

厚生労働行政推進調査事業費補助金

地域医療基盤開発推進研究事業

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究

(H27-医療-指定-016)

平成 2 8 年度

総括・分担研究報告書

研究代表者 木村通男 (浜松医科大学)

平成 2 9 年 (2 0 1 7 年) 3 月

目 次

・ 総括研究報告	
電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究……………	1
研究代表者 木村 通男	
・ 分担研究報告	
1. 研究ノートの電子化 および 臨床データバリデーションに関する検討……………	14
中島 直樹	
2. 研究用計測機器ユーザー認証等証拠性保全検討……………	32
澤 智博	
3. 測定、画像データ証拠性・安全運用環境検討……………	36
作佐部 太也	
4. 研究用計測機器ユーザー認証等証拠性保全検討……………	42
渡辺 浩	
・ 研究成果の刊行に関する一覧表……………	45

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究
(H27-医療-指定-016)

研究代表者 木村通男

浜松医科大学医学部附属病院医療情報部 教授

研究要旨: 目的: 医学研究における各種データを情報システムや電子カルテの技術・運用をあてはめることにより、誰により、いつ、どこで作成され、改定されたものかを明確にすることを自ら証せるものとするのが本研究の目的である。方法: 本年度は、a.) 測定データ、画像データのワークフローの検討 b.) 研究ノートの電子化のため、小規模施設用電子カルテの試用 c.) 臨床系データのヴァリデーション状態の把握のため、種類別に監査の有無の検討 d.) 研究用計測機器データへの署名の付加の実現性 について検討を行った。結果: a.) 測定データの証拠性向上: 次世代シーケンサー (以下、NGS という。) から出力されるデータの処理や解析を目的としたワークフロー管理システムについて、そのセキュリティ機能について調査した。複数名での利用を想定したワークフロー管理システムにおいては、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されていた。これらのセキュリティ管理機能は、研究データの証拠性保全を支援できる可能性があることが判明した。 b.) 研究ノートの電子化: 証拠性を有する研究を進めるために、電子研究ノートの実証実験を実施し、アンケート調査により一定の有用性を確認した。 c.) 臨床系データのヴァリデーション: 近年進んできた SS-MIX2 からの EDC は、導入から継続的なデータ品質管理を実施している限りにおいて、他よりも証拠性が強いことを示唆した。 d.) 研究用計測機器のユーザ認証など、証拠性保全: ハッシュ値の算出にかかる計算負荷は過大なものではないことがわかった。一方、改ざん防止システムの構築の必要性が高いことが明らかになるとともに、実現のためには多くの問題があることも明らかとなった。したがって、改ざん防止のシステムを構築するためには、研究者や学会などの学術組織だけでなく企業などとも連携した社会的な取り組みが必要であることがわかった。結論: a.) NGS から出力されるデータの処理や解析を目的とした複数名での利用を想定したワークフロー管理システムにおいては、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されており研究データの証拠性保全を支援できる可能性がある。 b.) 電子研究ノートの実証実験を実施し、アンケート調査により一定の有用性を確認した。 c.) SS-MIX2 からの EDC は、導入から継続的なデータ品質管理を実施している限りにおいて他よりも証拠性が強いことを示唆した。 d.) 実用的時間で電子署名を NGS データに処することが可能であることが判明した。NGS データの証拠性向上のため、特に改竄の検知を可能とするためには、NGS や公共データベースに対して電子署名技術と習慣を導入する必要があることが明らかになった。

A.研究目的

昨今、内外で、医学研究に対しての社会からの信頼が損ねられ事例が生じている。その多くは紙運用の時代のものであるが、今の研究用計測機器においても使用者が認証され、そのログが残されている施設は少なく、高価な機器でもそういう機能のないものを散見する。

しかしながら、科学研究以外の分野では、内外で、商取引、登記、資料保全などで、電子的な手法を用いる際の規範が定められている([1] [2] [3])。臨床応用に向けて、前のめりになりがちな医学研究では、この点が、ともすれば蔑ろにされてきた感があり現在の状況を惹起したと考えられる。このため、医学研究に携わる者が、その研究の道筋を証拠性をもって自ら証するための技術的ソリューション、運用的規範が、まさに今求められている。

本研究の目的は、医学研究における各種データを情報システム、電子カルテの技術・運用をあてはめることにより、誰により、いつ、どこで作成され、改定されたものかを明確にすることを自ら証せるものとする事である。関係する研究機器、記録系、外部サービスなどを試作を超え、実際に作成することが目的ではなく、研究指針、研究評価要件などのための技術的・運用的基準を定めることを目的とする。

[1] 民間事業者等が行う書面の保存等における情報通信の技術の利用に関する法律(平成 16 年法律第 149 号)

[2] 商業登記法(昭和 38 年法律第 125 号)

第 17 条第 4 項に規定する電磁的記録への記録の方式等

[3] CFR - Code of Federal Regulations
Title 21

B.研究方法

本研究の対象要素は、以下のとおりである：

- a. 測定データ、画像データの証拠性向上
- b. 研究ノートの電子化
- c. 臨床系データのヴァリデーション
- d. 研究用計測機器のユーザ認証など、証拠性保全
- e. ユーザ認証、外部保存などの外部サービスの利用
- f. 安全に運用するためのセキュアな環境

この内、本年度は、以下の 4 項目についての検討を行った。

- a. 測定データ、画像データの証拠性向上
複数のワークフロー管理システムについて、セキュリティ全般、ユーザー管理、アクセス権限管理、ログ管理に関するドキュメント類を精査した。

- b. 研究ノートの電子化
b-1 実験に用いるアプリケーションの選定
既に広く市販・活用されている Web 型電子カルテから、研究目的に合致しているシステム要件を持つ製品を選択した。

まず、市販の電子カルテの患者 ID を研究者個別の研究課題、研究者を記載医師、研究グループを診療科、と読み替える。
求められるシステム要件は

- ・ 複数診療科が設定可能、その属性によ

り患者データの参照権限が設定可能（他研究グループが別課題のデータを参照できないように）

- ・ Web 型のブラウザを利用したものであること（研究者の利便や費用を考慮して）
- ・ 会計機能、オーダ機能は必要なく、電子カルテ管理で最も手間がかかる各種項目（医事項目、検査項目など）のマスター管理は必要ない、などである。

MagicConnect®（NTT テクノクロス社）と呼ばれるアプリケーションでリモートアクセスコントロールを行った。これは、同社の保有するマジックコネクトサーバを介して https で暗号化通信を確立するものである。

b-2 実験（研究ノートとしての実際の試用）と使用後のアンケート調査

研究指導者 2 名（教員クラス）および九州大学医学研究院博士課程大学院生 2 名に対して研究の協力を依頼し、浜松医大病院に設置されたサーバに対する実際のアクセスなどを用いた研修会を九州大学病院内で実施した。

平成 28 年 3 月から 2 つの研究課題において研究ノートとして使用し、使用後にアンケート調査を行った。アンケートは研究者用と指導者用の 2 種類を作成した。

c. 臨床系データのヴァリデーション

c-1 EDC データ源としての SS-MIX2 標準化ストレージのデータの品質の検討

厚生労働省 / PMDA が行っている MID-NET 事業や各種の疾患コホート研究においては、EDC のデータ源として SS-MIX2 標準化ストレージを活用してい

る。それをデータ源とする場合、HIS から SS-MIX2 を介して各事業 DB への正確な移行が重要であるが、MID-NET ヴァリデーション事業では、PMDA と共同でこれを検証した。HIS データと事業 DB 間のデータ一致率を病名、処方、検査結果の 3 つ別々に調査し、その結果をフィードバックしてデータ品質を向上した。

c-2 考えられるデータ源別からの EDC の信頼性検討

HIS 周囲には各種の部門システムがあり相互に連携している。これらから、データ源の真正性や監査証跡の有無、およびデータ源との差分の比較の容易性を基にデータ源およびデータ回収法別（EDC・手入力別）にデータ信頼性を比較した。

d. 研究用計測機器データの証拠性保全

d-1 ハッシュ値の計算実験

データの改竄検出のための基礎となる情報処理はハッシュ値などの要約データの生成であり、どのような高度な改竄防止の仕組みを構築するにしても、ハッシュ値の生成と関連付けについての実現性があることは前提となる。そこで、NGS データについてのハッシュ値の生成を、NGS 関連で運用されるサーバと同程度の能力の計算機で実際に行いその時間を計測した。また、生成したハッシュ値と NGS データとの関連付けの方法として最も単純で信頼性の高い埋込みについて、実際にハッシュ値を NGS データに埋込む実験を行い、NGS データを解析するシステムに影響を及ぼさない方式を探索的に求めた。

d-2 NGS 関連環境の調査

ハッシュ値の生成・埋込みを行うのに、データ改竄の防止に適したタイミングを検討するために、実際の NGS の運用現場や公共データベース、研究者が利用するソフトウェア等についてインターネットや文献などによる資料に基づいて調査を行った。小規模ラボにも負担が少ない範囲で、どういう機能、運用、外部サービスが望まれるかについて明らかにした。

倫理的配慮

平成 28 年度の研究においては、特に個人情報を取扱うなどの倫理的な課題は発生しなかった。

C. 研究結果

a. 測定データ、画像データの証拠性向上

・ワークフロー管理システムのセキュリティについて

ワークフロー管理システムのセキュリティについては、そのソフトウェアの開発コンセプトと想定されるユースケースによって大きく異なる。個人での利用を想定しデスクトップアプリケーションとして開発されたワークフロー管理システムは、機動性や個人としてのユーザビリティに力点が置かれ作成中のワークフローに対する権限管理の機能等が存在していない。一方で、サーバーあるいはクラウドを前提したワークフロー管理システムでは、複数人による利用を前提としており、ユーザー認証、アクセス権限管理、システム使用ログ管理などの機能を備えている。

調査対象のワークフロー管理システム

においては、galaxy、GenePattern、KNIME においてセキュリティ全般を管理する機能が実装されていることを確認した。

・galaxy のセキュリティ管理機能

galaxy においては、security package が実装されドキュメントが公開されている。代表的な機能としては、データ・ライブラリ・フォルダー等に対する権限管理、ロールの設定、表示・非表示、等がある。また、サブモジュールとして、パスワード管理、バリデーションチェック機能が存在する。

ログの取得について異なる詳細レベルで可能となっている。また、ユーザーの使用状況についてレポートを生成する機能が実装されている。

・GenePattern のセキュリティ管理機能

GenePattern は、ウェブアプリケーションサーバーを基本としており、そのためアクセス管理に必要な機能が充実している。

主な機能として、アクセスフィルタリング、パスワード設定、ユーザー管理、ユーザー権限管理、SSL 対応、が実装されている。

ユーザー管理機能に関して、グループの権限設定が可能になっている。

ログ管理機能については、インストールログ、GenePattern ログ、ウェブサーバーログの三種類のログについて管理が可能な仕様となっている。

・KNIME のセキュリティ管理機能

KNIME には複数のプロダクトラインが

存在するが、KNIME Server は複数人による利用を想定しており、そのユースケースに応じたセキュリティ管理機能を備えている。

共有レポジトリにおいては、データ、ノード・メタノード、ワークフローにおいてアクセス権限を設定している。また、これらについて版管理機能を備えている。

ログ管理機能については、中間生成物の管理、トラッキングや監査機能を備えている。

b. 研究ノートの電子化

b-1 実験に用いるアプリケーションの選定

検討の結果、本研究では、NEC 社製の小病院用電子カルテシステム MegaOak SR for SaaS が、Web ブラウザをベースにしており、プラットフォームに特別なものを求めない点を重視し採用した。これは、外部サーバを利用した形式と施設内にサーバを置く形式が可能であるが、今回は、試用以後も使えるように固定費用の発生しない形式、つまり、浜松医大病院にサーバを置き、九州大学病院の研究グループからは電子証明書を用いた方法でのアクセスコントロールをおこなうこととした。浜松医大病院、九州大学病院の2施設でおこなったことから、施設内、外部サーバ利用をそれぞれ実現したこととなる。

b-2 実験(研究ノートとしての実際の試用)と使用後のアンケート調査

浜松医大病院内にサーバが設置された NEC 社製の小病院用電子カルテシステム MegaOak SR for SaaS に対して、安全性

を電子証明書で担保し、その範囲内で九州大学病院内、あるいは指導者や研究者の自宅などからアクセスできた。

研究者毎に異なる複数の研究課題について、研究ノートを設定し(患者 ID 機能を利用)、電子カルテの権限参照管理システムを用いて容易にアクセス権限管理を行うことが出来た。

