

検討テーマと主な論点(案)②

※第1回有識者会議で示した検討テーマのうち、
(4)から(6)までについて、事務局でまとめたもの。

平成30年6月14日

厚生労働省老健局・保険局

○ 主な検討テーマは、下記を想定。

〔 (1) 個人情報保護法制等との関係

(2) データの収集・利用目的、対象範囲

(3) 第三者提供

(4) 実施体制

(5) 費用負担

〔 (6) 技術面の課題（セキュリティの確保等を含む。）

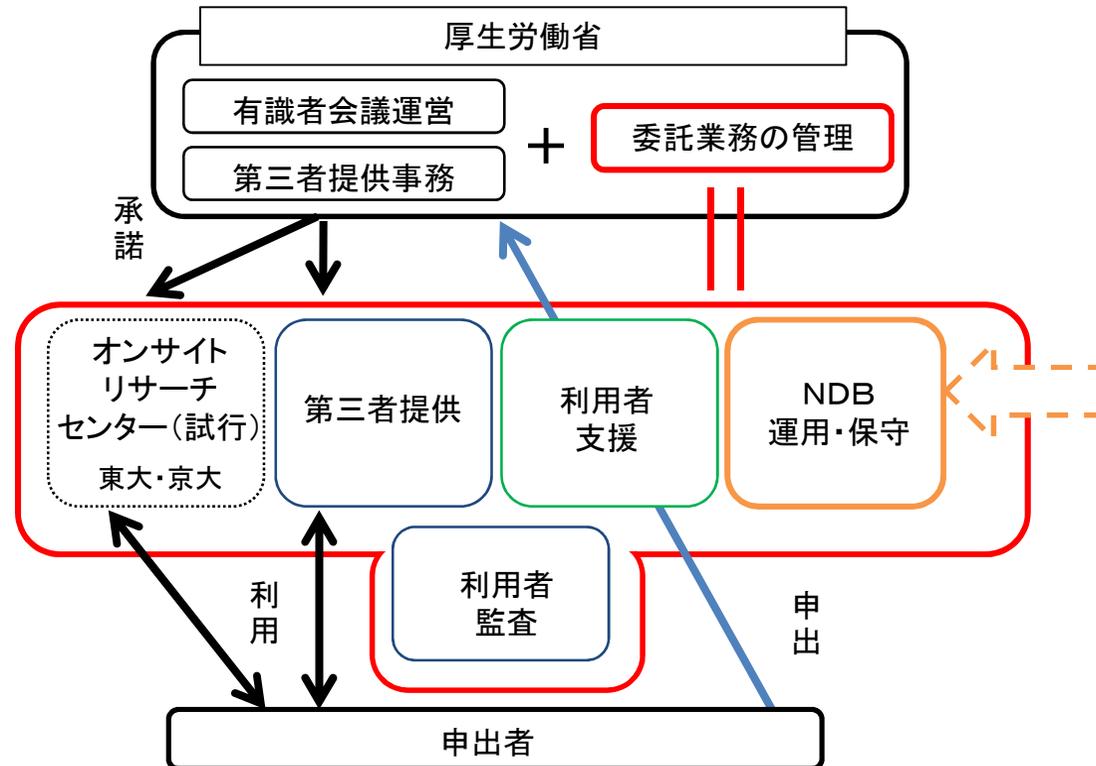
○ 上記以外の検討テーマも、必要に応じて追加して議論。

(4) 実施体制

現状（実施体制）

- NDB、介護DBともに、保有主体である国（厚生労働大臣）が責任主体となり運営。
※データベースの保守運用や、第三者提供の申出者からの受付は外部委託により実施。
- 先行して第三者提供を実施するNDBにおいては、上記に加え、利用者支援業務やオンサイトリサーチセンターの試行など、利用申出者の支援につながる取組を実施（外部委託）。
- データベースの構造や分析手法の開発に関しては、国の研究事業等において、外部の専門人材による研究の成果が一定程度蓄積。こうした成果が活用できれば、データベースの機能向上も期待される。

