

再生医療・遺伝子研究実用化と安全な医療情報の提供

PARI－国民医療ナンバー制度のあり方を考える

2013年2月14日

東京大学大学院情報学環 山本隆一

http://www.megabank.toho... 理念と概要：東北...

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(I) ヘルプ(H)

Google 東北メディカルメガバンク 検索 共有 詳細

東北メディカル・メガバンク機構は、地域医療支援、医療情報ICT化、バイオバンク構築事業を推進します。

東北メディカル・メガバンク機構
TOHOKU MEDICAL MEGABANK ORGANIZATION

お問い合わせ | アクセス | サイトマップ |
日本語 | English

サイト内検索 検索

トップページ HOME 地域の方々へ For Community 活動内容 Activities 進学ガイドと人材育成 Education and Training 制作物 Products 主な成果 Major Results

機構について

- ご挨拶 Greetings
- 組織・メンバー Organization and Members
- 理念と概要 Mission and Outline
- 背景と経緯 Background

HOME > 理念と概要

理念と概要 Mission and Outline

理念

東北メディカル・メガバンク機構は、未来型医療を築いて震災復興に取り組むことを目的に設置されました。機構は、東日本大震災の被災地の地域医療再建と健康支援に取り組みながら、医療情報とゲノム情報を複合させたバイオバンクを構築します。そして、構築するバイオバンクの情報とその解析結果に基づき新しい医療の創出を通じて、被災した東北地区への医療人の求心力向上、産学連携の促進、関連分野の雇用創出、そして被災地区の医療復興を成し遂げたいと考えています。

事業の概要

東北メディカル・メガバンク機構が取り組む主な事業は、医療情報とゲノム情報を組み合わせたバイオバンクの構築、地域医療情報連携基盤の構築、高度専門人材の育成の3つに分けられ、およそ10年に及ぶ事業を計画しています。構築されたバイオバンクから遺伝情報の解析を進め、ゲノム情報に基づいた未来型医療を築きます。

事業の推進にあたって、まず、東日本大震災により大きな被害を受けた地域における住民の長期健康調査とその結果の回付、地域の医療機関での勤務を組み入れた循環型医師支援システムの確立に取り組めます。地域医療情報連携基盤の構築のために、医療情報のICT化を推進し、二度と診療情報を失うことなく、地域で共通した医療記録にアクセスできる仕組みづくりを行います。

http://biosciencedbc.jp/ バイオサイエンス...

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(I) ヘルプ(H)

Google NBDC 検索 共有 詳細

ホーム NBDCについて 研究開発プログラム 公募概要 採用情報 広報 お問い合わせ リンク

新着情報 ニュース一覧 twitter RSS

2013/02/06 [【横断検索】 3つのデータベースが追加で検索できるようになりました。](#)

2013/02/05 [【メンテナンス】 2013年02月15日\(金\) 15:00~ 02月19日\(火\)17:00の間、サーバメンテナンスのため、JSNP \(日本人一塩基多型データベース\) サービスを停止します。](#)

2013/02/01 [【横断検索】 「KNApSack Family -DietNavi」が検索できるようになりました。](#)

2013/01/17 [2013年度 統合データベース講習会の受け入れ機関の募集を開始しました。](#)

2013/01/11 [「バイオインフォマティクス人材に関するアンケート調査」\(1月11日~2月9日\)を実施しています。アンケートへのご協力をよろしくお願いいたします。](#)

データベース横断検索

データベースのカタログ
[Integbioデータベースカタログ](#)