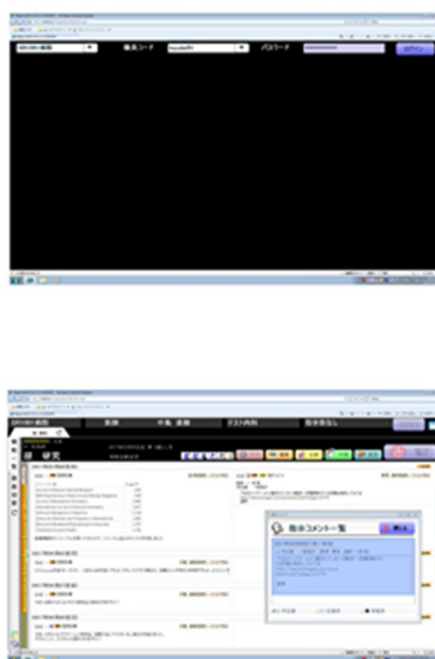


図1. 電子研究ノート画面。上)ログイン画面、下)電子研究ノート画面。電子カルテの患者 ID を研究者個別の研究課題、研究者を記載医師、研究グループを診療科、と読み替えて使用。

アンケートは研究者2名、指導者2名の少数の結果であったが、以下の結果を得た。

生物実験や化学実験などの Wet Labo と違い、統計学を含む情報系研究者は Dry

Labo と呼ばれる。今回の対象者は、Dry Labo の研究者と指導者であった。電子的に研究の半分以上の記録を記載していたのが 1 名で残り 1 名は記載していないと答えた。同時に 2 名の指導者も紙の研究ノートを確認はしていなかった。

電子研究ノートに参加して、4 名全員が記載の時間が増えたが、これまで記載や確認をしていなかった 3 名については新しい作業が発生したための当然の結果であるが、電子的に記載していたものにとっても、ログインなどの作業に手間がかかり、時間が増えたものと考えられる。なお、Dry Labo ではコンピューターが主たる研究業務の場であることが多く、計算式などの履歴をコンピューター上に残しながら仕事をする。そのため、Wet Labo に比較して Dry Labo では紙の研究ノートは作っていないことが多いと推定される。

研究者と指導者との間で、電子研究ノートの機能の充足の意見（複数回答可）は分かれた。指導者はともに機能は不足していると答えたが、リアルタイムのやり取りが困難、電子データファイルの管理（変更履歴や変更内容の管理）が難しい、などの意見であった。

紙の研究ノートと電子研究ノートの使用感の比較（複数回答可）では指導者も研究者も同じ傾向であった。両者を足し合わせると、紙ノートの利点は、自由度が高い（3 名）、使いやすい（2 名）であり、欠点は、紛失する危険がある（4 名）指導を得にくい（しづらい）（3 名）が多かった。一方で、電子研究ノートの利点は、記録の検索性が高い（3 名）が多く、課題は、画面がわかりにくい、操

作が複雑、ログインに手間がかかる、が多かった（各 3 名）。

改ざんの有無がわかる証跡が残っているかどうかについて指導者 2 名に聞いたところ、ともに証跡が残っている、と答えたが、これは電子カルテ機能を活用したためでもあろう。

その他の自由記載において、指導者からは「情報系（データ解析を含む）の研究では、多くの作業がコンピューター内で完結し、手作業の部分を除いて、ソースコードに作業内容が記録されるので、実験の再現も容易である。実験ノートという習慣がほとんどない一方で、電子データの管理が本質的に重要である。したがって、そこを証跡を含めて簡単に管理できる仕組みが必要だと考えられる。少なくともデータ解析の指導においては、電子カルテというよりも、GitHub のようなソフトウェア開発の仕組みを使った方が適していると思う。」という意見を得た。また、研究者からは、「デフォルトの枠が狭く、一行の文章を全文表示できない。マウス操作で枠を拡大できる。また、下までスクロールするスピードが遅く、文章が長いとき読みづらい。」という電子カルテ機能の制限に関する意見を得た。

c. 臨床系データのヴァリデーション

c-1 EDC データ源としての SS-MIX2 標準化ストレージのデータの品質の検討

HIS と事業 DB 間のデータ一致率は、データ品質管理作業前には、傷病名、処方、検査結果情報で、99.1、67.0、55.8%であった。データ不一致の原因は、連携バグ、マッピング不備、多様なオーダ種や

結果値の存在など様々であった。データ品質作業による是正の結果、99.9、100、100%となった。データ不一致の原因の一部は HIS と SS-MIX2 との連携が原因であったが、その原因も含めてほぼ全て見出し、解消することが可能であった。

C-2 考えられるデータ源別からの EDC の信頼性検討

平成 27 年度の本分担研究の結果を受けて、各種のデータ源およびデータ回収法別に、想定されるデータの信頼性に影響を与えるデータ源別の因子を表 1 に示す。

	真正性	監査証跡（移行）	HISとの差分の把握
HIS	○	○	—
SS-MIX2標準化ストレージ	×	○	○
SS-MIX2拡張ストレージ	×	○	△
各種部門システム	×～○	×～○	×～○
SS-MIX2以外の外部データ源	×	×～○	×～○

表 1 . データの信頼性に影響を与えるデータ源別の因子

病院情報システム(以下、HIS という。)は、真正性を確保していることが導入の条件である一方、各部門システムでは、HIS にデータ連携すれば HIS 側で真正性が確保されるためにそれ自体は必ずしも真正性を確保しているとは限らず、各病院の運用に依存する。また、HIS、部門システムを問わず業務用データベースであるため、そのデータ 2 次利用の場合は連携する外部集積データソースを利用する

ことが多い。したがって、外部集積データソースの真正性（特に正確性）が確保できていなければ、適切な解析結果を得ることはできない。

SS-MIX2 標準化ストレージは HIS の傷病名、処方注射、検査結果データなどが移行されているが、SS-MIX では HIS における監査証跡が移行され、さらに MID-NET 関連事業で確立された手法、開発されたツールを用いた HIS データとの差分チェックの応用が可能となりつつある。現在、SS-MIX2 標準化ストレージの導入が全国的にも増加し、これを特に MID-NET 事業や各種疾患コホート研究などで EDC の外部集積データソースとして用いることが急激に増えている。一方で、標準化ストレージに入っていないデータ項目の HIS から SSMIX2 拡張ストレージへの移行データに対する EDC では、HIS での真正性の確保は標準化ストレージと同様だが、監査証跡の移行は運用により異なり、MID-NET 関連事業での差分チェックの経験がない。また、これはデータ項目種毎に差分のチェックの方法と課題の解決を経ねば向上しない。一方、部門システムは、前述のようにその用途によって真正性の確保は様々であり、改竄や捏造が入り込む可能性がある。さらに、SS-MIX2 以外の外付データ源（ベンダーの DWH など）では真正性の確保や監査証跡の移行などは行われず、差分チェックも様々でありデータ品質は低下する。

以上の結果、各データ源からの EDC および手入力されたデータの信頼性は、

- 1.) HIS と連携した SS-MIX2 標準化ストレージからの EDC によるデータ回収

- 2.) HIS と連携した SS-MIX2 拡張ストレージからの EDC によるデータ回収
- 3.) HIS や真正性を確保した部門システムから連携した SS-MIX2 以外の外付データ源 (DWH など) からの EDC によるデータ回収
- 4.) 真正性を持たない部門システムからの SS-MIX2 以外 の外付データ源への出力
- 5.) 臨床研究用 DB へのデータ手入力の順にデータの信頼性は高いと想定された。 研究分担者、中島の担当であり、詳細は、その報告書に記載されている。
臨床研究で使われる項目を整理したところ、患者プロフィール情報、診療録情報：サマリー情報、連携情報、オーダ情報、医事会計情報、2 次利用目的登録情報に大別された。これらを調査したところ、人的な操作が入る可能性があるのは、
 - 1.) HIS あるいは、SS-MIX2 標準化ストレージを含めて、全ての医療情報システムから出力された後の 2 次利用用のデータ加工時
 - 2.) HIS への入力を行わないデータを部門システムから直接 EDC、あるいは、中間データストレージ (SS-MIX2 拡張ストレージ、DWH 含む) から EDC する際に、部門システム側に真正性が担保されていない場合に部門システム側、あるいは、DWH 上で操作 (SS-MIX2 拡張ストレージ上での操作は困難)
 - 3.) HIS への入力を行わないデータを手入力で臨床研究データベースへ入力する場合であった。
また、利用者認証や、真正性の確保は、

HIS の仕組みを使うことによってほぼ解決できることがわかった。その一方で、外部出力の仕組みの実装には、ベンダーの経験不足などにより、齟齬が生じやすく、バリデーションが必要であることが明らかになった。

d. 研究用計測機器データの証拠性保全

d-1 ハッシュ値の計算実験

ハッシュ値の生成にかかる時間は問題の無いものであった。また、ハッシュ値の NGS データの埋込みについても、解析システムに影響を与えない方式の実現性が確認できた。これらについておよび詳細および追加実験の結果は、分担研究者、作佐部の報告書に報告されている。

d-2 生成・埋込みのタイミング

ハッシュ値の生成・埋込みの処理を行うことができるタイミングについての検討、調査を行った。NGS 自身にそのような機能を付与することは理想的であるが、NGS の製造メーカーへの問い合わせなどは行っていない。

NGS を運用する施設において処理を実施する場合について検討した。

例えば 分担研究者の所属する機関においては、中間報告書において報告したが、生物学系研究者が独自に解析システムを構築し運用している。解析システムやシステムを構築した研究者の情報処理技術についてのスキルからみて、プログラムを提供すれば、NGS の出力直後にハッシュ値の生成・埋込みを行うプログラムを解析システムに組込むことは可能であると考えられる。より大規模な研究機

関であれば、情報処理の専門技術者や研究者も所属している場合もあり、プログラムの開発を含めて可能であろう。

NGS は運用しているが自前の解析システムを持たない機関においては、NGS データは解析業者に送付になるので、計算機操作についての高度なスキルを持たない研究者でも操作できるよう、操作が容易なプログラムを用意する必要がある。

NGS を持たず試料を委託先に送付して NGS の処理を外部に委託して行う場合には、受託業者が行うことになるが、問題は無いと考えられる。

何れの場合においても、ハッシュ値の生成および埋込みの方法について正確に記述された仕様を策定し配布することは必須である。可能であれば、実際に動作するプログラムを配布することが望ましく、特にオープンソースであればより望ましいと考えられる。

d-3 公共データベース

今日の遺伝子などの分子生物学に関わる研究では、研究成果のジャーナルへの投稿において、利用した NGS データを公共データベースへの登録公開が義務付けられることがある。この場合は、公共データベースへの登録前にハッシュ値は埋込まなければならない、公共データベースから取得するデータにも際にも保持されていないなければならない。

公共データベースとして、国際的に協調、集約の動きがあり、現在その中心となっているのが SRA である。SRA では独自のフォーマットで保存されている。必要に応じて解析システムに入力できる形

式に変換される。SRA をダウンロードして変換することもできるし、専用のツールにより目的のフォーマットを取得することもできる。従って、SRA 形式におけるハッシュ値の保持形態を検討する必要がある。

d-4 NGS シミュレータ

NGS の普及とともに、NGS をシミュレーションするシステム(以下 NGS シミュレータ)が開発されている。NGS シミュレータは仮想的な NGS データを生成するソフトウェアである。NGS シミュレータは、主に NGS データを解析するシステムの検証、評価に用いられる。NGS シミュレータについての包括的な調査がある。また、NGS の開発メーカーである Illumina 社が開発した高度なシステムである EAGLE についての調査を添付 3 にまとめた。

