

LDIによる二次利用：現状とこれから

2021年5月28日

株式会社 NTTデータ 製造ITイノベーション事業本部

西田 陽介

本日の内容は個人的見解であり、所属する団体等の公式な見解ではございません。

NTTデータについて

NTTデータは、より多様化するお客さまや社会の期待に応えるため、これまで以上に事業を跨った連携や、迅速な意思決定を図るため、事業本部レベルでの業務執行を可能とする体制とさせていただきます。



公共・社会基盤分野

国内外の公共機関向けビジネスを主な事業ドメインとする分野です。中央省庁や海外公共機関、地方自治体などの行政部門や、病院などのヘルスケア分野などに対し、社会的な基盤を担う高付加価値なITサービスを提供しています。

金融分野

国内外の金融機関向けビジネスを主な事業ドメインとする分野です。銀行、保険、証券など、金融業界のあらゆる業態に対し、ミッションクリティカルな大規模ITサービスを、グループ一体となって幅広く提供しています。

法人・ソリューション分野

製造業、流通・サービス業などの幅広い業種のお客さまに対し、事業活動を支える高付加価値なITサービスを提供する分野です。単納期・高品質・低コストといったお客さまのさまざまなニーズにこたえるベストソリューションを提供しています。

中国・APAC、北米、EMEA・中南米分野

海外市場において高付加価値なITサービスを提供する分野です。Global IT Innovatorとして、NTTデータの世界展開を司る役割を担っています。

技術革新統括本部

先進的なITサービスを支える基盤・ソリューションを生み出す分野です。グローバル事業の伸長や、急速なIT技術の進化に対応するために、グループ全体での生産性向上を図り、グローバルでの競争力強化を目指します。

1. 次世代医療基盤法におけるNTTデータの役割

次世代医療基盤法の制定の背景

1. インputのみならずアウトカムも含む医療情報の利活用

- レセプト（診療報酬明細書）は、インput（診療行為の実施に関する情報）を含むが、アウトカム（診療行為を実施した結果に関する情報）を含まない。
- 医療分野の研究開発に資するよう、カルテ（診療録）など、アウトカムを含む医療情報の利活用のための仕組みを整備することが求められた。

2. 医療情報の分散保有

- 我が国では、国民皆保険制度の下、医療情報が豊富に存在しているものの、医療機関が民間中心であるとともに、医療保険者が分立しているため、医療情報が分散して保有されている。
- 医療分野の研究開発に資するよう、医療情報を「集めて」「つなぐ」仕組みを整備することが求められた。

3. 改正個人情報保護法の施行

- 平成29年5月に施行された改正個人情報保護法では、
 - ① 病歴を始めとする要配慮個人情報を第三者に提供するに当たっては、学術研究等を除いては、いわゆるオプトイン（あらかじめ本人が同意すること）によらなければならない、いわゆるオプトアウト（本人が停止を求めないこと）によることができない
 - ② 特定の個人を識別できないように加工された匿名加工情報については、個人情報と比較して緩やかな規律で第三者に提供することができるものとされた。

個人情報保護法の特則となる次世代医療基盤法の制定

- 平成30年5月に施行された次世代医療基盤法では、オプトイン（あらかじめ本人が同意すること）のほか、一定の要件を満たすオプトアウト（あらかじめ通知（※）を受けた本人又はその遺族が停止を求めないこと）により、
 - ① 医療機関等から認定事業者へ要配慮個人情報である医療情報を提供することができる
 - ② 認定事業者から利活用者へ匿名加工医療情報を提供することができるものとされた。

引用：内閣府HP <https://www8.cao.go.jp/irvou>

（※）医療機関等の場合には、最初の受診時に書面により行うことを基本とする。

認定事業者

(一社) ライフデータイニシアティブ



豊富な研究実績を保有する有識者にて構成
アカデミア中心の運営による社会的信頼

受託事業者

(株) NTTデータ

社会インフラ構築で求められる高度なセキュリティ技術
多業界のシステム構築実績にて培った高いデジタル技術



NTT DATA

(サービス名) 千年カルテ

安全・安心



大規模な電子カルテ
データベース



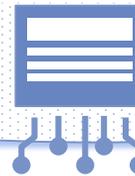
次世代医療基盤法に基づく厳格なセキュリティ管理
個人の特定リスクを低減する匿名加工技術