《参考》管理運営体制のイメージ(NDB)



主な関連研究

- ・新たなエビデンス創出のための次世代NDBデータ研究基盤構築に関する研究
- ・エビデンスの飛躍的創出を可能とする超速・超学際次世代NDBデータ研究基盤構築に関する研究
- ・診療現場の実態に即した医療ビッグデータ(NDB等)の解析の精度向上及び高速化を可能にするための人材育成プログラムの開発と検証に関する研究

(参考①) NDBのこれまでの取組

- 平成21年度の構築以来、レセプト情報、特定健診・保健指導情報を悉皆的に収集。
- サーバの増設による抽出速度向上や、利用者支援、オンサイトリサーチセンターの試行運用等により、利用環境の整備を実施。

<参考>：総提供件数 157件(平成30年3月末時点)

<参考①>NDBのこれまでの取組

H21年度	H25年度	H26年度	H27年度	H29年度	H30年度
データベース構築	利用者支援開始	データベース更改	オンサイトリサーチセンター(試行)	・サーバ増設(関西地区) ※抽出速度向上等 ・匿名処理機能の向上	H32年度データベース更改作業の開始 ※抽出速度向上等

<参考②>関連する主な取組

取組	開始時期	事業概要
NDBの運営・保守	H21～	情報の適切な管理のための措置を講じつつ、データベースを管理・運用。
利用者支援	H25～	利用者の申出までの手続の円滑化の支援窓口業務。データ利用に関するマニュアルも作成。
第三者提供	H25～	有識者会議の審査を経て利用が認められた研究者等に対し、NDBから対象データを抽出し、提供。
利用者監査	H23～	利用者への実地監査等を実施。
オンサイトリサーチセンター(試行)	H27～	NDBデータの利用の承諾を受けた研究者等が直接センターに出向き、データ集計作業を可能とする。

(参考②) 主な関連研究の概要

研究名	研究代表者	研究概要、目的
新たなエビデンス創出のための次世代NDBデータ研究基盤構築に関する研究	黒田 知宏 (京都大学)	<ul style="list-style-type: none">・大規模データの高速度検索、抽出を可能とするためのデータ処理基盤の開発・NDBデータ解析のスケラビリティを確保した効果的・効率的な基盤整備の検討・NDBデータ利用者の分析技術やレセプトに関する基礎知識の向上のための学習・試行環境の開発・NDBデータ利用者のデータ変換作業の負担軽減のための目的別データマートの整備
エビデンスの飛躍的創出を可能とする超速・超学際次世代NDBデータ研究基盤構築に関する研究	満武 巨裕 (医療経済研究機構)	<ul style="list-style-type: none">・研究システム基盤とオンサイト端末の設計、開発、評価を行い、大規模レセプト情報等の高速化処理の検証を行う。・研究システム基盤の利活用促進に向けた取組としての分析事例の蓄積、次世代NDB基盤構築と医療・介護政策等への利活用に関する提言を行う。
診療現場の実態に即した医療ビッグデータ(NDB等)の解析の精度向上及び高速化を可能にするための人材育成プログラムの開発と検証に関する研究	康永 秀生 (東京大学)	リサーチ・クエスチョンを解き明かす臨床研究・疫学研究・ヘルスサービスリサーチの実践能力や、大規模データベースを効果的・効率的にハンドリング能力を養成するプログラムを開発すること等を通じ、医療ビッグデータ研究の技術水準の向上を目指す。

論点V-1 (効果的・効率的な実施体制)

論点V-1 : 連結解析も念頭に、実施体制として求められる機能をどのように考えるか。

<現状・課題>

- ・ データベースの保守・管理が基本的な役割。
- ・ 第三者提供の開始に伴い、
 - 有識者会議の運営(H22. 10~)
 - 利用者への監査(H23~)
 - 利用者支援(H25. 4~)などの機能を順次付加。
- ・ 他方、実際の研究基盤としての利用のためには、技術の進歩を取り込みつつ、円滑なデータ処理のための最適なデータベース構造の確立や、データの迅速な解析を可能とするデータ解析機能の充実など、専門的・技術的な知見を取り込むことも求められる。