留意すべきは、NGS シミュレータは潜在的に、実際に NGS によって得られる NGS と区別できないようなリアルスティックなデータを生成し得るということである。すなわち、意図的なデータ改竄を行う上で NGS シミュレータは有効なツールになり得ると考えられる。

D.考察

a. 測定データ、画像データの証拠性向上
・次世代シーケンサー(NGS)の出力データと解析について

NGS から出力されるデータ(Raw Reads, fastq)は、そのままでは活用することができず、研究目的に応じてデータの Preprocess (前処理)やデータの normalization(正規化)を経て解析が可

能な状態となる。

これらの過程は試行錯誤を伴うこともあり、また、手順化された場合でも、その工程が複雑であり、対象となるデータ容量が大きく、複数の中間生成ファイルを伴うため様々な個所において、意図せぬエラーや意図的な改変が発生するリスクが存在する。更には、これら一連の作業は複数の研究者によって分担されることがあり、エラー発生リスクは高くなる。研究者は、これらのリスクを認識しており、様々な工夫によってエラーの発生の防止に努めているのが現状である。

・ワークフロー管理システム

workflow management system (ワークフロー管理システム) は、前述のリスクを軽減するため、また、研究の再現性を担保するためデータの処理から解析までの各タスクを管理プラットフォーム上で実施し、タスクの進捗や実施状況について可視化するとともにログ等のデータによる管理を可能としている。同システムは、pipeline framework と呼ばれることもある。

結果に示したように、ウェブアプリケーションの形態にて提供され、複数名での利用形態を想定して開発されたワークフロー管理システムでは、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されていた。これらは、他者から自己の成果物を保護するのみならず、自己の研究の正当性も示すことに利用できる可能性がある。

b. 研究ノートの電子化

研究ノートの電子化実験については、

1) 今回は Web 型電子カルテをカスタマイズなしで用いた、2) アンケート参加者が4名と少数であった、ことなどから結論を出すことは難しいが、意義のある試みであった。

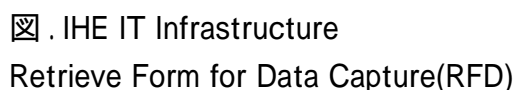
電子研究ノートは、画面がわかりにくく操作が煩雑でログインに時間がかかる、という課題が見える一方で、遠隔でも指導ができること、記録の検索性が高いことが、電子研究ノートの利点として挙げられた。機能的にはやはり不十分で、電子データファイルの管理が有効であること、記載の自由度が電子カルテよりも高い必要があること、記載したことが指導者・研究者相互にメールなどで把握できること、などが少なくとも必要であることが判明した。その一方で、紙の研究ノートではなく、遠隔地でもセキュアなコミュニケーションができる電子研究ノートの有用性も示された。今後の研究継続の参考としたい。

c 臨床データのバリデーションに関する検討

1 次利用目的の HIS データと 2 次利用目的のデータに齟齬が生じる可能性があるのは、(非意図的、あるいは改竄・捏造を問わず) 現実的には以下の経路である。

- ・ HIS あるいは部門システムからの直接の臨床研究用データベースへの EDC
- ・ HIS あるいは部門システムからの外付データ源への EDC
- ・ 真正性が担保されていない部門システムや外付データ源への操作
- ・ 外付データ源から臨床研究用データベースへの EDC

- 実際の臨床研究データの品質低下は、臨床研究 データベースへの移行の際、あるいは移行後の方が多いであろう。Integrating the Healthcare Enterprise (IHE) では、バリデーションの仕組み・タイミングと臨床介入研究や臨床観察研究としての正当性の保証について、図2のような EDC を活用したデータ回収のしくみを策定している。本研究結果を踏まえてデータ源の選択やデータ品質管理などの設計をするべきであろう。



d 研究用計測機器データの証拠性保全

NGS データのサイズは数ギガバイトから数十ギガバイトのである。今日では一般に市販されているノート型 PC でも大容量のものは 16GB 以上の場合の主記憶装置が搭載されている。従って、データを直接的に編集することすら困難ではなくなっている。

また、NGS データを処理するプログラムとしてオープンソースのものが多く配布されており、かつ、実際に主要な処理として用いられている。それらの多くは UNIX 系統の OS 上で動作し、その操作には GUI ではなくコマンドラインを用いるものが多い。プログラムの使用方法についての情報がインターネット上で掲載されているが、コマンドラインであればコ

ピー＆ペーストにより簡単に実行させることができる。

加えて、医学生物学系の研究者が Apple 社製のパーソナルコンピュータを好むことは頻繁に言及されるが、現在、それらのオペレーティングシステム(OS)は UNIX 系統であり、プログラムのインストールから実行についてのスキルの障壁は低いものとなっている。

つまり、NGS データに対して改竄する意思を持つものが十分な分子生物学的な知識を持つとは容易に想定できるが、そのような者が実際に改竄を実施しようとする場合に最初に障壁となると想定されるのが計算機関連の機材の調達やスキルの習得であるが、そのハードルが低くなっているということである。

2) ハッシュは NGS 内部で生成すべき

NGS データについてハッシュ値を埋込むことによって、埋込み以後の改竄を検知することはできるが、それを研究者に行わせることは、無意味である。NGS がデータを出力する段階でハッシュが埋込まれていなければならない。

3) ハッシュではなく電子署名でなければならない

NGS がハッシュ値を埋め込んだデータを出力したとしても、ハッシュと取除き、データを改竄し、再度ハッシュ値を埋込めば、やはり改竄はできてしまう。

この場合の改竄を検出するには、NGS は生成したデータのハッシュ値を研究者が操作できない所に半永久的に格納しておかなければならなくなる。これは全く現実的ではない。

この問題を解決するには公開鍵暗号技

術を用いた電子署名を作成し、それを NGS データに埋込む必要がある。この場合、改竄を検知するためには、電子署名を生成する際に用いた秘密鍵と対になる公開鍵を用いて電子署名を復号することになる。その為には何らかの公開鍵基盤(PKI)との連携が必要になる。

E. 結論

- a. NGS から出力されるデータの処理や解析を目的としたワークフロー管理システムについて、そのセキュリティ機能について調査した。複数名での利用を想定したワークフロー管理システムにおいては、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されていた。これらのセキュリティ管理機能は、研究データの証拠性保全を支援できる可能性がある。
- b. 証拠性を有する研究を進めるために、電子研究ノートの実証実験を実施し、アンケート調査により一定の有用性を確認した。
- c. 近年進んできた SS-MIX2 からの EDC は、導入から継続的なデータ品質管理を実施している限りにおいて、他よりも証拠性が強いことを示唆した。
- d. 実用的時間で電子署名を NGS データに処することが可能であることが判明したので、NGS データの証拠性を向上させるため、特に改竄を検知できるようにするためには、NGS や公共データベースに対して電子署名技術と習慣を導入する必要があることが明らかになった。

F. 健康危険情報

平成 28 年度の本研究においては、生命、健康に重大な影響を及ぼすと考えられる新たな問題、情報は取り扱わなかった。

G. 研究発表

1. 論文発表

木村通男，渡辺浩，澤智博，作佐部太也，中島直樹：証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化，第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，722-723，2016

中島直樹，渡辺浩，澤智博，作佐部太也，宇山佳明，山口光峰，木村通男：証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討，第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，714-717，2016

澤智博，渡辺浩，作佐部太也，中島直樹，木村通男：証拠性のある医学研究 次世代シーケンサー等のデータソースおよび解析ソフトウェアの検討，第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，718-719，2016

作佐部太也，大内雄矢，澤智博，渡辺浩，中島直樹，木村通男：証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用

についての評価と提案，第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，720-721，2016

2. 学会発表

木村通男，渡辺浩，澤智博，作佐部太也，中島直樹：証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化 第 36 回医療情報学連合大会，2016 年 11 月 24 日，横浜市

H. 知的財産権の出願・登録状況 (予定も含む)

- | | |
|----------|----|
| 1.特許取得 | なし |
| 2.実用新案登録 | なし |
| 3.その他 | なし |

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究

(H28-医療-指定-011)

- 研究ノートの電子化 -

および

- 臨床データのバリデーションに関する検討 -

研究分担者 中島 直樹

九州大学病院メディカルインフォメーションセンター 教授

研究要旨: 目的: 医学研究者自身が実施する研究の正当性を証明するため、情報システムに求める要件(各機器でこういった形でデータを持っておくべきか等)を明らかにするものである。方法: 研究1.既に広く市販・活用されているWeb型電子カルテから、市販の患者IDを研究者個別の研究課題、研究者を記載医師、研究グループを診療科、と読み替えて使用したための研究目的に合致しているシステム要件を持つ製品を選出した。研究者と指導者を九州大学病院で選定して、実際の複数の研究に対して1か月以上使用した。その後、研究者と指導者別にアンケート調査を行った。研究2.平成27年度研究の結果を基に、HIS・各種の部門システムからのデータ源の真正性や監査証跡の有無、およびデータ源との差分の比較の容易性を基に、データ源およびデータ回収法別(EDC・手入力別)にデータ信頼性を比較した。結果:研究1.個別の研究課題を持つ研究者2名と両研究課題に跨る研究指導者2名の計4名が実験に参加した。選択した電子カルテは場所を問わずアクセス可能であり、電子研究ノートとして使用可能であった。アンケートでは、一定の電子研究ノートの有用性が確認されたとともに利点・欠点が把握でき、必要な機能などを抽出することが出来た。研究2.各データ源からのEDCおよび手入力されたデータの信頼性は、1)HISと連携したSS-MIX2標準化ストレージからのEDCによるデータ回収、2)HISと連携したSS-MIX2拡張ストレージからのEDCによるデータ回収、3)HISや真正性を確保した部門システムから連携したSS-MIX2以外の外付データ源(DWHなど)からのEDCによるデータ回収、4)真正性を持たない部門システムからのSS-MIX2以外の外付データ源への出力、5)臨床研究用DBへのデータ手入力、の順にデータの信頼性は高いと想定された。結論: 証拠性を有する研究を進めるために、電子研究ノートの実証実験を実施し、アンケート調査により一定の有用性を確認した。また、近年進んできたSS-MIX2からのEDCは、導入から継続的なデータ品質管理を実施している限りにおいて、他よりも証拠性が強いことを示唆した。

研究協力者

九州大学病院メディカル・インフォメーションセンター

九州大学病院メディカル・インフォメーションセンター

特任助教 野原康伸

教室系技術職員 原田慶一

A. 研究背景と目的

近年、STAP 細胞問題をはじめとして、ディオバン医師主導臨床研究問題、J-ADNI（アルツハイマー研究プロジェクト）問題などのスキャンダルが立て続けに発生し、医学の基礎研究から臨床研究までの科学的信頼性が問題視されている。特に、研究の過程での「捏造」に関しては、研究ガイドラインが存在するものの、現実的には従来科学者の良心に委ねられる以外は、「研究ノート」という基本的には紙の記述物が存在するのみで、記載内容の真正性の保証や改竄の証拠性を担保するには不十分である。

そこで本研究では、医学研究者自身が実施する研究の正当性を証明するため、情報システムに求める要件（各機器でこういった形でデータを持っておくべきか等）を明らかにするものである。

ただし、本研究のフォーカスとしては、研究倫理や教育に大きく依存する臨床手法や研究手法に関する内容は対象外とする。

本研究班における分担研究者としての中島の役割は、平成 28 年度は、研究 1）研究ノートの電子化に関する実証研究を木村通男研究代表者と共に実施した。さらに平成 27・28 年度には、研究 2）臨床研究データ（特に処方・検査など）の品質管理、データを受け渡す際のプロトコル、およびそれらに対するバリデーション（人為的な操作の防止）を実施した。また、平成 28 年度には、臨床研究で使わ

