ポスター発表

[PB] ポスター B

2020年6月6日(土) 09:00 〜 16:30 ポスター会場(2) (e-poster)

[PB-28] 臨床情報データベースにおける直接的に識別可能な個人情報の匿名化
Anonymization Of Directly Identifiable Personal Information In Clinical Information Database

*中村 直毅[1]、佐藤 美喜子[2]、中山 雅晴[1,2] （1. 東北大学病院　メディカルITセンター、2. 東北大学大学院医学系研究科医学情報学分野）

*Naoki Nakamura[1], Mikiko Sato[2], Masaharu Nakayama[1,2] （1. Medical IT Center, Tohoku University Hospital, 2. Medical Informatics, Tohoku University Graduate School of Medicine）

臨床情報データベースにおける直接的に識別可能な個人情報の匿名化

中村　直毅[*1], 佐藤　美喜子[*2], 中山　雅晴[*1, *2]
[*1] 東北大学病院　メディカル IT センター,
[*2] 東北大学大学院医学系研究科医学情報学分野

Anonymization Of Directly Identifiable Personal Information In Clinical Information Database

Naoki Nakamura[*1] ,Mikiko Sato[*2], Masaharu Nakayama[*1, *2]
[*1] Medical IT Center, Tohoku University Hospital ,
[*2]Dept. of Medical Informatics, Tohoku University School of Medicine,

東北大学では、平成 29 年度から指定国立大学の第一陣の一つとして認定され、未来型医療の実現することを目標の一つに掲げ「ビッグデータメディシンセンター（BDMC: Big Data Medicine Center; BDMC）」を設立した。BDMC では、正確性・信頼性の高い臨床データと遺伝子・オミックスデータを連携させ、診療データの検索・分析を支援するための環境整備を進めている。著者らは、東北大学病院の電子カルテシステム、医事会計システム、部門システムなどが保持している診療データを収集する臨床情報データベースの整備を進めてきた。本稿では、この実名で構成されている臨床情報データベースを複製して、個人を直接的に識別可能な情報を匿名化して検索する仕組みを構築した。実際のシステムで個人が直接的に識別される可能性のある項目を匿名化し検索できることを確認した。複雑な匿名化の制御、匿名化の精度の評価、K 匿名性を考慮して同一属性のデータが K 件以上になる場合に、結果の出力を抑制する仕組みなどの確立は、今後の課題である。

キーワード　臨床情報データベース, DWH, 匿名化

1.　はじめに

東北大学では、平成 29 年度から指定国立大学の第一陣の一つとして認定され、未来型医療の実現することを目標の一つに掲げ「ビッグデータメディシンセンター（BDMC: Big Data Medicine Center; BDMC）」を設立した。BDMC では、正確性・信頼性の高い臨床データと遺伝子・オミックスデータを連携させ、基礎医学・疫学統計・情報処理・人工知能等の解析する臨床データの検索・分析を支援するための環境を研究者に提供する。

2.　目的

著者らが整備してきた病院情報システム、医事会計システムおよび部門システムが保持している臨床情報データベースシステムを基盤とし、同じインターフェースを介して、個人を直接的に識別可能な情報が匿名化されている臨床情報データベースを研究者に提供し、情報検索する環境を提供することを目的とする。

3.　方法

著者らがこれまで整備を進めてきた東北大学病院の電子カルテシステム、医事会計システム、部門システムなどが保持している臨床データを収集する臨床情報データベースを活用して、直接的に個人を識別できないデータベースシステムを構築した。次に実名のデータベースを複製し、直接的に個人を識別できる情報を匿名化もしくはデータの表示を抑制する仕組みを実現した。具体的には、データベース上にある個人を直接的に識別できる患者 ID はハッシュ化し、個人を直接的に識別できる他の項目は、非表示となるように構成した。

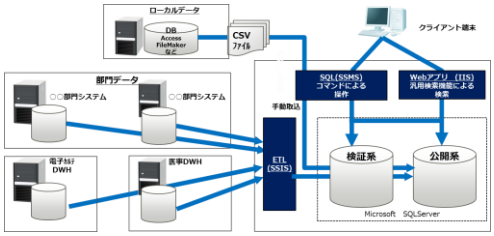

図 1 ：臨床情報データベースシステムの概要

4.　結果

最初に図 1 のように院内にある実名で臨床情報を検索・抽出をする統合的なデータベースシステムを構成した。検証系および公開系から構成され、データベースエンジンは、Microsoft 社の SQL サーバを利用している。データ収集の対象は、電子カルテシステム、医事会計システム、部門システム（救急システム、生理検査システム、病理システム、手術システム、オペラマスタ、SPD システム等）に加えて、診療科独自で Filemaker などで運用している診療科ローカルのデータベースと連携している。データ統合およびデータ変換は、SQL Server Integration Services(SSIS) を用いて、ETL（Extract、Transform、Load）処理によるデータ取り込み処理を行っている。診療科ローカルのデータベースについては、必要に応じて SQL Server