<検討の視点(例)>

- ・ 論点Ⅱ、Ⅲの議論も踏まえ、将来の解析ニーズの質・量両面の増加に対応するため、実施体制が備えるべき機能について、下記の両面を念頭に、どのように考えるか。
 - ① これまで取組が進められてきた、データベースの保守・管理、利用者支援の取組、第三者提供業務などの基本的な役割について、効果的・効率的に実施し、迅速に提供する機能
 - ② 利用ニーズの増加や解析ニーズの多様化・高度化に対応するため、データベースの構造改善やデータ解析機能の充実など、研究利用に応えるための取組について、効果的・効率的に実施する機能

※ 検討テーマ(5)(費用負担)の議論とも関係あり。

論点V-2（国と他の主体の役割分担）

論点V-2: 論点V-1の議論について、国に求められる役割や、国と他の主体との関係をどのように考えるか。

<現状・課題>

- ・データベースの保有主体である国(厚生労働省)が一義的な責任主体。
 - ※データベースの保守運用や、第三者提供の申出者からの受付は外部委託により実施。
- ・第三者提供の判断、NDBオープンデータの作成は、有識者会議の意見を聴き、国が行う。

<検討の視点(例)>

- ・データベースの保有主体が国(厚生労働省)であるという基本的な性格を踏まえて整理。
 - 第三者提供の判断等は国が自ら行う整理。
- ・論点V-1の議論を踏まえ、実施体制に求められる機能に即して、
 - ① 国が自ら担う機能
 - ② 他の主体との役割分担が適当な機能
 - ③ ②について、国の関与の在り方、他の主体に求められる要件について、検討してはどうか。

(5) 費用負担

費用負担に関する課題

<現状>

- NDB、介護DBともに、運用に要する費用は国が予算措置により対応。
- 第三者提供については、有識者会議の審議、個々の提供の判断を経て、提供範囲のデータ抽出作業(プログラム開発+データ抽出)を行い、提供。手数料等の利用者の費用負担は求めている。

<参考1:運営経費>

	NDB	介護DB
運営経費	約4.7億円(国費)	約2.8億円(国費)

<参考2:NDBデータの提供決定後の作業の流れ>

提供決定 → 提供範囲のデータの抽出のためのプログラム開発(①) → データ抽出(②) → 提供

(標準的な作業日数)

- ① プログラムの開発 …… SE4～5 人日
- ② プログラムに沿ったデータ抽出 …… SE2～3 人日

(第三者提供件数(実績))

約60件/年

<課題>

- 国が構築、保有し、一義的な責任主体であることを踏まえ、基本的な部分は国で対応。
- 連結解析を念頭に、利用ニーズの増加、多様化・高度化への対応が求められる。(論点V-1参照)
- 上記に伴い、第三者提供に要する費用も増加することが見込まれる。

※第三者提供については、公益性・提供に伴う受益(メリット)の双方が存在

⇒ 検討テーマ(2)～(4)の議論も踏まえ、上記の課題への対応の方向性と合わせ、国による対応を基本としつつ、費用負担について検討が必要。

(6) 技術面の課題
(セキュリティの確保等を含む。)

1. 匿名性を確保した状態での連結

- ・NDB、介護DBともに、データの収集に当たっては、個人の特定につながる情報を削除する処理を行った上で、国のデータベースに提出、収載されている。
- ・上記の処理の際、特定の情報項目を元に乱数を生成。この乱数を連結キーとして活用し、各データベース内の情報を、匿名情報としての性質を維持した上で、同一人の情報として識別・連結。

NDB → 保険者番号、記号・番号、生年月日、氏名を用いて乱数を生成。(注)留意点は参考資料1-3のP7、8を参照。

介護DB → 被保険者番号を用いて乱数を生成。(保険者番号と組み合わせて識別・連結。)