れる情報項目を整理し、人的な操作が入る可能性がある Electronic Data Capture（EDC）ケースを想定した。なお、本研究は「クリニカルバリデーション」ではなく「システムバリデーション」を対象とした。

B. 研究方法

B.1 研究 1. 研究ノートの電子化

B1-1 実験に用いるアプリケーションの選定

既に広く市販・活用されている Web 型電子カルテから、研究目的に合致しているシステム要件を持つ製品を選択した。

まず本研究では、市販の電子カルテの患者 ID を研究者個別の研究課題、研究者を記載医師、研究グループを診療科、と読み替える。求められるシステム要件は

- ・複数診療科が設定可能、その属性により患者データの参照権限が設定可能（他研究グループが別課題のデータを参照できないように）

- ・Web 型のブラウザを利用したものであること（研究者の利便や費用を考慮して）

- ・会計機能、オーダ機能は必要なく、電子カルテ管理で最も手間がかかる各種項目（医事項目、検査項目など）のマスター管理は必要ない、などである（詳細は木村通男研究代表者の報告書参照）

magicConnect[®]（NTT テクノクロス社）と呼ばれるアプリケーションでリモートアクセスコントロールを行った。これは、同社の保有するマジックコネクトサーバ

を介して https で暗号化通信を確立するものである。

B1-2 実験(研究ノートとしての実際の試用)と使用後のアンケート調査

研究指導者 2 名(教員クラス)および九州大学医学研究院博士課程大学院生 2 名に対して研究の協力を依頼し、浜松医大に設置されたサーバに対する実際のアクセスなどを用いた研修会を九州大学病院内で実施した。

平成 28 年 3 月から 2 名大学院生のそれぞれの研究課題において研究ノートとして使用し、使用後にアンケート調査を行った。

アンケートの内容を別紙 1 に示す(別紙 1 参照)。アンケートは研究者用と指導者用の 2 種類を作成した。

B. 2 研究 2: 臨床データのバリデーションに関する検討

B2-1 EDC データ源としての SS-MIX2 標準化ストレージのデータの品質の検討

厚生労働省/PMDA が行っている MID-NET 事業や各種の疾患コホート研究においては、EDC のデータ源として SS-MIX2 標準化ストレージを活用している。それをデータ源とする場合、HIS から SS-MIX2 を介して各事業 DB への正確な移行が重要であるが、MID-NET バリデーション事業では、PMDA と共同でこれを検証した。HIS データと事業 DB 間のデータ一致率を、病名、処方、検査結果の 3 つ別々に調査し、その結果をフィードバックしてデータ品質を向上した。

B2-2 考えられるデータ源別からの EDC の信頼性検討

HIS 周囲には各種の部門システムがあり相互に連携している。これらから、データ源の真正性や監査証跡の有無、およびデータ源との差分の比較の容易性を基に、データ源およびデータ回収法別(EDC・手入力別)にデータ信頼性を比較した。

B. 3 倫理的配慮

本研究においては、特に個人情報を取扱うなどの倫理的な課題は発生しなかった。

C. 研究結果

C.1 研究 1. 研究ノートの電子化

C1-1 実験に用いるアプリケーションの選定

検討の結果、本研究では、NEC 社製の小病院用電子カルテシステム MegaOak SR for SaaS が、Web ブラウザをベースにしており、プラットフォームに特別なものを求めない点を重視し、採用した。これは外部サーバを利用した形式と、施設内にサーバを置く形式が可能であるが、今回は試用以後も使えるように、固定費用の発生しない形式、つまり浜松医大にサーバを置き、九州大学の研究グループからは電子証明書を用いた方法でのアクセスコントロールをおこなうこととした。浜松医大、九州大学の 2 施設でおこなったことから、施設内、外部サーバ利用をそれぞれ実現したこととなる。

C1-2 実験(研究ノートとしての実際の試用)と使用後のアンケート調査

浜松医大内にサーバが設置された NEC 社製の小病院用電子カルテシステム MegaOak SR for SaaS に対して、安全性を電子証明書で担保し、その範囲内で九州大学病院内、あるいは指導者や研究者の自宅などからアクセスできた。

研究者毎に異なる複数の研究課題について、研究ノートを設定し（患者 ID 機能を利用）、電子カルテの権限参照管理システムを用いて容易にアクセス権限管理を行うことが出来た。

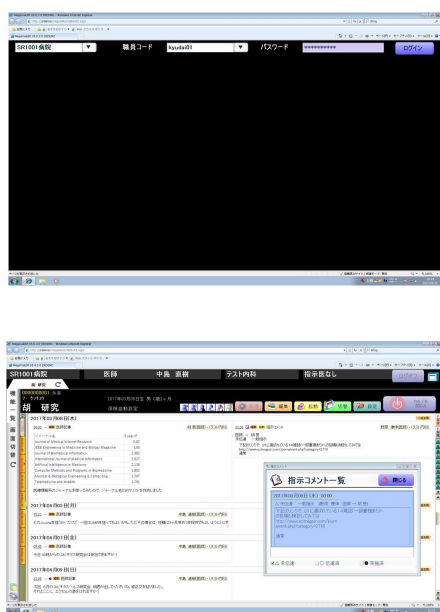


図 1．電子研究ノート画面。上）ログイン画面、下）電子研究ノート画面。電子カルテの患者 ID を研究者個別の研究課題、研究者を記載医師、研究グループを診療科、と読み替えて使用。

アンケートは研究者 2 名、指導者 2 名の少数の結果であったが、以下の結果を得た（結果は別紙 2 参照）。

生物実験や化学実験などの Wet Labo と違い、統計学を含む情報系研究者は Dry

Labo と呼ばれる。今回の対象者は Dry Labo の研究者と指導者であった。電子的に研究の半分以上の記録を記載していたのが 1 名で残り 1 名は記載していないと答えた。同時に 2 名の指導者も、これまでは紙の研究ノートを確認はしていなかった。なお、Dry Labo ではコンピューターが主たる研究業務の場であることが多く、計算式などの履歴をコンピューター上に残しながら仕事をする。そのため、Wet Labo に比較して Dry Labo では紙の研究ノートは作っていないことが多いと推定される。

電子研究ノートに参加して、4 名全員が記載の時間が増えたが、これまで記載や確認をしていなかった 3 名については新しい作業が発生したための当然の結果であるが、電子的に記載していたものにとっても、ログインなどの作業に手間がかかり、時間が増えたものと考えられる。

研究者と指導者との間で、電子研究ノートの機能の充足の意見（複数回答可）は分かれた。指導者はともに機能は不足していると答えたが、リアルタイムのやり取りが困難、電子データファイルの管理（変更履歴や変更内容の管理）が難しい、などの意見であった。

紙の研究ノートと電子研究ノートの使用感の比較（複数回答可）では指導者も研究者も同じ傾向であった。両者を足し合わせると、紙ノートの利点は、自由度が高い（3 名）、使いやすい（2 名）であり、欠点は、紛失する危険がある（4 名）、指導を得にくい（しづらい）（3 名）が多かった。一方で、電子研究ノートの利点は、記録の検索性が高い（3 名）が

多く、課題は、画面がわかりにくい、操作が複雑、ログインに手間がかかる、が多かった（各3名）。

改ざんの有無がわかる証跡が残っているかどうかについて指導者2名に聞いたところ、ともに証跡が残っている、と答えたが、これは電子カルテ機能を活用したためでもあろう。

その他の自由記載において、指導者からは「情報系(データ解析を含む)の研究では、多くの作業がコンピューター内で完結し、手作業の部分を除いて、ソースコードに作業内容が記録されるので、実験の再現も容易である。実験ノートという習慣がほとんどない一方で、電子データの管理が本質的に重要である。したがって、そこを証跡を含めて簡単に管理できる仕組みが必要だと考えられる。少なくともデータ解析の指導においては、電子カルテというよりも、GitHubのようなソフトウェア開発の仕組みを使った方が適していると思う。」という意見を得た。また、研究者からは、「デフォルトの枠が狭く、一行の文章を全文表示できない。マウス操作で枠を拡大できる。また、下までスクロールするスピードが遅く、文章が長いとき読みづらい。」という電子カルテ機能の制限に関する意見を得た。

C. 2 研究2：臨床データのバリデーションに関する検討

C2-1 EDC データ源としての SS-MIX2 標準化ストレージのデータの品質の検討

HIS と事業 DB 間のデータ一致率は、データ品質管理作業前には、傷病名、処方、検査結果情報で、99.1、67.0、55.8%であった。データ不一致の原因は、連携バ

グ、マッピング不備、多様なオーダ種や結果値の存在など様々であった。データ品質作業による是正の結果、99.9、100、100%となった。データ不一致の原因の一部は HIS と SS-MIX2 との連携が原因であったが、その原因も含めてほぼ全て見出し、解消することが可能であった。

C2-2 考えられるデータ源別からの EDC の信頼性検討(研究2)

平成 27 年度の本分担研究の結果を受けて、各種のデータ源およびデータ回収法別に、想定されるデータの信頼性に影響を与えるデータ源別の因子を表 1 に示す。

表 1 . データの信頼性に影響を与えるデータ源別の因子

	真正性	監査証跡 (移行)	HISとの差分の把握
HIS	○	○	—
SS-MIX2標準化ストレージ	×	○	○
SS-MIX2拡張ストレージ	×	○	△
各種部門システム	×～○	×～○	×～○
SS-MIX2以外の外部データ源	×	×～○	×～○

HIS では、真正性を確保していることが導入の条件である一方、各部門システムでは、HIS にデータ連携すれば HIS 側で真正性が確保されるためにそれ自体は必ずしも真正性を確保しているとは限らず、各病院の運用に依存する。また、HIS、部門システムを問わず業務用データベースであるため、そのデータ 2 次利用の場合は連携する外部集積データソースを利用することが多い。したがって、外部集積データソースの真正性（特に正確性）が確保できていなければ、適切な解析結果を得ることはできない。

SS-MIX2 標準化ストレージは HIS の傷病名、処方注射、検査結果データなどが移行されているが、SS-MIX では HIS における監査証跡が移行され、さらに MID-NET 関連事業で確立された手法、開発されたツールを用いた HIS データとの差分チェックの応用が可能となりつつある。現在、SS-MIX2 標準化ストレージの導入が全国的にも増加し、これを特に MID-NET 事業や各種疾患コホート研究などで EDC の外部集積データソースとして用いることが急激に増えている。一方で、標準化ストレージに入っていないデータ項目の HIS から SSMIX2 拡張ストレージへの移行データに対する EDC では、HIS での真正性の確保は標準化ストレージと同様だが、監査証跡の移行は運用により異なり、MID-NET 関連事業での差分チェックの経験がない。また、これはデータ項目種毎に差分のチェックの方法と課題の解決を経ねば向上しない。一方、部門システムは、前述のようにその用途によって真正性の確保は様々であり、改竄や捏造が入り込む可能性がある。さらに、SS-MIX2 以外の外付データ源（ベンダーの DWH など）では真正性の確保や監査証跡の移行などは行われず、差分チェックも様々でありデータ品質は低下する。

以上の結果、各データ源からの EDC および手入力されたデータの信頼性は、

1. HIS と連携した SS-MIX2 標準化ストレージからの EDC によるデータ回収
2. HIS と連携した SS-MIX2 拡張ストレージからの EDC によるデータ回収
3. HIS や真正性を確保した部門システムから連携した SS-MIX2 以外の外付

データ源（DWH など）からの EDC によるデータ回収

4. 真正性を持たない部門システムからの SS-MIX2 以外の外付データ源への出力
5. 臨床研究用 DB へのデータ手入力の順にデータの信頼性は高いと想定された。

D. 考察

D. 1 研究1. 研究ノートの電子化

研究ノートの電子化実験については、1) 今回は Web 型電子カルテをカスタマイズなしで用いた、2) アンケート参加者が 4 名と少数であった、ことなどから結論を出すことは難しいが、意義のある試みであった。

