

登録データベースシステムの設計開発及びデータ精度向上に関する検討

研究分担者 盛一 享徳（国立成育医療研究センター小児慢性特定疾病情報室 室長）

研究要旨

【目的】 現行システムにて登録されたデータの再利用に係る課題について検討した。

【背景】 2022年度中に指定難病患者データベースと小児慢性特定疾病児童等データベースが統合され登録システムが刷新される予定であるが、新システムでは現行登録システムの登録データを処理できないことから、現行登録システムが停止後も登録データの利活用を可能とするためのリポジトリデータベースの構築が急務となった。

【結果】 現行登録データは、階層構造をもつキー・バリュー型データとして登録されている。中間データであるJSON形式データを利用し、それを取り込むことで可能な限り、登録データの階層構造を維持しつつ、登録データの二次利用のための任意のデータ抽出にも対応できる実装が可能となると考えられた。将来的には、技術的に最適となる階層構造を持つオブジェクトを定義してのデータ保持が望まれた。

【考察】 実現可能性と将来性を図りつつ、具体的な実装に向けて取り組んでゆきたい。

A. 研究目的

2015（平成27）年に改正法が施行された児童福祉法（昭和22年法律第164号）に基づく「小児慢性特定疾病その他の疾病にかかっていることにより長期にわたり療養を必要とする児童等の健全な育成に係る施策の推進を図るための基本的な方針」（平成27年厚生労働省告示第431号）において、国は、

- ・ 小児慢性特定疾病児童等についての臨床データ（以下「小児慢性特定疾病児童等データ」）を収集し、小児慢性特定疾病児童等に係る医学的データベースを構築すること
- ・ 小児慢性特定疾病児童等データベースの構築及び運用に当たっては、個人情報の保護等に万全を期すこと
- ・ 小児慢性特定疾病に関する調査及び研究に

有効活用できる体制に整備するとともに、指定難病患者データベースと連携することとされた。これに基づき、小児慢性特定疾病児童等データは、2015（平成27）年以降の医療意見書データは、新しく構築されたシステムにて収集・蓄積されてきた。2020（令和2）年より研究利用に限り、小児慢性特定疾病児童等データの提供が開始され、厚生労働省の審査会により使用申請が認められた難病研究班等へ適宜データ提供されている。

これまで別々のシステムとして運用されてきた指定難病患者データベースと小児慢性特定疾病児童等データベースは、2022（令和4）年度中にシステムが刷新され一つに統合される予定であるが、当初は現行システムの登録データを新システムが取り込み、現行システムにて記録されたデータも新システムから提供

される予定であったが、新システムと現行システムとの構造が全く異なることから、新システムが現行登録データ取り込むことは、開発途中で断念された。現行システムは 2022 年度末には停止されることから、これまでの登録データの提供が頓挫する可能性がでてきた。

本研究は、①現行システムの登録データを取り込みデータ提供が実施できる、②登録システムが出力したデータを取り込み再構成することで、登録システムの世代を越えてもシームレスにデータ提供が行えるリポジトリデータベースの構築について検討を行った。

B. 研究方法

2022 (令和 4) 年 3 月末現在、小児慢性特定疾病登録データベースは、公称告示疾病 788 (包括的病名を含めた実質の疾病数は 848 疾病) に達し、それぞれの疾病について新規申請用と継続申請用の二種類の医療意見書が存在している。また成長ホルモン治療申請用医療意見書が別途 12 種類用意されている状態となっている。さらに対象疾病の追加等により同一疾病であっても医療意見書の版の更新があり、現在 5 つの異なるバージョンが存在している。

医療意見書は、大きく基本情報と疾病情報部分に分かれており、疾病情報はさらに臨床所見、検査所見、合併症、経過 (治療) のドメインを持ち、内部の項目は概念ごとにグループ化され、階層構造を持っている。各項目は同一の内容であれば疾患を越えて横断的に統一されており、理論上は疾患別の集計だけでなく、疾患横断的な項目別の集計も可能となっている。各項目は項目名 (キー名) と値 (バリュー) の対になっているキー・バリュー型として記録されている。