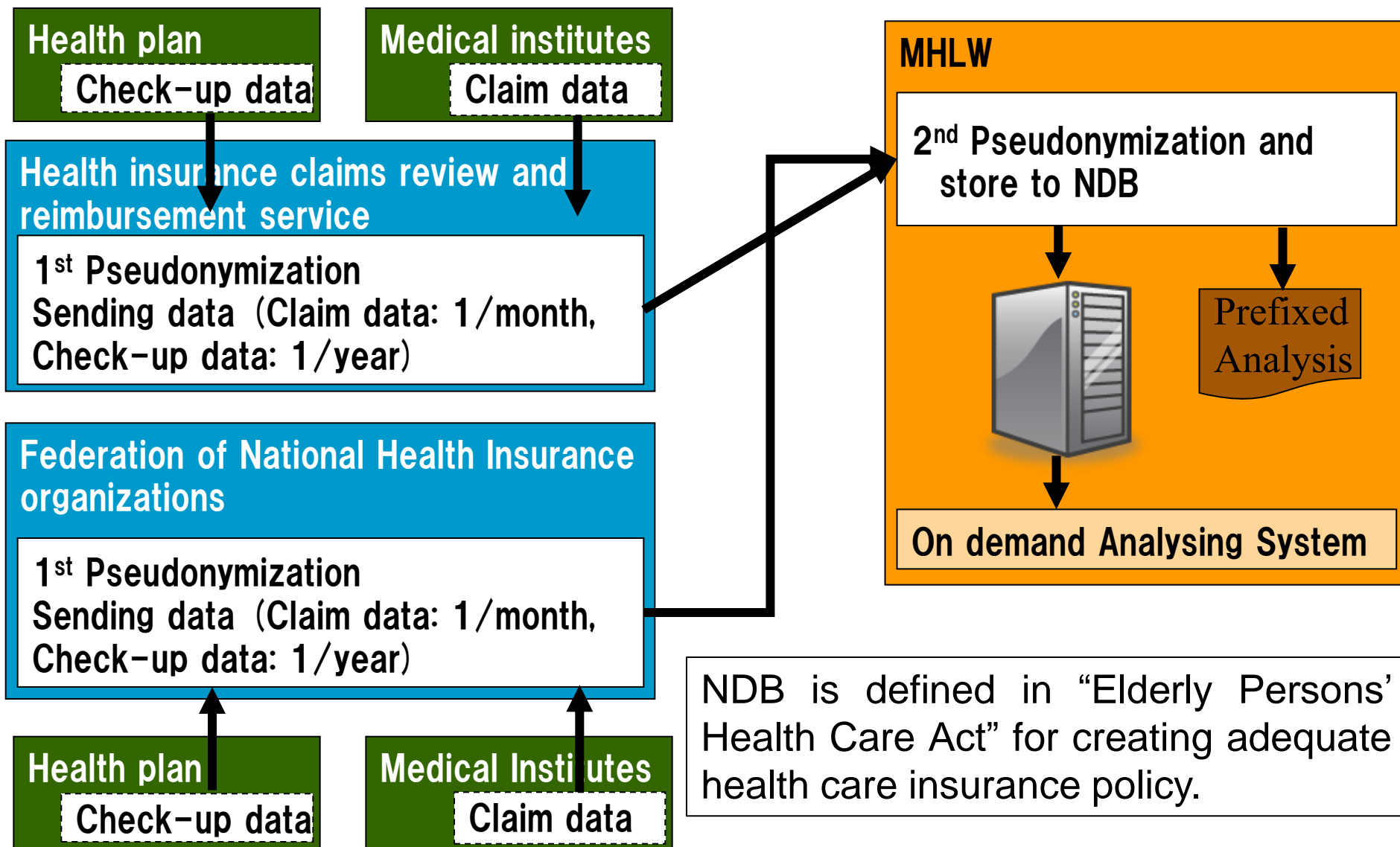
コンテンツ
[生命科学系 学協会 カタログ](#)
[生命科学系主要プロジェクト一覧](#)
[生物アイコン](#)
[Webリソースポータルサイト](#)
[ゲノム解析ツール リンク集](#)
[ライフサイエンス 新着論文レビュー](#)
[ライフサイエンス 領域融合レポ](#)

広報
[NBDC広報サイト](#)
[パンフレット\(PDF:2.78MB\)](#)

アーカイブ
[生命科学系データベースアーカイブ](#)

開発ツール
[TogoDB](#)
[TogoWS](#)
[DBCLS Galaxy](#)