電子研究ノートは、画面がわかりにくく操作が煩雑でログインに時間がかかる、という課題が見える一方で、遠隔でも指導ができること、記録の検索性が高いことが、電子研究ノートの利点として挙げられた。機能的にはやはり不十分で、電子データファイルの管理が有効であること、記載の自由度が電子カルテよりも高い必要があること、記載したことが指導者・研究者相互にメールなどで把握できること、などが少なくとも必要であることが判明した。その一方で、紙の研究ノートではなく、遠隔地でもセキュアなコミュニケーションができる電子研究ノートの有用性も示された。今後の研究継続の参考としたい。

D. 2 研究2: 臨床データのバリデーションに関する検討

1 次利用目的の HIS データと 2 次利用目

的のデータに齟齬が生じる可能性があるのは、（非意図的、あるいは改竄・捏造を問わず）現実的には以下の経路である。

- ✓ HIS あるいは部門システムからの直接の臨床研究用データベースへの EDC
- ✓ HIS あるいは部門システムからの外付データ源への EDC
- ✓ 真正性が担保されていない部門システムや外付データ源への操作
- ✓ 外付データ源から臨床研究用データベースへの EDC
- ✓ HIS などに蓄積した臨床データを手入力で臨床研究用データベースへ入力する際、あるいは入力後の操作

これらから、手入力をできるだけ減らし EDC を重視すること、部門システムも可能な限り 真正性を確保すること、監査証跡をできるだけ移行すること、データ品質検証が行いやすい外付データ源を用いること（SS-MIX2 標準化ストレージなど）、臨床研究用データベースも真正性を担保すること、などが 2 次利用する臨床データの信頼性を高める上では重要ということがわかる。なお、HIS や部門システムからの EDC に関しては、MID-NET 事業以外に、日本循環器学会（平成 26 年度～）、日本糖尿病学会（平成 27 年度～）、あるいは 日本腎臓学会（平成 27 年度～）が構築している各種の疾患コホート研究などがあり、毎年増加している。これらはいずれも SS-MIX2 標準化ストレージや HIS からの EDC を行っており、意図的なデータ操作は入りにくい。

実際の臨床研究データの品質低下は、臨床研究 データベースへの移行の際、あ

るいは移行後の方が多いであろう。Integrating the Healthcare Enterprise（IHE）では、バリデーションの仕組み・タイミングと臨床介入研究や臨床観察研究としての正当性の保証について、図 2 のような EDC を活用したデータ回収のしくみを策定している。本研究結果を踏まえてデータ源の選択やデータ品質管理などの設計をするべきであろう。

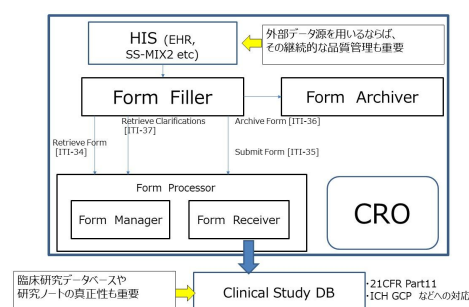


図 2 . IHE IT Infrastructure
Retrieve Form for Data Capture(RFD)

図 2 の Form Filler を SS-MIX2 標準化ストレージなどと連携し、また利用者認証、真正性担保を行うことにより、データの質を確保することが可能となる。今後は 21CFR Part11 や ICH GCP への対応の重要な位置づけとなるであろう。

E. 結論

証拠性を有する研究を進めるために、電子研究ノートの実証実験を実施し、アンケート調査により一定の有用性を確認した。また、近年進んできた SS-MIX2 からの EDC は、導入から継続的なデータ品質管理を実施している限りにおいて、他よりも証拠性が強いことを示唆した。

F. 健康危険情報

平成 28 年度の本研究においては、生命、健康に重大な影響を及ぼすと考えられる新たな問題、情報は取り扱わなかった。

G. 研究発表

1. 論文発表

- 1) 中島直樹，渡辺浩，澤智博，作佐部太也，宇山佳明，山口光峰，木村通男：証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討 第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，714-717，2016.
- 2) 澤智博，渡辺浩，作佐部太也，中島直樹，木村通男：証拠性のある医学研究 次世代シーケンサー等のデータソースおよび解析ソフトウェアの検討 第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，718-719，2016
- 3) 作佐部太也，大内雄矢，澤智博，渡辺浩，中島直樹，木村通男：証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用についての評価と提案 第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，720-721，2016
- 4) 木村通男，渡辺浩，澤智博，作佐部太也，中島直樹：証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化 第 36 回医療情報学連合大会抄録集 722-723，2016

2. 学会発表

- 1) 中島直樹，渡辺浩，澤智博，作佐部太也，宇山佳明，山口光峰，木村通男：証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討 第 36 回医療情報学連合大会，2016 年 11 月 24 日，横浜市

H. 知的財産権の出願・登録状況

(予定も含む)

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 特になし

別紙1

1) 研究者向けアンケート

生産性について確認

1 - 1 . これまでの研究生活の中で研究ノートに記載していましたか？（一つを選択）

A. 紙ノートに記載していた

B. 電子的に記録していた

C. 記載していない

【1 - 1 で A, B を選択した方のみ回答して下さい】

1 - 2 . 研究ノートはどの頻度で記載していましたか？（一つを選択）

必ず記載していた（研究内容全て）

ある程度記載していた（研究内容の半分以上）

あまり記載していなかった（研究内容の半分未満）

1 - 3 . 今回、電子研究ノート研究に参加して、この研究ノートを記載しましたか？（一つを選択）

必ず記載した（研究内容全て）

ある程度記載した（研究内容の半分以上）

あまり記載しなかった（研究内容の半分未満）

全く記載しなかった

1 - 4 . 電子研究ノートを利用して、記録に係る時間は増えましたか。

記録に係る時間は増えた（今まで： 分 電子研究ノート： 分）

記録内容が増えた

操作、入力の手間が増えた

その他（ ）

記録を行う時間は減った（今まで： 分 電子研究ノート： 分）

記録内容が減った

入力効率が良くなった

その他（ ）

以前と変わらない

記録が増えたが、効率が良くなった。

記録は減ったが、手間が増えた。

記録、手間ともに変わらない

有効性について確認

2 - 1 . 研究ノートとして利用するうえで、必要な機能を満たしていると思いますか。

はい

いいえ

2 - 2 . (2 - 1 でいいえと答えた方) 研究ノートとして足りない機能を挙げてください。

(自由記載)

満足性について確認

3 - 1 . 紙の研究ノートの良いと思う点を挙げてください。(複数回答可)

使いやすい

自由度が高い

記録をすぐに確認できる

指導が得られやすい

どこでも記載ができる

その他

(自由記載)

3 - 2 . 紙の研究ノートの悪いと思う点を挙げてください。(複数回答可)

使いにくい

自由度が低い

記録を確認しづらい

指導が得られにくい

紛失する危険がある

その他

(自由記載)

3 - 3 . 電子研究ノートを使ってみて、良いと思った点を挙げてください(複数回答可)

画面 (インターフェイス) が分かりやすい

操作が簡単

自由度が高い

記録の検索性が高い
指導が得られやすい
機能が豊富
どこでも記載ができる
その他
(自由記載)

3 - 4 . 電子研究ノートを使ってみて、今後の改良が必要と思う点を挙げてください (複数回答可)

画面 (インターフェイス) が分かりにくい
操作が複雑
自由度が低い
記録の検索性が悪い
指導を受けにくい
機能が足りない
ログインに手間がかかる
その他
(自由記載)

#####

安全性について確認
(研究者向けは省略)

#####

その他

4 . その他なんでも気が付いたことを記載してください。
(自由記載)

1) 指導者向けアンケート

生産性について確認

1 - 1. これまでの研究指導生活で研究ノートを確認していましたか？（一つを選択）

- A. 紙ノートを確認していた
- B. 電子的に確認していた
- C. 状況により、紙と電子的記録を使い分けていた
- D. 確認していない

【1 - 1でA,B,Cを選択した方のみ回答して下さい】

1 - 2. 研究ノートはどの頻度で確認していましたが？（一つを選択）

- 必ず確認していた（研究内容全て）
- ある程度確認していた（研究内容の半分以上）
- あまり確認していなかった（研究内容の半分未満）

【1 - 1でA,B,Cを選択した方のみ回答して下さい】

1 - 3. これまでの研究指導生活で研究ノートに指導内容を記載していましたが？（一つを選択）

- 必ず記載していた（研究内容全て）
- ある程度記載していた（研究内容の半分以上）
- あまり記載していなかった（研究内容の半分未満）
- 全く記載していなかった

1 - 4. 今回、電子研究ノート研究に参加して、この研究ノートを確認しましたか？（一つを選択）

- A. 必ず確認した（研究内容全て）
- B. ある程度確認した（研究内容の半分以上）
- C. あまり確認しなかった（研究内容の半分未満）（理由： ）
- D. 全く確認しなかった（理由： ）

【1 - 4でA,B,Cを選択した方のみ回答して下さい】

1 - 5. 今回の電子研究ノートに指導内容を記載しましたが？（一つを選択）

- 必ず記載した（指導内容全て）
- ある程度記載した（指導内容の半分以上）
- あまり記載しなかった（指導内容の半分未満）
- 全く記載しなかった

- 1 - 6 . 電子研究ノートを利用して、研究ノートの確認に係る時間は増えましたか。
- 確認に係る時間は増えた（今まで： 分 電子研究ノート： 分）
- 記録内容が増えていた
- 操作、入力の手間が増えた
- その他
- 確認を行う時間は減った（今まで： 分 電子研究ノート： 分）
- 記録内容が減っていた
- 確認効率が良くなった
- その他
- 以前と変わらない
- 記録が増えたが、効率が良くなった。
- 記録は減ったが、手間が増えた
- 記録、手間ともに変わらない

- 1 - 7 . 電子研究ノートを利用して、指導記録の記載に係る時間は増えましたか。
- 記録に係る時間は増えた（今まで： 分 電子研究ノート： 分）
- 記録内容が増えた
- 操作、入力の手間が増えた
- その他
- 記録を行う時間は減った（今まで： 分 電子研究ノート： 分）
- 記録内容が減った
- 入力効率が良くなった
- その他
- 以前と変わらない
- 記録が増えたが、効率が良くなった。
- 記録は減ったが、手間が増えた。
- 記録、手間ともに変わらない

有効性について確認

- 2 - 1 . 研究ノートとして利用するうえで、必要な機能を満たしていると思いますか。
- はい
- いいえ
- 2 - 2 . (2 - 1 でいいえと答えた方) 研究ノートとして足りない機能を挙げてください。
(自由記載)

満足性について確認

3 - 1 . 紙の研究ノートの良いと思う点を挙げてください。(複数回答可)

使いやすい

自由度が高い

記録をすぐに確認できる

指導しやすい

どこでも記載ができる

その他

(自由記載)

3 - 2 . 紙の研究ノートの悪いと思う点を挙げてください。(複数回答可)

使いにくい

自由度が低い

記録を確認しづらい

指導がしづらい

紛失する危険がある

その他

(自由記載)

3 - 3 . 電子研究ノートを使ってみて、良いと思った点を挙げてください(複数回答可)

画面(インターフェイス)が分かりやすい

操作が簡単

自由度が高い

記録の検索性が高い

指導しやすい

機能が豊富

どこでも記載ができる

その他

(自由記載)

