

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 23 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26670346

研究課題名(和文) CDISC標準対応症例データレポジトリシステムの構築と運用・評価

研究課題名(英文) Development and evaluation of a clinical data repository supporting CDISC-conformant dataset.

研究代表者

木内 貴弘 (Kiuchi, Takahiro)

東京大学・医学部附属病院・教授

研究者番号：10260481

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：個別症例生データをCDISC ODM形式で構築しXMLデータベースシステムに導入した。データはデータ表現にある程度の齟齬があっても取り込み可能であった。研究者がXQueryの技能を有せずとも簡単に検索可能なフロントエンドWebアプリケーションおよびフロントエンドアプリケーションを開発した。検索結果は抽出・保存し新たなデータとしてXMLデータベースシステムに再取込可能であった。これら成果を利活用することで研究者が特殊な技能を有せずとも簡単にメタアナリシスや再統計解析可能な環境を提供できる。

研究成果の概要(英文)：In this year, we built some example datasets of a clinical trial including case reports of individual subjects, in CDISC ODM format. Then we confirmed that we can import the example dataset into XML database which was deployed in last year. We tested some of variations in the method of expression for clinical data, and got successful results in much of cases. Then we developed a web interface and a desktop application for the XML database. With this application, clinical researchers can search complex data structure from the XML database, without experiences of XQuery XML query language. The search results can also be re-import into the XML database. As the result, clinical researchers can use the XML database as a reliable and easy-to-use network storage, which can be an infrastructure of reproducible research.

研究分野：医歯薬学

キーワード：CDISC 症例データレポジトリ

1. 研究開始当初の背景

過去の臨床・疫学研究の個別症例生データは、少数の共同研究者のみに共有され、1)個別症例生データを活用したメタアナリシス、2)第三者による再統計解析の検証は不可能であった。近年、臨床研究電子化の国際規格 CDISC(Clinical Data Interchange Standards Consortium)標準の確立によって、これらが「理論上は可能」となっている。しかし、匿名化した個別症例生データを CDISC 標準形式を用いて必要な研究者間で自由に共有するレポジトリの構築・運用はまだ行われていなかった。

2. 研究の目的

本研究の目的は、各分野の異なった研究グループの異なった臨床・疫学研究の個別症例生データをデータベースとして集積して、逐次データ抽出、メタアナリシス、統計解析に利用可能な CDISC 標準により臨床・疫学研究症例レポジトリの構築・運用評価を行うことであった。

従来は、数多くの実施されてきた過去の臨床・疫学研究の個別症例生データは少数の共同研究者のみが共有していた。また少数の共同研究者のみによる共有においても、個別研究毎に変数名・変数型がバラバラで、複数の研究をまたがった統計課軌跡は簡単ではなかった。これらのため、当該の研究者が統計解析した結果が公表されるのみであり、1)個別症例生データを活用したメタアナリシス、2)第三者による統計隔世の検証は困難であった。

本レポジトリは、研究者の利便性を考慮して、XML データベースネイティブだけではなく、Web によるユーザインターフェイスで提供されるものとし、必要に応じて医学研究者が無料で自由に症例データ登録、既登録データ検索ができる。

3. 研究の方法

(平成 26 年度)

運用テスト用の模擬症例データを作成し CDISC ODM(Operational Data Model)仕様に基づいて、これらの各臨床試験のサンプル XML データを作成する。XML データベースサーバーシステムを構築し作成したサンプルデータを XML データベースシステムサーバーに蓄積し XPath、XQuery を用いて、試験的に様々な検索を実施する。この試験検索結果に基づいて、一般利用者向けのフロントエンド Web アプリケーションのシステム仕様の策定を行う。このソフトウェアは、臨床研究の個別症例生データを、研究分野や患者背景情報に基づいて、検索結果を表示し、ダウンロード可能な小規模ソフトウェアである。

(平成 27 年度)

平成 26 年度に実施したフロントエンド Web アプリケーションの仕様策定を実施し、これに基づいてシステム開発を実施する。開

発したフロントエンド Web アプリケーションの運用・評価を実施しその結果をもとにフロントエンド Web アプリケーションの修正と再評価を実施してアプリケーションの向上を図る。

4. 研究成果

個別症例生データを作成しそのデータを CDISC ODM 形式で構築した。CDISC ODM 形式は CDISC 標準群のうち特に臨床データの交換・蓄積・保存のための標準で臨床データの取り扱いは XML 形式を行うように仕様が策定されている。UMIN センターでは既に UMIN がサービスを提供している医師主導型臨床試験の支援システムである INDICE において医療機関に保存されている電子カルテ等よりの臨床個別症例データの電子的送付の機能提供において既に対応しているデータ形式である。CDISC 標準群のうち臨床データのデータ交換にはこの CDISC ODM 形式か CDISC ODM をベースとした DataSet-XML 形式になるが、DataSet-XML 形式は臨床データ交換でも特に新薬承認申請での利用を意識した設計となっており承認申請時以外においては DataSet-XML に比較して汎用的が必要となる今回の個別症例データの取り扱いとして CDISC ODM 形式が適していると考えた。

