

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になりアリ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

・ 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。

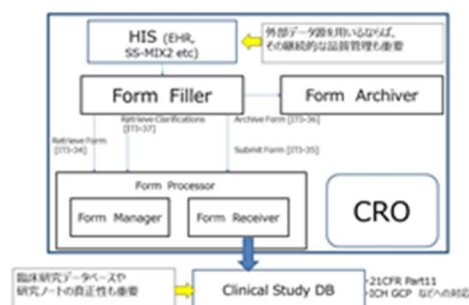


図 . IHE IT Infrastructure  
Retrieve Form for Data Capture(RFD)

図の 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。

#### d 研究用計測機器データの証拠性保全

##### 1) 改竄の可能性

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系システムであり、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めれば、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 |