

小児慢性特定疾病児童等登録データの提供システム等に関する検討

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】 現行システムが出力した登録データを取り込み、階層構造を再現するとともに登録データを保持し、再出力ができるリポジトリデータベースのプロトタイプの開発を目的とした。

【方法・結果】 現行システムが疾患ごとに出力する XML 形式の階層構造ファイルをインポートし、階層構造を再現した。tsv 形式の登録データ出力ファイルをインポートし、階層構造を持つデータとして疾病情報を保持した。二次元配列の tsv 形式データとして、再度疾病情報を出力できることを確認した。

【結論】 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。本研究では、XML 形式の階層構造ファイルから階層構造を再現し、json 形式のキー・バリュー型データとして保持できることが確認された。また二次元配列 tsv 形式ファイルとして登録データを出力できることが確認された。これにより現行登録データをデータ提供する機能が実装できたことが確認された。今後は実データの取り込みと出力確認、登録データの追加・編集機能の実装を目指したい。

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およ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および研究

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、システムが刷新され一つに統合される予定であり、小児慢性特定疾病側の登録システムは、2023（令和 5）年中の稼動開始が見込まれている。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断念された。現行システムは、新システム稼働後は運用が停止されることから、これまでの登録データの提供に影響を及ぼす可能性がでてきた。

本研究は、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、また将来的には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超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開発を目的とした。

B. 研究方法

2023（令和 5）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。また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、項目数や値が登録開始後に追加・変更される可能性が高く、固定長データではなく可変長データとしての扱いが望ましいこと等から、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持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、開発の容易さから 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。次期システムではシミュレーションの結果、データ規模から縦持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。しかしデータベース自体にはデータの階層構造や関連付けを表現できないことから、別途、データ項目整理票と呼ばれる項目間のグルーピングや階層構造を定義する巨大な構造ファイルが必要となっている。

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、現行登録データの出力した JSON 形式データのシステムへの取り込みが断念されたことから、現行システム停止後のデータ提供に対応するための新たなリポジトリデータベースの開発が急務となった。今回のように登録システムは、周期的に入れ替えが生じる可能性があるため、

登録システムの世代を超えた登録データのシームレスな保持が望まれることとなった。リポジトリデータベースの開発に当たっては、

- 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か
- ③ 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超えてシームレスなデータ抽出を行える実装が可能か

について検討した。本年度は①の実装に向けてプロトタイプを作成を行い、試験データを取り込むことで実データの取得に向けた準備を行った。

(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

現行システムは、現在までに大きく 5 回データ構造マスタのバージョンが更新されている。データ構造マスタの更新は、新規追加疾病に併せて実施される。またマスタ更新が行われた際に、項目の追加や修正が行われることもある。このため異なるマスタ間では項目の不一致が生じる場合がある。とくに 2018 年に新規申請用と継続申請用に異なる様式が用意され、2 疾患群が追加されたことから、2015 年のバージョンと 2018 年以降のバージョンでは、マスタの構造は大きく異なっている。

小児慢性特定疾病は、疾患群の下に区分という分類があり、その中に対象疾病が並んでいる階層構造を持っている。新規に疾病が追加される際には、既存の区分の中もしくは新たな区分が新設され、その中に追加される。新規追加疾病は既対象疾病の並びの間に挿入されることが多いため、追加疾病が生じる度に、しばしば既対象疾病を含めて告示番号がずれることが多い。データ構造マスタの更新では、告示番号の変更も追隨している。

現行システムでは、データ構造はデータ項目整理票と呼ばれるデータ定義が記載された excel 形式ファイルを登録システムの機能により階層構造を持った XML 形式ファイル(図 1)に変換して実装されている。この XML 形式ファイルから、登録システムのデータ入力を行うためのブラウザ入力画面 (HTML 形式) が定義されると同時に、non-RDB データベースの構造定義にも利用される。本研究では、この XML 形式のファイルから疾病ごとの階層構造を取得することとした。