- ・現在は、両データベース間の情報連結はできないが、両制度のレセプト等で共通して収集される情報(例. 氏名、性別、生年月日)を元に、共通の乱数を生成し、匿名情報としての性質を維持した上で、同一人の情報として識別・連結することは技術的には可能。
- ・更に、技術面の環境整備等に応じて、両データベース間の情報連結の手法として、匿名情報としての性質を維持した上で、識別・連結する精度の向上につながる方策(個人単位被保険者番号(医療保険)の活用等)を取り入れることも選択肢。

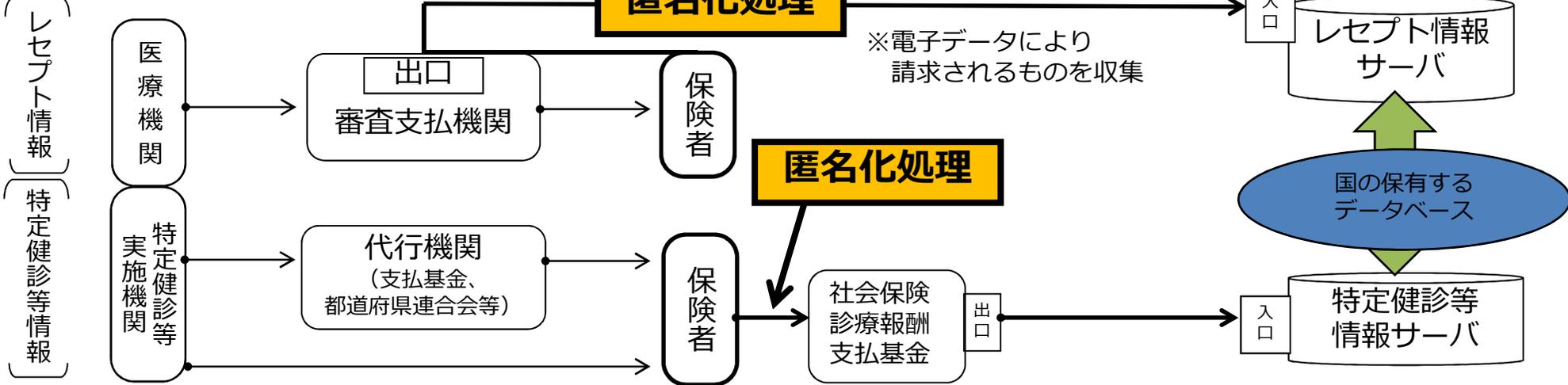
⇒ 両データベースの匿名性を維持した上での連結解析のために用いる技術的手法について検討が必要。

2. 1以外の技術的課題

- ・必要なセキュリティの確保や、迅速な提供や利用ニーズの増加、解析ニーズの多様化・高度化に対応するための機能の確保等、検討テーマ(2)~(4)の議論に応じ、必要な技術的対応の内容についての検討が必要。

