

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