匿名加工された
医療情報



医療分野の研究開発促進

高度なIT技術

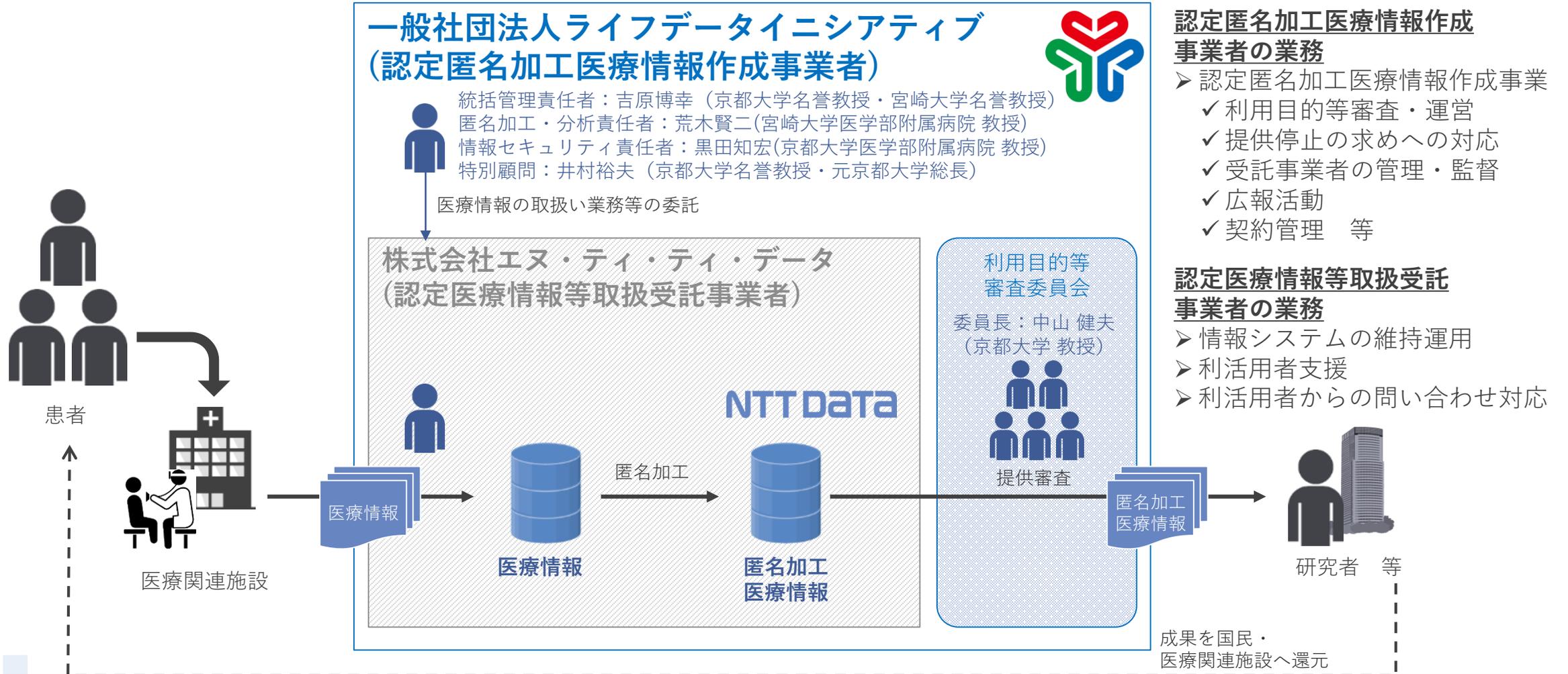


先進医療を担う特定機能病院等を中心に数多くの
医療機関から医療情報を収集
薬剤や治療の効果測定、副作用解析に必要な検査
値を含む電子カルテ由来の情報を保持

疾病予測や未病・予防解析、リスクファク
ター特定など新たな価値の創出に貢献

ビッグデータ化した診療情報を用い、必要なAI解析（機械
学習・深層学習・自然言語処理）などを支援

次世代医療基盤法に基づく医療情報利活用は、事業主体であるライフデータイニシアティブと、システムの運用を委託されるNTTデータの両社により推進されます。



2. 千年カルテ2次利用について～研究～



千年カルテ2次利用DB取り扱いデータ一覧

千年カルテ2次利用DB取り扱いデータ

患者IDで紐付け

- 患者情報
(**患者ID**を含む)
- 保険情報

- 診断情報
- 診療行為**
- 医薬品情報
- その他

- 入退院情報
- 診断群分類 (DPC)
- がん患者情報
- その他

レセプトデータ

DPC調査データ

【構造化データ】

- 患者情報[Pi]
(**患者ID**を含む)
- 健康保険[Hi]
- 診断履歴[Rd]**
- 検歴情報[Lb]
- バイタルサイン[Vs]
- 体温表[Fs]
- 処方箋[Ps]

- 注射記録[Ing]
- 生活習慣情報[Ln]
- 基礎的診断[Bc]
- 手術記録[Sg]
- 透析[Hd]
- 汎用処置[Pr]

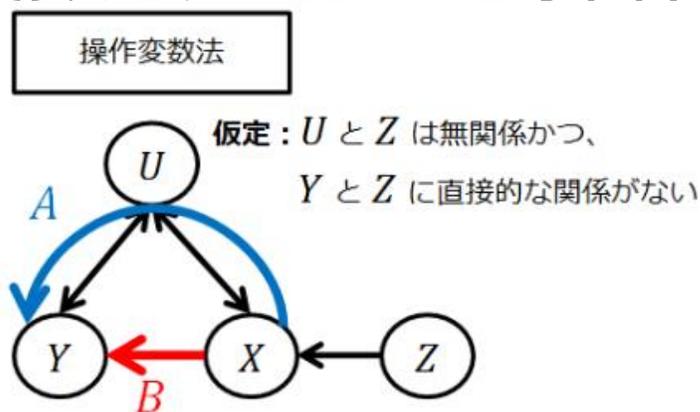
【非構造化データ】

- 経過記録[Pc]
- 臨床サマリ[Sm]**
- 報告書[Rp]
- 紹介状[Re]
- 初診時特有情報[Fcl]