3 - 4 . 電子研究ノートを使ってみて、今後の改良が必要と思う点を挙げてください（複数回答可）

画面（インターフェイス）が分かりにくい

操作が複雑

自由度が低い

記録の検索性が悪い

指導を行いにくい

機能が足りない

ログインに手間がかかる

その他

（自由記載）

安全性・信頼性について確認

4 - 1 . 電子研究ノートに改竄の疑いがないことが分かる十分な証跡が残っていますか。

はい

いいえ

【4 - 1 でいいえと答えた方のみ回答して下さい】

4 - 2 . 不足している証跡があれば、記載して下さい。

（自由記載）

その他

5 . その他なんでも気が付いたことを記載してください。

（自由記載）

【参考資料】

http://www.meti.go.jp/policy/it_policy/softseibi/index.html

システム及びソフトウェア品質の見える化、確保及び向上のためのガイド

http://www.meti.go.jp/policy/it_policy/softseibi/metrics/product_metrics.pdf

表 1-5 ISO/IEC 9126-4 の利用時の品質に係る品質特性とメトリクス例

	品質特性	メトリクス例
利用時の品質特性	有効性	<ul style="list-style-type: none">・作業有効性・作業完了度・誤り頻度
	生産性	<ul style="list-style-type: none">・作業時間・作業効率性・経済生産性・生産的な割合・相対的利用者効率性
	安全性	<ul style="list-style-type: none">・利用者の健康及び安全性・システム利用で影響を受ける人の安全性・経済的損害・ソフトウェア損害
	満足度	<ul style="list-style-type: none">・満足度尺度・満足度質問表・自由裁量の使用

別紙2

「e実験ノート」のアンケート調査						
確認項目	No.	設問・アンケート内容		回答・選択肢の内容	回答数	詳細・自由記載
生産性について (指導者向け)	1-1	これまでの研究指導生活の中で研究ノートを確認していましたか？(一つを選択)	指導者	A 紙ノートを確認していた B 電子的に確認していた C 状況により、紙と電子的記録を使い分けていた D 確認していない	0 0 0 2	
	1-2	[1 - 1でA,B,Cを選択した方のみ回答] 研究ノートはどの頻度で確認していましたか？(一つを選択)	指導者	必ず確認していた(研究内容全て) ある程度確認していた(研究内容の半分以上) あまり確認していなかった(研究内容の半分未満)		
	1-3	[1 - 1でA,B,Cを選択した方のみ回答] これまでの研究指導生活で研究ノートに指導内容を記載していましたか？(一つを選択)	指導者	必ず記載していた(研究内容全て) ある程度記載していた(研究内容の半分以上) あまり記載していなかった(研究内容の半分未満) 全く記載していなかった		
	1-4	今回、電子研究ノート研究に参加して、この研究ノートを確認しましたか？(一つを選択)	指導者	必ず確認した(研究内容全て) ある程度確認した(研究内容の半分以上) あまり確認しなかった(研究内容の半分未満)(理由:) 全く確認しなかった(理由:)	0 0 2 0	(理由:研究者が研究ノートを書くことに慣れていない、アクセスに時間と手間がかかる、など) (理由: 未記入)
	1-5	[1 - 4でA,B,Cを選択した方のみ回答して下さい] 今回の電子研究ノートに指導内容を記載しましたが？(一つを選択)	指導者	必ず記載した(指導内容全て) ある程度記載した(指導内容の半分以上) あまり記載しなかった(指導内容の半分未満) 全く記載しなかった	0 0 2 0	
	1-6	電子研究ノートを利用して、研究ノートの確認に係る時間は増えましたか、	指導者	確認に係る時間は増えた(今まで: 分 電子研究ノート:分) 確認を行う時間は減った 以前と変わらない	2 0 0	(今まで:0分 電子研究ノート:3分) (今まで:分 電子研究ノート:10分) 記録内容が増えた1, 操作・入力の手間が増えた2, その他0
	1-7	電子研究ノートを利用して、指導記録の記載に係る時間は増えましたか、	指導者	確認に係る時間は増えた(今まで: 分 電子研究ノート:分) 確認を行う時間は減った 以前と変わらない	2 0 0	(今まで:0分 電子研究ノート:10分) (今まで:分 電子研究ノート:10分) 記録内容が増えた1, 操作・入力の手間が増えた1, その他0
生産性について (研究者向け)	1-1	これまでの研究生活の中で研究ノートを記載していましたか？(一つを選択)	研究者	A 紙ノートに記載していた B 電子的に記録していた C 記載していない	0 1 1	
	1-2	[1-1でABを選択した方のみ] 研究ノートはどの頻度で記載していましたか？(一つを選択)	研究者	必ず記載していた(研究内容全て) ある程度記載していた(研究内容の半分以上) あまり記載していなかった(研究内容の半分未満)	0 1 0	
	1-3	[1-1でABを選択した方のみ] 今回、電子研究ノート研究に参加して、この研究ノートを記載しましたか？(一つを選択)	研究者	必ず記載した(研究内容全て) ある程度記載していた(研究内容の半分以上) あまり記載していなかった(研究内容の半分未満) 全く記載しなかった	0 0 1 0	
	1-4	[1-1でABを選択した方のみ] 電子研究ノートを利用して、記録に係る時間は増えましたか、	研究者	記録に係る時間は増えた(今まで: 分 電子研究ノート:分) 記録を行う時間は減った(今まで: 分 電子研究ノート:分) 以前と変わらない	1 0 0	(今まで:10分 電子研究ノート:20分) 記録内容が増えた0, 操作・入力の手間が増えた1, その他0
	2-1	研究ノートとして利用するうえで、必要な機能を満たしていると思いますか、	研究者	はい いいえ	0 2	
	2-2	(2 - 1でいいえと答えた方)研究ノートとして足りない機能を挙げてください。(自由記載)	指導者	はい いいえ	2 0	(自由記載) 例えば研究ノートに記載した、というお知らせメールをemailがlineなどで共有すること、そうでなければリアルタイムのやり取りができない。 (自由記載) データ解析においては、ソースファイルなどの電子データの管理が本質的に重要で、実験手順を記載した研究ノートに相当すると考えられる。電子カルテではファイルの添付は一応出来るが(使い方をよく知らないだけかもしれない)、ファイルのどこをどのように変えたか等のファイル管理が簡単にできないというのは難しい。

満足性について	3-1	紙の研究ノートの良いと思う点を挙げてください。(複数回答可)	指導者	使いやすい	1	
				自由度が高い	2	
				記録をすぐに確認できる	1	
				指導しやすい	0	
				どこでも記載ができる	0	
				その他	0	
			研究者	使いやすい	1	
				自由度が高い	2	
				記録をすぐに確認できる	0	
				指導が得られやすい	0	
				どこでも記載ができる	1	
				その他	0	
	3-2	紙の研究ノートの悪いと思う点を挙げてください。(複数回答可)	指導者	使いにくい	0	
				自由度が低い	0	
				記録を確認しづらい	0	
				指導がしづらい	2	
				紛失する危険がある	2	
				その他	2	(自由記載) どこでも記載が出来ない (自由記載) 遠隔地との間だと、紙の研究ノートは事実上使えない
			研究者	使いにくい	0	
				自由度が低い	0	
				記録を確認しづらい	1	
				指導が得られにくい	2	
				紛失する危険がある	2	
				その他	0	
	3-3	電子研究ノートを使ってみて、良いと思った点を挙げてください(複数回答可)	指導者	画面(インターフェイス)が分かりやすい	0	
				操作が簡単	0	
				自由度が高い	0	
				記録の検索性が高い	1	
				指導しやすい	0	
				機能が豊富	0	
			研究者	どこでも記載ができる	1	
				その他	1	(自由記載) 電子カルテと同じレベルのセキュリティがあるという安心感
				画面(インターフェイス)が分かりやすい	1	
				操作が簡単	0	
				自由度が高い	0	
				記録の検索性が高い	2	
	3-4	電子研究ノートを使ってみて、今後の改良が必要と思う点を挙げてください(複数回答可)	指導者	指導が得られやすい	1	
				機能が豊富	0	
				どこでも記載ができる	0	
				その他	0	
			研究者	画面(インターフェイス)が分かりにくい	1	
				操作が複雑	1	
				自由度が低い	0	
				記録の検索性が悪い	1	
				指導を行いにくい	0	
				機能が足りない	1	
安全性・信頼性について	4-1	電子研究ノートに改竄の疑いがないことが分かる十分な証拠が残っていますか。	指導者	はい	2	
				いいえ	0	
	4-2	[4 - 1でいいえと答えた方のみ回答して下さい] 不足している証拠があれば、記載して下さい。 (自由記載)				
その他	5	その他なんでも気が付いたことを記載してください。	指導者			(自由記載) 情報系(データ解析を含む)の研究では、多くの作業がコンピューター内で完結し、手作業の部分を除いて、ソースコードに作業内容が記録されるので、実験の再現も容易である。実験ノートという習慣がほとんどなく、電子データの管理が本質的に重要である。したがって、そこを簡単に管理できる仕組みが必要だと考えられる。少なくともデータ解析の指導においては、電子カルテというよりも、GitHubのようなソフトウェア開発の仕組みを使った方が適していると思う。
	4	その他なんでも気が付いたことを記載してください。	研究者			(自由記載) デフォルトの枠が狭く、一行の文章を全文表示できない。マウス操作で枠を拡大できる。 また、下までスクロールするスピードが遅く、文章が長いとき読みづらい。

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究

(H27-医療-指定-016)

- 研究用計測機器ユーザー認証等証拠性保全検討 -
次世代シーケンサー等のデータ解析ソフトウェアの検討

研究分担者 澤 智博

帝京大学医療情報システム研究センター 教授

研究要旨： 医学研究における各種データ、特に次世代シーケンサーから出力されるデータについて、解析ソフトウェア、主にワークフロー管理システムに関して、証拠性保全を検討するための調査を実施した。次世代シーケンサーから出力されるデータを解析するためのワークフロー管理システムについて、セキュリティ全般、ユーザー管理、ログ管理について証拠性保全の視点でプロパティや機能を精査し、証拠性保全のために必要な要素について検討した。

A.研究目的

医学研究におけるデータ解析について、電子カルテの証拠性保全の技術・運用の適用可能性を検討するため、次世代シーケンサーから出力されるデータに対するワークフロー管理システムについて調査する。

- KNIME
- OnlineHPC
- UGENE
- VisTrails

B.研究方法

複数のワークフロー管理システムについて、セキュリティ全般、ユーザー管理、アクセス権限管理、ログ管理に関するドキュメント類を精査する。

・調査対象のワークフロー管理システム

- Anduril
- BioBIKE
- Cuneiform
- Discovery Net
- Galaxy
- GenePattern

倫理的配慮

平成 28 年度の分担研究においては、特に個人情報を取扱うなどの倫理的な課題は発生しなかった。

C.研究結果

・ワークフロー管理システムのセキュリティについて

ワークフロー管理システムのセキュリティについては、そのソフトウェアの開発コンセプトと想定されるユースケースによって大きく異なる。個人での利用を想定しデスクトップアプリケーションとして開発されたワークフロー管理システムは、機

動性や個人としてのユーザビリティに力が置かれ作成中のワークフローに対する権限管理の機能等が存在していない。一方で、サーバーあるいはクラウドを前提したワークフロー管理システムでは、複数人による利用を前提としており、ユーザー認証、アクセス権限管理、システム使用ログ管理などの機能を備えている。

調査対象のワークフロー管理システムにおいては、galaxy、GenePattern、KNIMEにおいてセキュリティ全般を管理する機能が実装されていることを確認した。

・ galaxy のセキュリティ管理機能

galaxy においては、security package が実装されドキュメントが公開されている。代表的な機能としては、データ・ライブラリ・フォルダー等に対する権限管理、ロールの設定、表示・非表示、等がある。また、サブモジュールとして、パスワード管理、バリデーションチェック機能が存在する。ログの取得について異なる詳細レベルで可能となっている。また、ユーザーの使用状況についてレポートを生成する機能が実装されている。

・ GenePattern のセキュリティ管理機能

GenePattern は、ウェブアプリケーションサーバーを基本としており、そのためアクセス管理に必要な機能が充実している。主な機能として、アクセスフィルタリング、パスワード設定、ユーザー管理、ユーザー権限管理、SSL 対応、が実装されている。ユーザー管理機能に関して、グループの権限設定が可能になっている。ログ管理機能については、インストールロ