また現行システムから二次利用データ用に出力される登録データのファイルは、1 レコードが 1 意見書に対応するタブ区切りのテキスト (tsv 形式) ファイルとなっている(図 2)。本研究では、この tsv 形式のファイルから 1 意見書ごとのレコードを取得することとした。データ構造マスタが異なると、項目定義が異なる場合があることから、データ構造マスタのバージョンごとに同一疾病で最大 5 種類の tsv 形式ファイルが出力されることに対応した。

この他に現行システムでは、疾患群マスタ (疾患群のリストに相当)、疾患カテゴリマスタ (対象疾病の区分のリストに相当)、疾患マスタ (対象疾病のリストに相当)、項目名称マスタ (ID 化されている項目に対応する項目名のリスト) などの管理用マスタが存在するため、それらも取得した(図 3)。各テーブルとその関連図を図 4 に示す。

基本情報部分は階層構造なしの JSON 形式データとして保持し(表 1)、疾病情報に当たる部分は、XML 形式の階層構造ファイルから取得した構造を階層あり JSON 形式として保持した(表 2)。

以上を実装したシステム概要を図 5 に示した。

D. 考察

現行システムは、医療情報の特殊性が考慮された上で、階層構造を持つ key-value 型データ形式によるデータ保持が最適であると判断され実装された。階層構造情報と項目情報が定義されているデータ項目整理票と呼ばれる定義

書から、XML 形式の階層構造を中間ファイルとして作成し、ここから入力用画面、登録データベースの構造、紙の医療意見書レイアウトが一元的に作成できる仕組みとなっている。

本研究では、出力された XML 形式ファイルをインポートし、データ構造を取得できることを確認できた。また登録データ二次利用およびシステムデータのバックアップとして出力される二次元配列の tsv 形式ファイルの疾病データをインポートし、取り込んだデータ構造と合わせて、JSON 形式データとして保持できることが確認できた。また JSON 形式データとして保持した疾病データを現行システムと同様に、二次元配列の tsv 形式ファイルとして、出力できることも確認された。以上から、今回のプロトタイプでは予定どおりの実装が行えたと判断され、現行システムと同様のデータ提供への道筋ができたと考えられた。

E. 結論

現行システムの出力データを構造ファイルとともに取得することで、階層構造の再現と出力データの保持を行うことができた。今後は 2023（令和 5）年度中に、現行システムが未だ稼動している段階で、実データの取り込みおよび再出力を行い、現行システムの出力結果と同様のデータ出力が可能であることを確認するとともに、現行システム停止後のデータ追加・編集も念頭において、登録データの編集機能の実装を目指す予定である。

F. 研究発表

論文発表/学会発表
なし/なし

G. 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

1) データフォーマット (xml形式)