Management Studio を用いて手動で登録している。現時点で本データベースに取り込まれているデータは、表 1 の通りである。

表 1 ：データ収集対象のテーブル数

システム	電子カルテ	医事	手術	救急	病理	生理検査	オペラマスタ	SPD	ローカルデータ
テーブル数	128	90	42	12	6	7	1	1	28

本システムでは、Web アプリを通して、抽出する項目や条件を指定して情報検索する機能や予め設定した条件をテンプレートとして登録して検索する機能を備えている。また、職種や利用者ごとに表示されるメニューや検索可能な項目を制限し、診療情報の検索範囲も制限することも可能である。

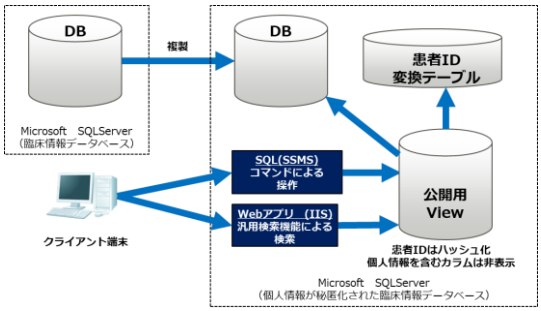

図 2:臨床情報データベースの匿名化の概要

個人を直接的に識別できてしまう情報が匿名化されている臨床情報データベースを実現するにあたり、システムの改修を最小限に抑えるため、臨床情報データベースを直接拡張するのではなく、データベースを複製し、現行のアプリケーションのユーザインターフェースは維持したまま、表示・非表示を柔軟に制御するシステムとして構成した。

具体的には、複製したデータベースには、個人を直接的に識別できる患者 ID は、そのまま表示するのではなく、患者 ID を一定のルールで非可逆のハッシュ化するテーブルを用意し、ハッシュ化された ID を利用するようにした。また、患者 ID 以外の個人を直接的に識別可能な項目に対しては、データベース上に Web アプリおよび SQL クライアントによる表示や検索の可否を制御する公開用の View を新たに用意して、この View の制御に従って表示制御するようにした。

実際のシステムでは、個人を直接的に識別可能な項目である、氏名、郵便番号住所、電話番号、保険情報を非表示にするよう公開用の View を構成した。個人を直接的に識別可能な情報の表示が抑制された形で検索する際の臨床情報データベースの画面およびその結果を図 3、図 4、図 5 に示す。これらの図が示すように患者 ID をハッシュ化し、個人を直接的に識別される項目を非表示にすることができ、個人を直接的に識別可能な情報の表示の柔軟な制御可能であることを確認した。

データベースの複製に際しては、システムを単純にするため、複製元と複製先のデータベースをリアルタイム同期するのではなく、複製元でデータのエクスポート、複製先でデータをインポートする構成とした。システムのインポート中は運用停止になることが懸念されたため、正副のデータベースを用意し、インポート完了後、参照するデータベースを切り替えることにより、インポートの際に停止時間を最小限に抑えるように工夫した。

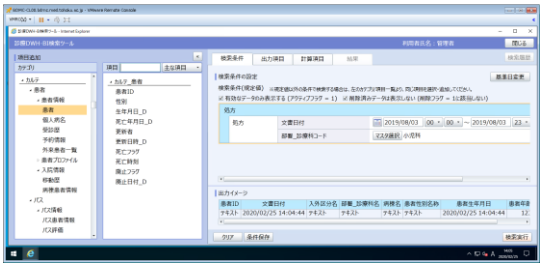

図 3:検索項目指定の画面
（個人を直接的に識別が可能な項目の表示の抑制）

図 4:検索条件の画面

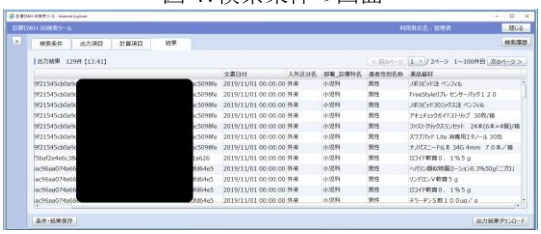

図 5:検索結果の画面サンプル
（患者 ID のハッシュ化）

5. 考察

電子カルテ、医事システムに留まらず院内の部門システムおよび診療科で独自に管理しているデータベースを統合的に検索できる環境整備を進め、2020 年 1 月からは、第 9 次病院情報システムの統合データベースシステムとして運用を開始した所である。従来の仕組み[1]においては、別システムとして個人を識別できる項目を匿名化にするように構成されている。一方、本システムでは、実名のデータベースを複製し、複製元のデータベースと同じ操作で利用できるようにし、個人の特定に直接的に繋がる項目の表示を公開用の Viewを用いて柔軟に制御することができ、連携システム対象が増えた際にも対応が可能である。

現時点では、項目の表示・非表示に限定した匿名化に留まっており、個人が間接的に識別できてしまう懸念が残っている。今後、より複雑な匿名化の制御、匿名化の精度の評価、K 匿名性を考慮して同一属性のデータが K 件以上になる場合に、結果の出力を抑制する仕組みなどの確立が課題である。

6. おわりに

個人を直接的に識別可能となる項目を匿名化する臨床情報のデータベースについて述べた。今後は、本システムを提供しながら機能を見直し、未来型医療の実現の貢献に寄与したい。

参考文献

畠山ら, RYOMA：診療データに基づく匿名化解析用データベース構築, 医療情報学 28(Suppl.), 273-275, 2008