XML データベースサーバーシステムとして BaseX を用いたサーバーシステムを構築した。このサーバーシステムに前述の CDISC ODM 形式の個別症例生データを蓄積した。BaseX が持っている GUI モードを利用しその中で XQuery 等を用いて様々な試験検索を実施した。それらの検索の実施結果を用いてフロントエンド Web アプリケーションの仕様を策定した。その XQuery を用いた様々な検索をした結果は、XQuery は強力で多様な検索ができ標準規格化済みではあるもののその多様な検索を実施するには XQuery の文法と記述を習得する必要がありメタアナリシスが目的の医学研究者に検索の目的にその技能の習得を課すのは大きな負担になると思われた。

そこで、フロントエンド Web アプリケーションの仕様設計方針として XQuery の記述をフロントエンド Web アプリケーション内に隠蔽化し利用者である医学研究者が XQuery のコマンドや記述を意識しなくても利用可能なものにした。このフロントエンド Web アプリケーションを運用・評価したところ狙い通りに利用者が XQuery を全く意識することなく検索キーワードを指定し検索ボタンを押下するだけの簡単な方法で複数の試験を跨った検索が実現できたことより目標は達せられたと考える。

(開発したフロントエンド Web アプリケーション)



(検索結果例)



(XQuery 構文例)

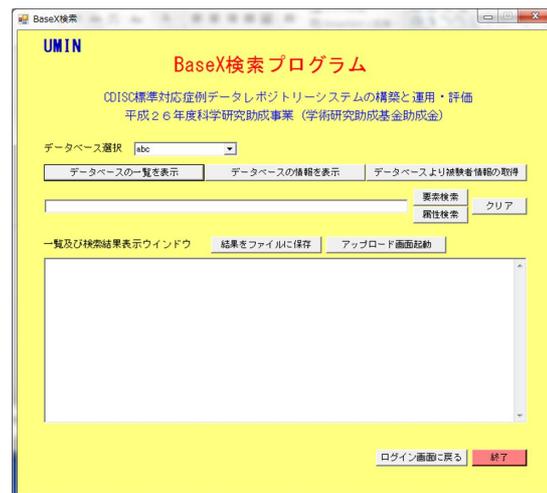
```
<response>{
  for $result in
  db:open($dbname)//*[text() contains text
  {".*" || $term || ".*"} using wildcards]
  return $result
}</response>
```

研究計画時点には無かった追加の成果として利用者およびサーバ側の多様なシステム環境に対応する目的で前述のフロントエンド Web アプリケーションに加えてフロントエンドアプリケーション(非 Web アプリケーション)も開発した。これはフロントエンド Web アプリケーションと同等以上の機能を

実装し提供するが Web アプリケーションではない通常の実行可能形式のアプリケーションとした。

本研究においてフロントエンド Web アプリケーションを開発し、様々な検索を実施し運用評価を行った結果フロントエンド Web アプリケーションでの提供形態ではその言語環境設計においてセキュリティ上の考え方からサーバーリソースへのアクセスが通常のアプリケーションに比較して意図的に制限されている。勿論、本研究計画立案時にもその点は考慮しており、様々であっても単純な検索などでは全く問題ないがサーバ側も多様な環境、例えば医療機関ごとにサーバが運用されているような場合などを想定するとフロントエンド Web アプリケーション形態ではそれらの環境対応性に限界が発生する可能性も考えられた。また利用者を IT スキルが高くない医学研究者を想定したことから簡便に利用可能なフロントエンド Web アプリケーション形態を候補にあげ検討したが通常のフロントエンドアプリケーションでもインストーラーを作成することによりそれ程煩雑でないシステムを構築することができた。インストール自体は1度のみ実行すればよいだけであるからインストール後のソフトウェア起動においては逆にフロントエンド Web アプリケーションより簡単になった。

(開発したフロントエンドアプリケーション)



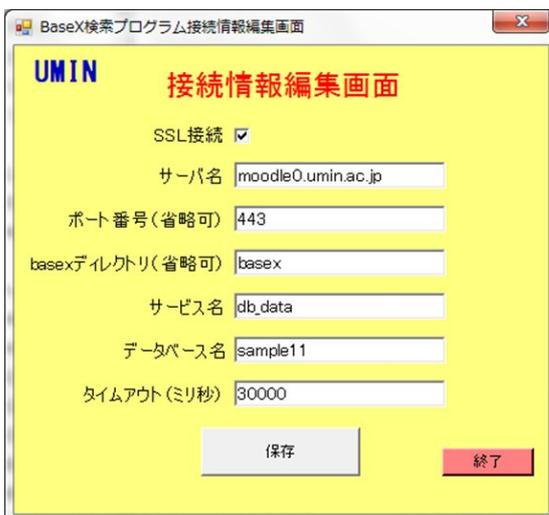
(検索結果例)