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE disease_def [
<ELEMENT disease_def (category)*>
<ELEMENT category (category)*>
<!ATTLIST category name CDATA #REQUIRED
item_id ID #REQUIRED
iso CDATA #REQUIRED
if_child CDATA #REQUIRED
group_id CDATA #REQUIRED
pre_disp CDATA #REQUIRED
post_disp.lst CDATA #REQUIRED
input_type CDATA #REQUIRED
add_attr CDATA #REQUIRED
type (文字列|数値|列挙|年月|年月日) #REQUIRED
dimension (単一|複数) #REQUIRED
required CDATA #REQUIRED
readonly CDATA #REQUIRED
link_item_id CDATA #REQUIRED
input_cnd CDATA #REQUIRED
]>
]
disease_def id="D10_001N-2019a" name="10 X連鎖重症複合免疫不全症【初回申請用】"
<category name="基本情報" iso="basic_information_tab">
<category name="基本情報" iso="basic_information_section">
<category name="受診日" item_id="00_524_002" iso="recipient_date" input_type="ymd" add_attr="ng-init=&quot;data[&apos;00_524_002&apos;]=Date()&quot;" type="年月日" grp_id="00_524" />
<category name="受給者番号" item_id="00_002_001" iso="recipient_number" input_type="text" add_attr="ime-off" type="文字列;20" grp_id="00_002">
<category name="受付種別" item_id="00_002_002" iso="recipient_number_type" input_type="radio" add_attr="ng-init=&quot;data[&apos;00_002_002&apos;]=&quot;" type="列挙;新規登録" />
</category>
<category name="氏名" iso="basic_information_name_category" grp_id="00_003">
<category name="漢字" item_id="00_003_001" iso="basic_information_name_family_name_kanji" input_type="text" add_attr="" type="文字列;12" />
<category name="漢字" item_id="00_003_002" iso="basic_information_name_last_name_kanji" input_type="text" add_attr="" type="文字列;30" />
<category name="ふりがな" item_id="00_003_003" iso="basic_information_name_family_name_kana" input_type="text" add_attr="" type="文字列;20" />
.
.
<category name="小児慢性特定疾病指定番号" item_id="00_999_007" iso="signature_signature_number" input_type="text" add_attr="" type="文字列;20" />
<category name="記載年月日" item_id="00_524_001" iso="signature_signature_date" input_type="ymd" add_attr="" type="年月日" />
</category>
</category>
</disease_def>
```

図 1. 疾病情報部分の階層構造データ

1) データフォーマット (tsv形式)

固定項目														以降、疾病群に含まれる疾病の最大Item_idを列挙					
カラム部	sheetId	versionNumber	groupNumber	category	diseaseNumber	diseaseName	mode	status	patientClusterId	jichitaiName	newUpdated	obesityDegree	sd0fHeight	00_002	00_002_001	00_002_002
データ部	1111-1111-0000-0000	2019a	6	1	1	若年性特発性関節炎	継続	初期登録データ入力完了	111111	東京都	継続			11111111	3	
データ部	1111-2222-0000-0000	2019a	6	4	14	混合性結合組織病	継続	名寄せ確認中	222222	埼玉県	継続			22222222	3	

図 2. 出力データサンプル (tsv) 形式

1) データフォーマット (tsv形式)

```
a. 疾患群マスタ (ファイル名: t_group.tsv)
group_id group_number group_name group_name_e group_name_k is_disabled registered_at regist_user_id updated_at update_user_id version_number normalized_name
2015010001 1 悪性新生物群 0 2015a
2015010002 2 慢性腎疾患 0 2015a

b. 疾患カテゴリマスタ (ファイル名: t_category.tsv)
category_id group_id category_number category_name category_name_hlw category_name_e category_name_k is_disabled registered_at regist_user_id updated_at update_user_id version_number normalized_name
2015010001 2015010001 1 白血病 白血病 0 2015a
2015010002 2015010001 2 骨髄異形性症候群 骨髄異形性症候群 0 2015a

c. 疾患マスタ (ファイル名: t_disease.tsv)
disease_id category_id disease_number disease_name disease_name_hlw disease_name_hlw_e disease_name_k is_disabled registered_at regist_user_id updated_at update_user_id version_number normalized_name
2015010001 2015010001 1 80 前駆胚細胞急性リンパ性白血病 前駆胚細胞急性リンパ性白血病 0-cell precursor lymphoblastic leukaemia ぜんくびーさいぼうきゅうせいりんぱせいけつびょう 0 2015a
2015010002 2015010001 2 77 成熟胚細胞急性リンパ性白血病 成熟胚細胞急性リンパ性白血病 Mature B-cell lymphoblastic leukaemia せいしゅくびーさいぼうきゅうせいりんぱせいけつびょう 0 2015a