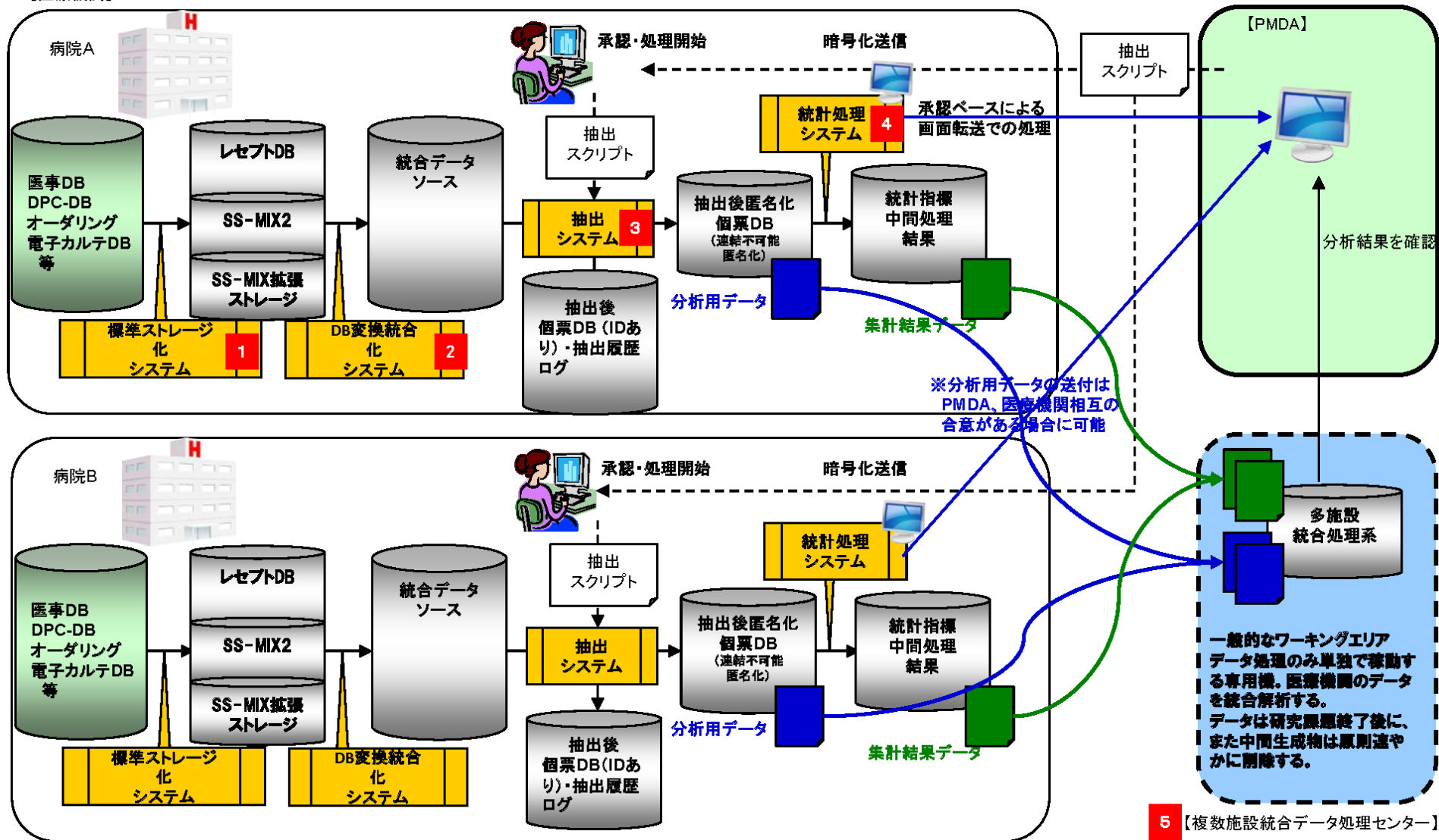
National Claims and specific screening information database system of Japan (NDB)



Sentinel Project in Japan (so-called)

- 集計結果データ** ■ 統合データソースから抽出した集計結果情報(主に、発生頻度、件数、記述統計、クロス集計結果等)
- 分析用データ** ■ 統合データソースから抽出した分析用データ(主に、データをマージ・ソート処理が必要な統計処理等)

【医療機関】



5 【複数施設統合データ処理センター】

利活用とプライバシー保護

- ＞ 個人の特定可能性が低いデータセットであれば法的あるいは運用（倫理綱領等）で対応可能
- ＞ 稀少疾患等ではそもそも特定性を下げることは難しい。
- ＞ 公益とのリスクのバランスが重要
匿名化を再考しなければならない
- ＞ リスクがどうしても排除できない場合は？
状況によっては秘匿計算が利用できる？

匿名化すれば個人情報ではない

では匿名化とは何か？

- ▶ **医療・介護事業者における個人情報の適切な取り扱いのためのガイドライン**
当該個人情報から、当該情報に含まれる氏名、生年月日、住所等、個人を識別する情報を取り除くことで、特定の個人を識別できないようにすることをいう。
- ▶ **疫学研究に関する倫理指針**
個人情報から個人を識別することができる情報の全部又は一部を取り除き、代わりにその人と関わりのない符号又は番号を付すことをいう。資料に付随する情報のうち、ある情報だけでは特定の人を識別できない情報であっても、各種の名簿等の他で入手できる情報と組み合わせることにより、その人を識別できる場合には、組合せに必要な情報の全部又は一部を取り除いて、その人が識別できないようにすることをいう。
- ▶ **匿名データの作成・提供に係るガイドライン(統計法)**
匿名化の基準：
調査票情報の特性は統計調査ごとに異なることから、各統計調査について一律に匿名化の基準を設定することは困難である。