フロントエンド Web アプリケーションは Web ブラウザのみで動作し軽量なためダウンロードも容易でインストール不要であるから利用者がより簡便に利用可能となる。

一方フロントエンドアプリケーション(非 Web アプリケーション)はマイクロソフト社製 WindowsPC でのみ動作しサーバからダウンロード後 PC にインストール作業が必要であり利用に際してはフロントエンド Web アプリケーションよりも若干手間を要するが本システムの内部設定変更作業を実施することにより検索対象のサーバやデータベースを柔軟に切り替えたりサーバへの通信の暗号化のオン/オフ、その他通信パラメータの設定も可能な設計としたので将来のより高度な環境への対応が可能となっている。

(環境設定画面)



コンピュータ言語や PC 環境的にも一般的に販売されているアプリケーションと同等

のものを採用しているため、フロントエンド Web アプリケーションに比べてより深く高機能化が可能であると共に、セキュリティ的にも強固な環境を実現・提供可能と考えられる。利用者の簡便性優先の場合はフロントエンド Web アプリケーションの方が対応性が高く、逆に若干簡便性は低下するが強固で高機能化が優先される場合はフロントエンドアプリケーション(非 Web アプリケーション)が有用と考えられた。

本フロントエンドアプリケーションの開発においては過去に実施した科学研究助成事業の挑戦的萌芽研究「CDISC 標準を活用した死体検案書の施設別及び全国集計データベースの構築」において開発した法医学データベースシステムのクライアントシステムの開発ノウハウが役に立った。

本研究で構築したシステムを利活用することで例えば各医療機関ごとにデータ表現形式にある程度の齟齬が含有されるようなデータであっても取り込み可能であって利用者はそれら複数試験を検索対象とし指定しても簡単な操作で検索が可能なが確認された。

検索結果は抽出・保全することも可能でその抽出し保存したデータを新たな XML データベースのデータとして利用することも可能であった。つまり元の多数の症例データから条件に合致した症例データのみを抽出しその結果を別のデータベースとして簡単に構築できることになり、新データベースは元データベースのサブセットとして元のデータには影響を与えない独立したデータベースとして操作可能になる。これにより疾患や年齢層、既往歴の条件により様々なデータベースを簡単に生成可能になった。独立したデータベースとなれば抽出生成された新データベース側を変更や更新処理を実施しても独立性が保たれる。また複数の構造が異なる元データに対して同一の条件合致により抽出した複数のそれぞれ構造が異なる新たなデータを新データベースに予め構造定義作業不要にて取り込み可能になるので一般的なリレーショナルデータベースに比較して簡単に特定条件に合致した統合データベースを構築できることになる。それは分野や構造が異なるデータに対してのメタアナリシスが可能になったことを意味する。

これら成果を踏まえると利用者は XQuery 等の技能を習得せずとも簡単にメタアナリシスや再統計解析が可能な環境が提供できた。

引き続き今後は実運用に向けてまず小規模な実際の個別症例生データの蓄積・検索から運用を開始し、将来は大規模データの取り扱い方法、それらデータのアップロードや検索結果の見易さや取り扱いの方法、検索パフォーマンスを改良しつつ、サーバの保守運営方法を研究設計し実運用へのノウハウを蓄積してゆく方法が良いだろうと考えた。

5. 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

木内貴弘、CDISC 標準の概要と実装の留意点、PHARMSTAGE、査読無、5 巻、2014、4-8

〔学会発表〕(計 4 件)

木内貴弘、Thial development of Clinical and epidemiological data repository system using CDISC ODM.CDISC Interchange Noth America.2014 Bethesda(USA)

大津洋、医療・臨床研究領域での CDISC 標準、SAS ユーザー総会、2014 東京大学伊藤国際学術研究センター(東京・文京区)

木内貴弘、The Development of Clinical Research Date Using CDISC ODM.2015CDISC International Interchange.2015 Chicago(USA)

大津洋、臨床研究における ICT 標準化の取り組みと現状・将来像について 第7回日本臨床試験学会学術総会 名古屋国際会議場(名古屋・名古屋市)

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1)研究代表者

木内 貴弘(TAKAHIRO KIUCHI)
東京大学・医学部附属病院・教授

研究者番号：10260481

(2)研究分担者

大津 洋(HIROSHI OHTSU)
国立国際医療研究センター・医療情報解析
研究部・臨床疫学研究室・室長
研究者番号：40372388

(3)研究分担者

岡田昌史(MASAFUMI OKADA)
東京大学・医学部附属病院・講師
研究者番号：70375492

(4)連携研究者

石川ひろの(HIRONO ISHIKAWA)
東京大学・医学部附属病院・准教授
研究者番号：40384846