として有力な候補である。HL7 CDA による XML 文書内にスキャンされた文書イメージデータを内包できる規格であり、上述の署名された同意文書のスキャン運用との親和性が高い。ただし、ヘッダ情報に相当する文書のメタ情報は、スキャン時にある程度マニュアルにより指定する必要があり、今後の検討課題となる。

2.4 今後の課題

PSI サービスのような1検索に対して無数のレコードがヒットし、第三者に流通される場合、これらの流通を確認するトレーサビリティ機能が必要となる。しかしながら、電子カルテのデータをワンショットで PHR へ複製するといった場合と違い、検索のたびに無数の診療データが特定の研究グループへ暴露する。台帳管理技術としては Block Chain が昨今注目を浴びているが、実現性も含めて検討を進めたい。

3. 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取り組み、現在の状況について記載した。

おもにセキュリティ技術としてのクラウドサービス、分散データの秘匿検索・集合演算、プライバシー保護を前提としたリスクアセスメントへの統合・運用を目指すものであり、本プロジェクトで開発したプロトタイプが、診療情報における収集、解析、利活用を促進するためのセキュアな情報基盤の創出、実用化を目指すものである。

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。また、患者データの利用にあたっては、東京大学医学部倫理審査委員会の承認（審査番号：11787、課題名：患者横断検索が可能な SS-MIX2 標準化ストレージ向けプラットホームの開発）をえて行った。

参考文献

1) Miyaji A, Nakasho K, Nishida S. Privacy-Preserving Integration of Medical Data. Journal of Medical Systems. 2017;41(3):37.
2) 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.
3) 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.