2019010043 2019010282 33 10 MECP2重複症候群 MECP2重複症候群 MECP2 duplication syndrome えむいーれーひーつーちようふくしゅうこうぐん 0 2019a
2019010044 2019010282 34 23 武内-小崎症候群 武内-小崎症候群 Takenouchi-Kozaki syndrome たけのうち-こさきしゅうこうぐん 0 2019a
```

図 3. 管理用マスタ例 (tsv 形式)

1. テーブル一覧

NO	テーブル名	説明
1	取込データ	登録データ/医師意見書データ
2	整理票データ(メイン)	整理票の項目ID、項目名称の構造データ(xmlファイルの取込内容)
3	整理票データ(サブ)	登録・編集画面を構成する際の整理票データをもとにした設定データ(xmlファイルの取込内容)
4	疾患群マスタ	整理票バージョンごとの疾患群データ
5	疾患マスタ	整理票バージョン、疾患群ごとの疾患データ

2. テーブル関連図

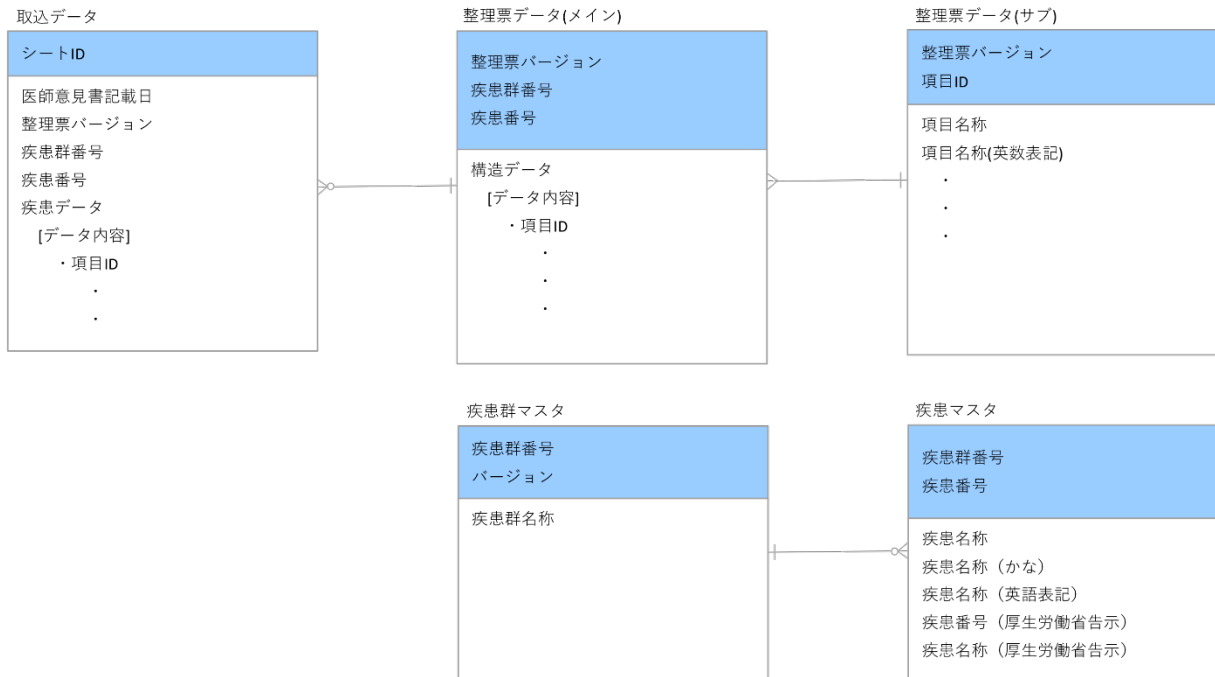


図 4. テーブル一覧およびテーブル関連図

表 1. 登録データイメージ (基本情報)

3. 登録データイメージ(マスタは省略)

データ項目	登録内容 例)	備考
シートID	1111-2222-0000-0000	先頭9桁(9999-9999)がスタンプ番号
医師意見書記載日	2020/10/01	
整理票バージョン	2019a	
疾患群番号	6	
疾患番号	1	
疾患データ	[<pre> "sheetId": "1111-2222-0000-0000", "versionNumber": "2019a", "groupNumber": 6, "categoryNumber": 1, "diseaseNumber": 