ISS-A-010

医療機関名称をキーとした医療オープンデータ結合方式の検討

荒木　舞那†　　松尾　仁司†
† 高崎健康福祉大学　大学院　健康福祉学研究科　医療福祉情報学専攻

1. はじめに

「情報を利活用した社会」を実現するため，政府や行政が収集している公的統計データを広く活用できるよう，オープンデータの公開が推進されている．しかし，全国レベルで記載が統一され，ID が付与された医療機関のリストは公開されていない．保健医療分野での公的統計は，根拠となる法律ごとに独立したオープンデータとして公開されているため，データ間の連携に関しては考慮されておらず，医療機関名称や ID も統一されていないという課題があった．複数のオープンデータを統合して利活用するためには，名称表記のゆれや粒度を調節し，全国共通の医療機関コードを各オープンデータに付与する必要がある．

2. 全国保険医療機関データの作成

厚生労働省が管理している都道府県ごとの保険医療機関指定のコード内容別医療機関一覧表（以下，医療機関データ）を基に全国保険医療機関データを作成した．47 都道府県別に作成された医療機関データには表 1 のような区切り文字の違いや記載ミスなどによる表記のゆれが見られた．

表1. 表記ゆれの例

項目	表記のゆれ
医療機関コード	″-″　″,″　″・″　″.″　″″　″　″の区切り記号
法人名	(医)(社)などの省略系の有無
住所	都道府県表記の有無
診療科目	1614 種の表記の違い

表記ゆれを吸収して，全国統一表記のリストを作成するプログラムを開発し，94,268 の病院・診療所(医科)を含む医療機関リストを完成させることができた．先頭 2 桁に都道府県コードを追加し，全国保険医療機関データの医療機関コードとした．全国保険医療機関データには医療機関コード，医療機関名，市町村コード，開設者，医師数等全 20 項目のデータが含まれる．この医療機関コードを他のオープンデータに付与することで統一したコードを持つデータとの結合が可能となる．

3. オープンデータの結合方法

全国保険医療機関データをマスターとし，病院を対象として各オープンデータに ID を付与する方法を検討した．今回，オープンデータは病院の機能や規模が分かる病床機能報告[1]と医療提供状況が分かる DPC 調査[2]を用いた．まず DPC 調査データ,病床機能報告データを対象に全国保険医療機関データの医療機関名称をキーとして突合を行った.単純に名称で突合した場合，表記のゆれで付与できない医療機関が多くあった．また，同一都道府県に同一名称が複数存在する場合は名称だけでは特定ができない．そのため，オープンデータの医療機関名称に前処理を施して，全国保険医療機関データの名称との完全一致，部分一致，不一致を判断し，同一都道府県に複数候補がある場合は，市区町村で特定するプログラムを開発した．図 1 にプログラムの処理の流れを示す．

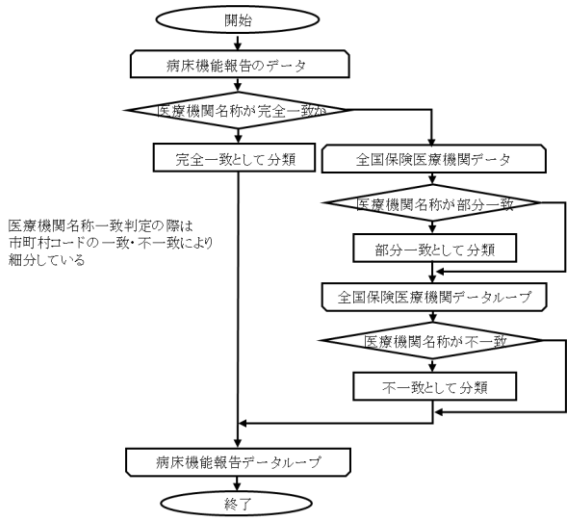

図1. 医療機関名称をキーとした突合処理(病床機能報告)

4. 結果

表 2 に病床機能報告，DPC 調査データにおいて，医療機関名称をキーとして一致分類をしたデータ件数を示す．全国保険医療機関データの医療機関コードを自動的に付与できたのは，完全一致・部分一致(単一)のデータであり，病床機能報告の 94.5%，DPC 調査データの 96.2%であった．

表 2. オープンデータ結合結果

	病床機能報告	DPC 調査
完全一致	5755	4172
部分一致(単一)	900	943
部分一致(重複)	7	3
不一致	379	198

参考文献

[1]　令和２年度病床機能報告の報告結果(厚労省)
[2]　DPC 導入の影響評価に関する調査(厚労省)

Copyright © 2023 IEICE