グ、GenePattern ログ、ウェブサーバーログの三種類のログについて管理が可能な仕様となっている。

・ KNIME のセキュリティ管理機能

KNIME には複数のプロダクトラインが存在するが、KNIME Server は複数人による利用を想定しており、そのユースケースに応じたセキュリティ管理機能を備えている。共有レポジトリにおいては、データ、ノード・メタノード、ワークフローにおいてアクセス権限を設定している。また、これらについて版管理機能を備えている。

ログ管理機能については、中間生成物の管理、トラッキングや監査機能を備えている。

D. 考察

・ 次世代シーケンサーの出力データと解析について

次世代シーケンサー (NGS) から出力されるデータ (Raw Reads, fastq) は、そのままでは活用することができず、研究目的に応じてデータの Preprocess (前処理) やデータの normalization (正規化) を経て解析が可能な状態となる。

これらの過程は試行錯誤を伴うこともあり、また、手順化された場合でも、その工程が複雑であり、対象となるデータ容量が大きく、複数の中間生成ファイルを伴うため様々な個所において、意図せぬエラーや意図的な改変が発生するリスクが存在する。更には、これら一連の作業は複数の研究者によって分担されることがあり、エラー発生リスクは高くなる。研究者は、これらのリスクを認識しており、様々な工夫によってエラーの発生の防止に努めてい

るのが現状である。

・ワークフロー管理システム

workflow management system (ワークフロー管理システム) は、前述のリスクを軽減するため、また、研究の再現性を担保するためデータの処理から解析までの各タスクを管理プラットフォーム上で実施し、タスクの進捗や実施状況について可視化するとともにログ等のデータによる管理を可能としている。同システムは、pipeline framework と呼ばれることもある。

結果に示したように、ウェブアプリケーションの形態にて提供され、複数名での利用形態を想定して開発されたワークフロー管理システムでは、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されていた。これらは、他者から自己の成果物を保護するのみならず、自己の研究の正当性も示すことに利用できる可能性がある。

E. 結論

次世代シーケンサーから出力されるデータの処理や解析を目的としたワークフロー管理システムについて、そのセキュリティ機能について調査した。複数名での利用を想定したワークフロー管理システムにおいては、ユーザー管理、アクセス権限管理、ログ管理の機能が実装されていた。これらのセキュリティ管理機能は、研究データの証拠性保全を支援できる可能性がある。

F. 健康危険情報

平成 28 年度の本研究においては、生命、

健康に重大な影響を及ぼすと考えられる新たな問題、情報は取り扱わなかった。

G. 研究発表

1. 論文発表

Nakata Y, Watanabe Y, Narimatsu H, Yoshimura T, Otake H, Sawa T. Surgeons' efficiency change is a major determinant of their productivity change. Int J Health Care Qual Assur. 2016 May 9;29(4):417-424.

澤智博. クラウドシステム・仮想化技術は医療情報システムを高度化させる鍵となるか. 月刊新医療. 43(7):78-82. 2016

中田善規, 成松宏人, 吉村達也, 大嶽浩司, 澤智博. 医療経済学を応用した手術室運営. 麻酔 65(S):S114-S121. 2016

澤智博. 人工知能による自動診断. Clinical Neuroscience. 34(8):928-31. 2016

澤智博, 渡辺浩, 作佐部太也, 中島直樹, 木村通男: 証拠性のある医学研究 次世代シーケンサー等のデータソースおよび解析ソフトウェアの検討第 36 回医療情報学連合大会, 医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2), 718-719, 2016

中島直樹, 渡辺浩, 澤智博, 作佐部太也, 宇山佳明, 山口光峰, 木村通男: 証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討 第 36 回医療

情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，714-717，2016.

作佐部太也，大内雄矢，澤智博，渡辺浩，中島直樹，木村通男：証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用についての評価と提案第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，720-721，2016

木村通男，渡辺浩，澤智博，作佐部太也，中島直樹：証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化 第 36 回医療情報学連合大会抄録集 722-723，2016

2. 学会発表

澤智博，渡辺浩，作佐部太也，中島直樹，木村通男：証拠性のある医学研究 次世代シーケンサー等のデータソースおよび解析ソフトウェアの検討第 36 回医療情報学連合大会，2016 年 11 月 24 日，横浜市

H. 知的財産権の出願・登録状況

(予定も含む)

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 特になし

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究

(H27-医療-指定-016)

- 測定、画像データ証拠性・安全運用環境検討 -

遺伝子解析研究への証拠性の導入

研究分担者 作佐部 太也

藤田保健衛生大学医療科学部臨床工学科 准教授

研究要旨: 目的: 遺伝子解析研究における証拠性を確保する上で、データ改ざんの防止に暗号技術の導入するための方法を検討することを目的とする。方法: 塩基配列データについてのハッシュ値を計算する実験を行い、また、改ざん防止のためのシステム構築の可能性についての調査を行った。結果: ハッシュ値の計算の実施には技術的な問題のないことが分かった。一方、改ざん防止システムの構築の必要性が高いことが明らかになるとともに、実現のためには多くの問題があることも明らかとなった。結論: 改ざん防止のシステムを構築するためには、研究者や学会などの学術組織だけでなく企業などとも連携した社会的な取り組みが必要であることがわかった。

A.研究目的

今日においてゲノムの塩基配列解析研究における次世代シーケンサー(NGS)は、基礎研究を超えて臨床研究のツールとなりつつある。またNGSが生成する塩基配列データ(以下NGSデータ)を活用するための各種の基盤の整備も進んでおり、統計処理やデータベースなどで高度な技術が投入され、更には、それらを自由に利用する文化が根付つつある。

このことは研究の発展という観点からは良好な状況と考えられるが、研究における不正の防止という観点からはリスクを孕んだ状況でもある。特にデータの改ざんという不正の防止についての検討が必要と考えられる。これは画像処理ツールの利用の普及が研

究における不正を少なからず助長したという指摘¹⁾からも、類似の現象がNGSデータにおいても起こることは推測しうることである。

そこで、本分担研究においては、前年度の調査を踏まえ、NGSデータの証拠性確保のため暗号化技術の導入のための方法について、調査研究および提案を行う。

B.研究方法

1) ハッシュ値の計算実験

データの改ざんを検出するための基盤となる情報処理はハッシュ値などの要約データの生成であり²⁾、どのような高度な改ざん防止の仕組みを構築するにしても、ハッシュ値の生成と元データとの関連付けについての実現性は前提となる。そこ

で本分担研究ではNGSデータについてのハッシュ値の生成を実際に行いその時間を計測した。また、生成したハッシュ値とNGSデータとの関連付けの方法として最も単純で信頼性の高い埋込みについて、実際にハッシュ値をNGSデータに埋込む実験を行い、NGSデータを解析するシステムに影響を及ぼさない方式を探索的に求めた。

2) NGS関連環境の調査

ハッシュ値の生成・埋込みをどのように行うことがデータ改竄の防止に効果的であるかについて検討するために、実際のNGSの運用現場や公共データベース、研究者が利用するソフトウェア等について文献やインターネットなどによる資料に基づいて調査を行った。

倫理的配慮

平成 28 年度の分担研究においては、特に個人情報を取扱うなどの倫理的な課題は発生しなかった。

C.研究結果

1) ハッシュ値の計算実験

ハッシュ値の計算実験については、添付1、添付2によって報告した。NGS関連で運用されるサーバと同程度の能力の計算機を用いて全エクソームで30秒程度、全ゲノムでも10分以内であった。元来NGSデータは巨大であり、どのような処理にしても分単位の時間がかかることから、ハッシュ値の計算にかかる時間はNGSデータの解析処理の中で行っても全体的な処理時間への影響はほとんどないと考えられる。また、ハッシュ値のNGSデータの埋込みについても、解

析システムに影響を与えない方式の実現性も確認できた。

2) 生成・埋込みのタイミング

ハッシュ値の生成・埋込みの処理を誰が、また、どのようなタイミングで行うことができるかについて調査、検討を行った。

なお、NGS自身にそのような機能を付与することは理想的であるが、NGSの製造メーカーへの問い合わせなどは本分担研究では行っていない。

NGSを運用する施設において処理を実施するケースについて検討した。

一例として分担研究者の所属する研究機関においては生物学系研究者が独自に解析システムを構築し運用している(中間報告書参照)。解析システムの構成やシステムを構築した研究者の情報処理技術についてのスキルからみて、プログラムを提供すれば、NGSの出力直後にハッシュ値の生成・埋込みを行うプログラムを解析システムに組込むことは可能であると考えられる。

より大規模な研究機関であれば、情報処理の専門技術者や研究者も所属している場合もあり、プログラムの開発を含めて可能であろう。

NGSを運用しているが自前の解析システムを持たない研究機関においてはNGSデータを解析業者に送付する前に、研究者がハッシュ値の計算と埋め込みを行わなければならない。従って計算機操作についての高度なスキルを持たない研究者でも操作できるよう、操作が容易なプログラムを提供する必要がある。また、NGSを持たず試料を委託先に送付してNGSの処理を外部に委託して行う場合には受託業者が行うことになるが、技術的には問題は無いと考えられる。

何れの場合においても、ハッシュ値の生成および埋込みの方法について正確に記述された仕様を策定し配布することは必須である。可能であれば、実際に動作するプログラムを配布することが望ましく、特に解析システムがオープンソースのプログラムによって構築される場合が多いことから、オープンソースとして配布するのが望ましいと考えられる。

3) 公共データベース

NGSを用いて取得した塩基配列情報の解析に基づく研究では、研究成果の公表の際には際には取得したNGSデータの公共データベースへの登録公開が義務付けられることがある。また、登録されたNGSデータは当該研究の証拠としてだけでなく、以後の別の研究において参照され再利用されることになる。

したがって、公共データベースへの登録前にハッシュ値は埋込まなければならない、公共データベースから取得するデータにも保持されていなければならない。

公共データベースとして、国際的に協調、集約の動きがあり現在その中心となっているのは、INSDC (The International Nucleotide Sequence Database Collaboration) の活動として米国のNCBI (The National Center for Biotechnology Information)、欧州のEMBL-EBI (The European Bioinformatics Institute)、日本のDDBJ (DNA Data Bank of Japan) が協調して蓄積、管理しているSRA (Sequence Read Archive)である³⁾。

SRAへの登録の際にはNGSデータはそのまま、研究や実験に関する情報は別のデータとして登録する。

SRAの内部ではNGSデータは独自のフォ

ーマット(SRA形式)で保存されており、利用するためには解析システムに入力できるFastQ形式などに変換する必要がある。変換のタイミングとしては、利用者がSRA形式のデータをダウンロードして専用のツールにより変換する方法と、専用のダウンロードツールにより自動的に目的のフォーマットとしてダウンロードする方法がある。

従ってSRAの運営組織に対して埋め込んだハッシュ値を維持できるようシステムの改造を要請する必要がある。

4) NGSシミュレータ

NGSの普及とともに、NGSをシミュレーションするシステム(以下NSGシミュレータ)が開発されている⁴⁾。NGSシミュレータは仮想的なNGSデータを生成するソフトウェアである。NGSシミュレータについての調査研究によると、塩基配列の変異について指定してNGSデータを生成できる機能をもつNGSシミュレータがある。これは、NGSデータの解析システムの挙動を詳細に検証するために重要な機能である。一方、別の観点からみると、NGSシミュレータは、実際のNSGによって得られるNGSデータと区別できないようなリアルスティックなデータを生成する可能性があるということである。すなわち、意図的な改ざんを行う上でNGSシミュレータは有効なツールになり得ると考えられる。

D.考察

1) 改ざんへの障壁の低下

NGSデータのサイズは数ギガバイトから数十ギガバイトのである。今日では一般に市販されているノート型PCでも大容量のものは16GB以上の場合の主記憶装置が搭載されている。従って、データを直接的に編集する