1, "diseaseName": "若年性特発性関節炎", "mode": "継続", "status": "初期登録データ入力完了", "patientClusterId": 111111, "jichitaiName": "東京都", "newUpdated": "継続", "obesityDegree": -0.7, "saOfHeight": -2, "00_002": "1111111", "00_002_001": "1111111", "00_002_002": "3",]</pre>	json形式(階層なし)

表 2. 登録データイメージ (疾病情報)

3. 登録データイメージ (マスタは省略)

データ項目	登録内容 (例)	備考
整理票バージョン	2019a	
疾患群番号	10	
疾患番号	1	
申請種別	N	
構造データ	<pre>[{ "iso": "basic_information_tab", "name": "基本情報", "sections": [{ "iso": "basic_information_section", "name": "基本情報", "categories": [{ "iso": "recipient_date_category", "name": "受診日", "items": [{ "\$type": 1, "value": "20190328", "items": [], "itemId": "00_524_002" }] }], "categoryId": "00_524" }] }, { "iso": "recipient_number_category", "name": "受給者番号", "items": [{ "\$type": 1, "value": "0503144" }] }] </pre>	json形式(階層あり)

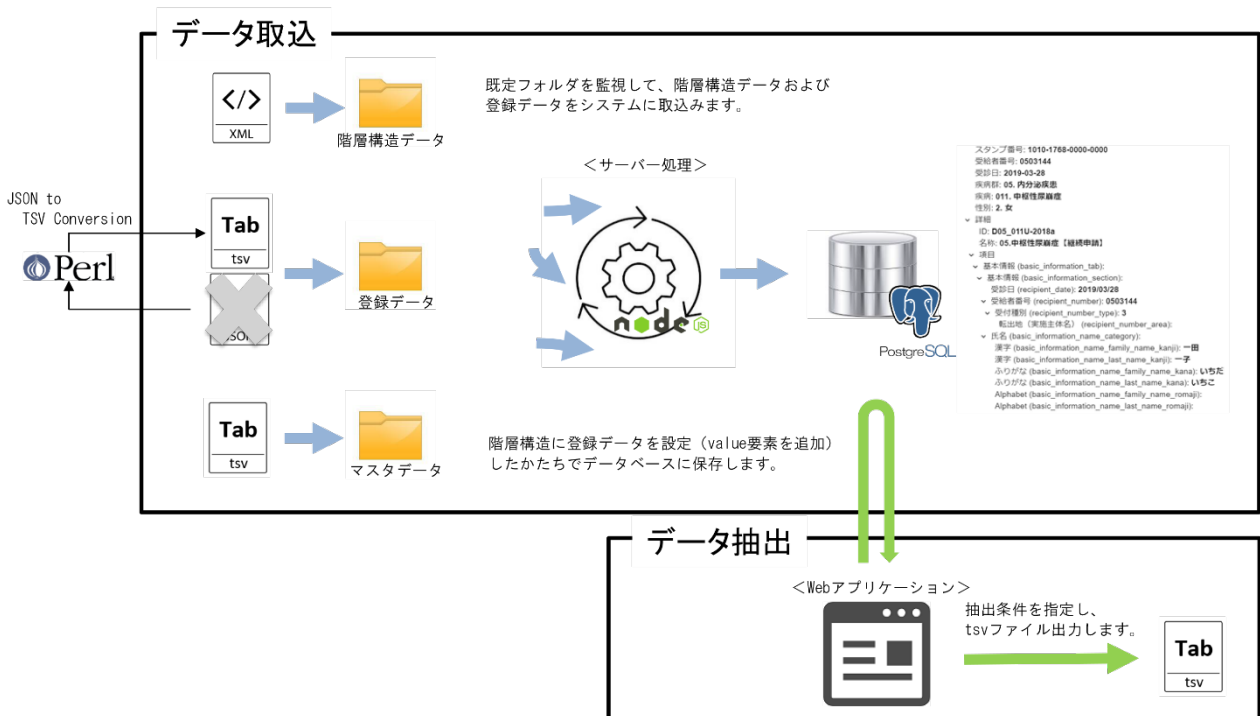


図 5. システム概要