(参考①) NDBの収集経路と匿名化処理

① 収集経路

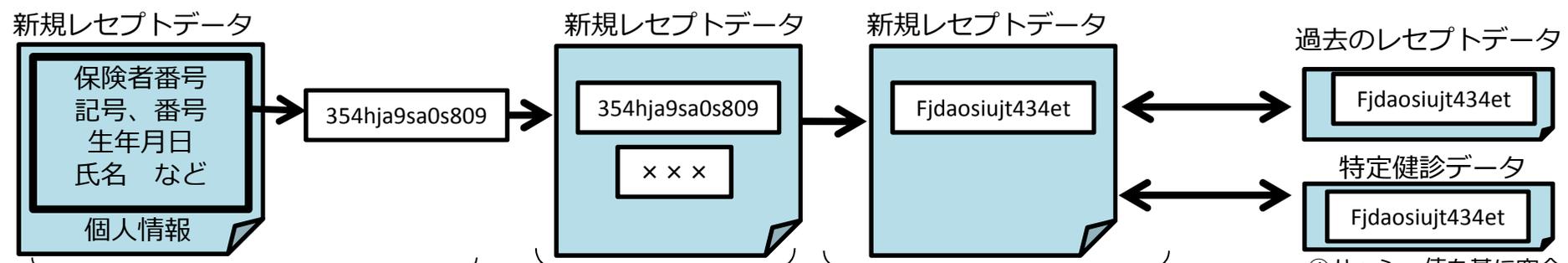


② 匿名化処理について

・「ハッシュ関数」を用い、**個人特定につながる情報を削除 (=匿名化)**。下図のように、同一人物の情報を識別・突合し、保管。

特徴

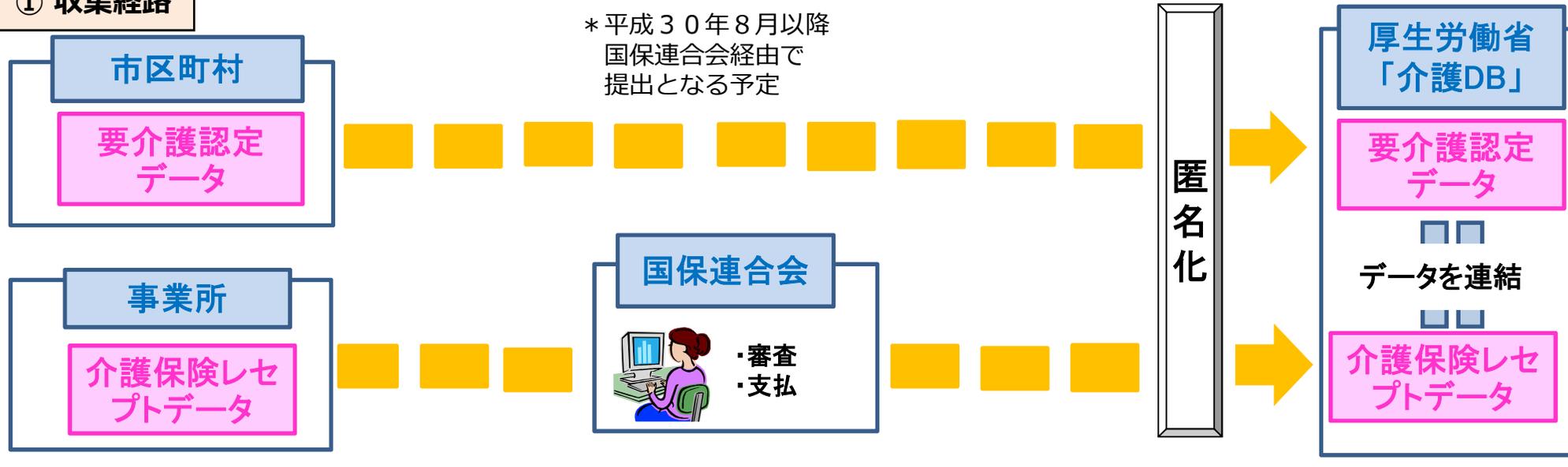
- ① 与えられたデータから固定長の疑似乱数（ハッシュ値）を生成。
 - ② 異なるデータから同じハッシュ値を生成することは極めて困難。
 - ③ **生成された値（ハッシュ値）からは、元データを再現することは出来ない。** 対応表も作成しない。
- ※ 個人情報（氏名、生年月日等）を基にしてハッシュ値を生成し、それをIDとして用いることで個人情報を削除したレセプト情報等について、同一人物の情報として特定することが可能。



- ① 個人情報をもとにハッシュ値を生成
- ② 個人情報を削除。ハッシュ値のみ残し、運用管理業者が独自キーを発生。
- ③ 1次ハッシュ値と独自キーに基づき2次ハッシュ値を作成。
- ④ ハッシュ値を基に突合

(参考②) 介護DBの収集経路と匿名化处理

① 収集経路



② 匿名化处理について