医療情報は頻繁にマスタが更新され、項目は医学的概念ごとにまとまって処理されることが望まれることから、実装の前に固定的なスキーマ (項目の二次元設計図) を必須とするリレーショナルデータベース (RDB) の構造は、医療情報を処理する上で最適とは言えない構造であると言われている。このため現行の小児慢性特定疾病児童等データベースは、キー・バ

リュー型でかつ階層構造をもつデータをそのまま保持するために、non-RDB と呼ばれるデータベースに最終的に格納される設計となっている。

一方、新しい登録データベースは、開発の用意さから RDB をベースとする設計に戻っているが、頻繁なマスタ更新に対応するため、通常のテーブル構造とは異なる縦持ちテーブルが採用された。一般的な横持ちテーブルは、一つのレコードについて、全ての項目が横一列に並び全体像が分かりやすい反面、項目追加などの構造変化に劇的に弱いという欠点がある。一方、縦持ちテーブルの場合は、一つのレコードに一つの項目 (一つのキーとバリューの組み合わせ) しか持たないため、項目が自由に追加できる利点がある。しかし項目の全体像がテーブルの中には存在しないことになり、横持ちテーブルの際の 1 レコードを再現 (ビューを作る) するために、大量の処理が必要となる、という欠点もある。今回はシミュレーションの結果、データ規模から縦持ちテーブルで実装したとしても、現実的な時間内で処理できると判断され、実装されることとなった。

現行データは、最終的には XML 形式で non-RDB に記録されるが、JSON 形式の中間ファイルを作成しており、この中間ファイルから記録データの全出力が予定されている。しかし、JSON 形式データを縦持ちテーブルの新システムへの取り込みが断念されたことから、①階層構造を維持したデータ保持は可能か、②既存の JSON 形式の登録データを取り込み、データ抽出依頼に対応できるよう任意の抽出を実現するにはどのような実装が望ましいか、③一定期間が経過後、登録システムの刷新が行われる可能性が高く、今回と同様の事態が定期的に生じることが予想されることから、登録システムの構造によらず、異なる登録システムから登録データを取得・蓄積し、登録システムの世代を越えてシームレスなデータ抽出を行える実装が可能かどうか、④国レベルで医療情報伝達には、FHIR 規格が利用されつつあることから、将来的な FHIR 規格に準拠した出力に対応できるよう

な実装は可能かどうか、について検討を行った。

(倫理面の配慮)

本研究は、個人を特定しない内容を用いて実施しており、特別な倫理的配慮は必要ないものと判断した。

C. 研究結果

登録システム側が RDB であっても non-RDB であっても対応できるようにするためには、受け手側は non-RDB もしくは JSON 形式のような階層構造をハンドルできる DB であることが望ましいと考えられた。現行の小児慢性特定疾病児童等データベースの階層構造を技術的に最適な形で再現する場合、データベース構造を階層構造を持つオブジェクトとして定義することになるが、既存の構造データのデータ量が極めて多いことから、短期間での対応は難しいと考えられた。しかし将来的に FHIR 規格につなげることを考えた場合は、データベースの階層構造化に進むことが望ましいと思われた。

データ利用申請に対応する、という直近の課

題に対応するには、JSON 形式のデータをそのまま取り込める環境を整備し、JSON 形式のデータから、データ抽出に対するリクエストに対応するのが当面現実的であると判断された。結果として、将来的な機能と実装可能な機能を考え、利用する DB としては、InterSystems IRIS for Health および Postgres SQL を併用するのがよいと考えられた。

D. 考察

システム更新に伴う登録済データの処理に対する課題について検討を行った。工期が限られていることから、現実的な解決策を実施しつつ、将来的なデータ利活用の恒常性も視野に入れて引き続き対応することとした。

E. 研究発表

なし。

F. 知的財産権の出願・登録状況(予定を含む。)

特許取得/実用新案登録/その他
なし/なし/なし