ことすら困難ではなくなっている。

また、NGSデータを処理するプログラムとしてオープンソースのものが多く配布されており、かつ、実際に主要な処理として用いられている。それらの多くはUNIX系統のOS上で動作し、その操作にはGUIではなくコマンドラインを用いるものが多い。プログラムの使用方法についての情報が書籍やインターネット上で掲載されているが⁵⁾、特にインターネットから場合、コマンドラインであればコピー＆ペーストにより簡単に実行させることができる。

加えて、医学生物学系の研究者がApple社製のパーソナルコンピュータを好むことは頻繁に言及されるが、現在、それらのオペレーティングシステム(OS)はUNIX系統であり、プログラムのインストールから実行についてのスキルの障壁は低いものとなっている。

NGSデータに対して改ざんする意思を持つものが十分な分子生物学的な知識を持つとは容易に想定できる。一方、そのような者が実際に改竄を実施しようとする場合に最初に障壁となると想定されるのが計算機関連の機材の調達やスキルの習得である。そして今日、そのハードルが低くなってきているということである。

更にはNGSシミュレータを悪用することにより高度なねつ造が可能になると考えられる。

NGSデータについてハッシュ値を埋込むことによって、埋込み以後の改竄を検知する技術は実現できるが、運用としてそれを研究者に行わせることは無意味である。なぜならハッシュと取除き、データを改竄し、再度ハッシュ値を埋込むことにより、改ざんの検出はNGSデータ単体ではできなくなってしまう。

う。

従って、NGSによってハッシュ値が埋込まれていなければならない。

2) 電子署名の必要性

NGSによりハッシュ値がNGSデータに埋込まれたとして、実際に改竄を検出するにはNGSは生成したハッシュ値を研究者が操作できない所に半永久的に格納しておかなければならなくなる。これは全く現実的ではない。

この問題を解決するには公開鍵暗号技術を用いた電子署名を作成し、それをNGSデータに埋込む必要がある。この場合、改ざんを検知するためには、電子署名を生成する際に用いた秘密鍵と対になる公開鍵を用いて電子署名を復号することになる。その為には公開鍵基盤(PKI)との連携が必要になる。日本においては医療関連機関向けにHPKIが運用されているが、NGSデータは国際的に流通するデータであるため一国の組織としてではなく国際的な組織としてPKIを運用する必要がある。またPKIシステムに対する情報処理および管理上の負荷は非常に大きくなるものと予想される。

E. 結論

NGSデータの証拠性を向上させるため、特に改竄を検知できるようにするためには、学術組織だけでなくNGSメーカーや解析のためのソフトウェアやサービスを提供する営利組織、国際的な学術協調組織にまでまたがった取り組みが必要であることが明らかになった。

参考文献

1. 榎木 英介, “生命科学の研究倫理

- なぜ不正が絶えないのか？”，KEIO SFC JOURNAL Vol.15 No.1 2015, pp.
3. Cryptography Second Edition: protocols, algorithms, and source code in C”, John Wiley & Sons Inc. 1996, p38.
 4. Rasko Leinonen, Hideaki Sugawara, Martin Shumway, “The Sequence Read Archive”, Nucleic Acids Research, 2011, Vol. 39, Database issue D19-D21.
 5. Merly Escalona, Sara Rocha, David Posada, “A comparison of tools for the simulation of genomic next-generation sequencing data”, NATURE REVIEWS GENETICS Vol. 17, pp.459-469.
 6. 清水厚志, 坊農秀雅, “細胞工学別冊 次世代シーケンサーDRY 解析教本”, 学研メディカル秀潤社, 2015.

F. 健康危険情報

平成 28 年度の本研究においては、生命、健康に重大な影響を及ぼすと考えられる新たな問題、情報は取り扱わなかった。

G. 研究発表

1. 論文発表

作佐部太也，大内雄矢，澤智博，渡辺浩，中島直樹，木村通男：証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用についての評価と提案第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，

340-362.

2. Bruce Schneier, “Applied 720-721, 2016

澤智博，渡辺浩，作佐部太也，中島直樹，木村通男：証拠性のある医学研究 次世代シーケンサー等のデータソースおよび解析ソフトウェアの検討第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，718-719, 2016

中島直樹，渡辺浩，澤智博，作佐部太也，宇山佳明，山口光峰，木村通男：証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討 第 36 回医療情報学連合大会，医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2)，714-717, 2016

木村通男，渡辺浩，澤智博，作佐部太也，中島直樹：証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化 第 36 回医療情報学連合大会抄録集 722-723, 2016

2. 学会発表

作佐部太也，大内雄矢，澤智博，渡辺浩，中島直樹，木村通男：証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用についての評価と提案第 36 回医療情報学連合大会, 2016 年 11 月 24 日, 横浜市

H . 知的財産権の出願・登録状況

(予定も含む)

- | | |
|-----------|----|
| 1. 特許取得 | なし |
| 2. 実用新案登録 | なし |
| 3. その他 | なし |

電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究
(H27-医療-指定-016)

- 研究用計測機器ユーザ認証等証拠性保全検討 -

研究分担者 渡辺 浩

国立長寿医療研究センター 治験・臨床研究推進センター 室長

研究要旨：国立長寿医療研究センターの治験臨床研究推進センターの医療情報室では研究者の研究支援基盤構築プロジェクトとして「電子カルテ領域の臨床データを匿名化モジュールを介して施設基盤レベルでデータを匿名化する仕組み」を試験的に導入している。今回の木村班の研究に関しても有用なソリューションであるためこれを報告する。

A. 研究目的

今回の班研究では、臨床研究用のデータを如何に研究者に安全に活用させるかコンセプトになっている

B. 研究方法

長寿医療研究センターは、病院と研究所から成り立っており、病院情報システムはクローズドなネットワーク、研究は外部との接続も可能な情報系のネットワークとして独立して存在している。2012年にバイオバンクシステムが導入された際には匿名化されつつ連結可能なデータソースのバンキングが必要になった。バイオバンクは研究機関内の部門であり、検体を管理する検体管理モジュールは、研究棟内の情報系ネットワークに存在しており、一方病院から送られた検体を処理するには電子カルテ上の患者属性、病名、検査結果、処方内容など、各種情報をサンプルのカatalogとして必要であったが、直接アクセスをさせることはポリシー上相応しくないと

考えた。そこで、病院系と情報系ネットワークの間にいずれにもアクセス可能な特殊なDMZネットワークを置き、ここに患者IDとバイオバンク協力者IDの変換や番号照合を行う匿名化システムを導入した。同時にこの機能を匿名化モジュールとして、本幹システムから分離独立させた。

倫理的配慮

今回の分担研究上、特に個人情報を取扱うなどの倫理的な課題は発生しなかった。

C. 研究結果

昨年度においては、本研究における電子カルテからのデータ抽出時に必要なデータ匿名化作業に際し、国立長寿医療研究センターで導入されている「電子カルテ領域の臨床データを匿名化モジュールを介して施設基盤レベルでデータを匿名化する仕組み」が有用であるとした。今年度はこの仕組みを実際に木村班研究で想定される仕組みに適應させた際の仕様などを検討

した。この仕組みには

1. 患者の ID を含んだ電子カルテ等診療情報システムネットワーク内のサーバ
2. DMZ（ネットワークの非武装地帯）領域内に患者 ID と匿名化 ID の対応表を管理するサーバ
3. システム利用ネットワーク内に置かれる基本サーバ

の 3 種類が必要とされる。今年度はまずはこれらのサーバの実際の基本設計を行う予定であったが、発展的研究を実際に開始後、現存のサーバ等の基盤環境、現存のデータ処理ソフトにて対応可能なことが判明し、購入予定していた部品、ソフトは不要となった。

D. 考察

今年度は 実装のための具体的な検討を行ったが今後はより詳細な仕様を決定する必要がある。外部の複数施設を跨いだ匿名化の場合、対象者が重複していることもあり得るためこれを解決する必要がある。現在マイナンバー（そのもので無く）を利用した医療福祉に関するユニークな国民番号発行も想定されているので、これらを用いた IHE の PIX・PDQ プロファイル等を利用した仕組みづくりが必要と思われる。

E . 結論

電子カルテデータの匿名化処理をインフラレベルで構築することには臨床研究推進の支援として期待が持てる。

F . 健康危険情報

今回の分担研究では、生命、健康に重大な影響を及ぼすと考えられる新たな問題、

情報は取り扱わなかった。

G . 研究発表

1 . 論文発表

澤智博, 渡辺浩, 作佐部太也, 中島直樹, 木村通男: 証拠性のある医学研究 次世代シーケンサー等のデータソースおよび 解析ソフトウェアの検討第 36 回医療情報学連合大会, 医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2), 718-719, 2016

中島直樹, 渡辺浩, 澤智博, 作佐部太也, 宇山佳明, 山口光峰, 木村通男: 証拠性のある医学研究 病院情報システムからの EDC データ源に関する検討 第 36 回医療情報学連合大会, 医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2), 714-717, 2016.

作佐部太也, 大内雄矢, 澤智博, 渡辺浩, 中島直樹, 木村通男: 証拠性のある医学研究 - 次世代シーケンサーからのデータの証拠性確保における暗号技術の利用についての評価と提案第 36 回医療情報学連合大会, 医療情報学 第 36 回医療情報学連合大会論文集 36(Suppl.2), 720-721, 2016

木村通男, 渡辺浩, 澤智博, 作佐部太也, 中島直樹: 証拠性のある医学研究 Web 型小病院向け電子カルテシステムを用いた研究ノートの電子化 第 36 回医療情報学連合大会抄録集 722-723, 2016

2. 学会発表

なし

H. 知的財産権の出願・登録状況

- | | |
|-----------|----|
| 1. 特許取得 | なし |
| 2. 実用新案登録 | なし |
| 3. その他 | なし |

研究成果の刊行に関する一覧表

雑 誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
<u>中島直樹</u> , <u>渡辺浩</u> , <u>澤智博</u> , <u>作佐部太也</u> , 宇山佳明, 山口光峰, <u>木村通男</u>	証拠性のある医学研究 病院情報システムからのEDCデータ源 に関する検討	医療情報学 第36回医療情 報学連合大会 論文集	36(Suppl.2)	714-717	2016
<u>澤智博</u> , <u>渡辺浩</u> , <u>作佐部太也</u> , <u>中島直樹</u> , <u>木村通男</u>	証拠性のある医学研究 次世代シーケンサー等のデータソース および 解析ソフトウェアの検討	医療情報学 第36回医療情 報学連合大会 論文集	36(Suppl.2)	718-719	2016
<u>作佐部太也</u> , 大内雄矢, <u>澤智博</u> , <u>渡辺浩</u> , <u>中島直樹</u> , <u>木村通男</u>	証拠性のある医学研究 － 次世代シーケンサーからのデータの 証拠性確保における暗号技術の利用に ついての評価と提案	医療情報学 第36回医療情 報学連合大会 論文集	36(Suppl.2)	720-721	2016
<u>木村通男</u> , <u>渡辺浩</u> , <u>澤智博</u> , <u>作佐部太也</u> , <u>中島直樹</u>	証拠性のある医学研究 Web型小病 院向け電子カルテシステムを用いた研究 ノートの電子化	医療情報学 第36回医療情 報学連合大会 論文集	36(Suppl.2)	722-723	2016
Nakata Y, Watanabe Y, Narimatsu H,Y oshimura T, O take H, <u>Sawa T</u>	Surgeons' efficiency change is a major determinant of their productivity change	International jou rnal of health care quality assurance 29(4)	29(4)	417-424.	2016
<u>澤智博</u>	クラウドシステム・仮想化技術は医療情 報システムを高度化させる鍵となるか	月刊新医療	43(7)	78-82	2016
中田善規, 成松宏人, 吉村達也, 大嶽浩司, <u>澤智博</u>	医療経済学を応用した手術室運営	麻酔	65(S)	S114-S121	2016
<u>澤智博</u>	人工知能による自動診断	Clinical Neuroscience	34(8)	928-31	2016