匿名化すれば個人情報ではない

では匿名化とは何か？

> HIPAA Privacy rule (U.S.) § 164.514

次のいずれかの場合、個人が識別できないとして良い。

(1)一般的に受け入れられ、統計的かつ科学的な原則及び方法に関して適切な知識及び経験を持った者が、情報の利用を行うものが、単独または合理的に入手可能な情報と照らし合わせることで、個人が特定されるリスクを評価し、リスクが十分低いことを判断した分析の経過および結果を文書化した場合。

(2)以下のあげる19項目を本人、親類、雇用者、世帯員に関する情報から除いた場合。

名前、州より小さい範囲の住所、2万人以下に限定される郵便番号の上位3桁(2万人以下に限定される場合は000に)、すべての日付(年は除く)と89歳以上の年齢、電話番号、FAX番号、電子メールアドレス、SSN、カルテ番号、保険番号、口座番号、免許証番号、車両番号およびシリアル番号、装置の識別番号およびシリアル番号、URLs、IPアドレス、生体識別情報(指紋等)、顔写真を含む識別可能な写真、連結可能または不可能匿名化のために付与した番号を除く固有の数字・特徴またはコード

利活用試行例 レセプト等情報データベース(NRDB)

- 氏名○生年月日の「日」○保険医療機関の所在地及び名称○カルテ番号等○国民健康保険一部負担金減額、免除、徴収猶予証明書の証明書番号○被保険者証(手帳)等の記号・番号○公費受給者番号は削除、保険医療機関番号は保持
- 保険者番号・記号番号・生年月日・性別からハッシュ値①を生成、氏名・生年月日・性別からハッシュ値②を生成して格納している。
- 疫学研究指針で言う連結不可能匿名化とは考えていない。
- 公益性の確認、目的外利用の禁止、地域情報、医療機関情報の原則提供禁止、安全管理の確認、公表形式の審査(患者の場合は10人未満に特定されてはいけない、等)

匿名化と仮名化 (Anonymization and Pseudonymization)

- ＞ 仮名化:個人を識別できる情報を元に戻れない変換(一方向変換)で、一意の識別子に置き換えた状態。対応表が存在し、それを使うことで、個人の情報に戻ることができる状態。

対応表が安全に管理されていれば、個人が識別される恐れは低い。

連結可能匿名化と同じ意味で、対応表を破棄すれば、あるいは初めから対応表を作らなければ連結不可能匿名化と呼ぶ。

ISO/TS25374 Health Informatics – Pseudonymization (2008)

連結不可能匿名化は一般に言う匿名化と同じ意味。
つまり匿名化とは何かという問は残る。

医療・介護・福祉情報における匿名化(二次利用)

- 多くの場合、HIPAA Privacy Rule § 164.514の(2)番目の規則のように個人の識別につながる項目を明確にした上で、除去すれば問題はない。
- しかし……
- 精度が落ちすぎて公益性の高い調査さえできないこともある。(日付、地理情報……)
リスクを十分下げれば利用可能では？
- 稀少疾患、頻度が低い医療行為、少数しか用いられない薬品など、個人の識別につながる可能性を排除できない。
特に身近な人の場合、関連情報が入手しやすくリスクが増える。
- リスクの評価が必要な場合がある。(HIPAA Privacy Rule § 164.514 (1))
完全匿名化が自明でない場合はリスク評価を行うべきか。

匿名化のリスク評価を導入すべきではないか

> 最悪の場合、何人に特定されるか？

最小特定人数 (BIN) L. Sweeney, 1998, Medinfo Proceeding

K-匿名性 (k-anonymity) L. Sweeney, 2002

もっとも多く用いられており、ツールも豊富。

ただし計算量は比較的大きい。

> その項目はどの程度多様か？

l-多様性 (l-diversity) A. Machanavajjhala et.al., 2006

> 分布に著しい偏りがないか？

t-近似性 (t-closeness) N. Li et.al., 2007

「匿名化」ではリスクを下げるできない場合

- ＞ 秘匿計算は解決策となるか？
- ＞ 準同型暗号(暗号化状態で演算可能)を使う
 - いくつかの暗号が準同型性を示すことが知られている
 - Pailler暗号(加算準同型)、RSA暗号(乗算準同型)
- ＞ 秘密分散を用いる
 - Shamirの (k,n) しきい値法による秘密分散
 - 中国人剰余定理による秘密分散
 - n 個に分散された情報は $k-1$ 以下の情報では情報値は0で論理的に元の情報は回復できない。
 - n 個に分散した状態で演算可能

Conclusions

- 個人特定性が無視できないデータセットを用いる場合は匿名性のリスク評価が必要
k-anonymity, l-diversity, t-closeness
- リスクが大きい場合は秘匿計算を用いるべき
少なくとも計算時点では個票の値を知ることはできない。
- ただし、検索、計算を繰り返すことで特定性が増大することもあり、プロセスの事前審査は必要。