電子カルテデータ (MML)

精緻なアウトカム定義、Criteria定義が期待される。

- アウトカムバリデーションのPPV向上
- 感度分析のバリエーション期待
- 傾向スコア、共変量の多様性
- 操作変数法における操作変数Zの選択肢



Bの効果が存在するならば、Zが変動するとXを通じた間接的な効果のみによってYも変動する。この時、Zの変動でUは変動しないので、Aの効果は生じない。これを利用し、Bの効果のみを推定する方法が操作変数法。

電子カルテの臨床サマリには、どういった記載があるのか、サンプル情報を提示します。行為しか把握ができないレセプトに対し、電子カルテと行き来することで、検査結果や患者背景、治療内容を把握することが可能となります。

臨床サマリ情報のサンプル（一部）

※ダミーデータ

<mmlSm : history>

4/01 左乳房に疼痛自覚し、〇〇病院に受診された。妊孕性温存を希望され、当科へ紹介。

4/18 母が乳癌に罹患していたため、遺伝性乳がん卵巣がん症候群を疑い、遺伝カウンセリングを希望された。遺伝子検査の結果、**BRCA陰性**が判明した。

⋮

<mmlSm : clinicalRecord>

#右C領域 閉経前浸潤性乳管癌(cT1cN0M0,cStageI,luminal-B type)

⋮

6/22 右乳房部分切除術・センチネルリンパ節生検施行。
術後補助化学療法(TC4コース)の方針。

6/29 退院。

<mmlSm : testResults>

⋮

Immunohistology: **ER: Positive(100%) PgR: Negative(0%)**
HER2(IHC): Negative(score 0) Ki-67:21%(106/500)

非構造化データの課題

→表記揺れ

例、骨粗鬆症

- DXA ? DEXA ?
- TRACP-5b ? TRAP – 5 b ?
- BRONJ → ONJ

→Feasibility時に表現系の辞書を詰めていく必要あり。

東京大学ITHC 清水研研究内容

レセプト

電子カルテ

電子カルテ及びレセプトから取得した1000名の乳癌患者うち、入院症例600名のサブタイプ検査数、検査結果記載数について整理しました。

調査対象	症例数 入院患者	ホルモン受容体		現在研究中の内容のため、ダミーデータ		
		ER	PgR	ER or PgR	HER2遺伝子	
ソース	電子カルテ	電子カルテ	電子カルテ	レセプト	電子カルテ	レセプト
A病院	100	50	50	20	50	20
B病院	200	250	200	80	250	80
C病院	300	200	150	250	200	250
合計	600	500	400	350	500	350

※電子カルテ(診断履歴[Rd],臨床サマリ[Sm])及びレセプト(診療報酬レコード[IY])を用いて集計

■レセプト算定件数と電子カルテへの記載件数の乖離について

i、AB施設は 電子カルテ>レセプト

理由: 転院前情報、セカンドオピニオンやコンサル結果など、他施設で実施した診療行為の情報がサマリー中に記載されている可能性がある。

ii、C施設は 電子カルテ<レセプト

理由:

- ①ERなどの自然言語の辞書はA、B施設で作ったきりでC施設の情報を含めた辞書としていなかったため、AB施設で使われず、C施設で使われるタームの抽出漏れの可能性
- ②MMLの他項目に記載されている可能性
- ③MMLで抽出していないその他領域に記載されている可能性。

Source: 「千年カルテ」よりNTTデータ調査

検査値データの課題

→施設間データ項目の不整合

- ・各施設基準値の多様性
- ・異なる単位 例、Ca : mg/dl、mEq/L
- ・病院独自の検査基準

→NTTデータ側で検査値マスターを作成し紐づけ可能。
ユーザー見解を基に連結対象を調整する運用。

【参考】利活用事例の紹介

次世代医療基盤法に基づくデータ利活用について下記のプレスリリースを行いました。



<https://www.nttdata.com/ip/ja/news/release/2020/121402/>

■ 研究概要

○ 目的

次世代医療基盤法に基づく匿名加工医療情報から、がん患者さんにおける臨床アウトカムを評価するための方法論に関する研究を行う

○ 内容

- RWDには医師の診療録などテキストデータが多く含まれる電子カルテの情報があり、それらをどのように臨床アウトカムの評価に資するデータとして取得するか検討する
- 検討結果より、電子カルテ情報から臨床アウトカムを生成する方法・アルゴリズムを開発する
- さらに、解析を行うために必要な情報を、全国の施設が含まれたデータベースからどの程度取得できるのかを確認する

現時点で何ができるか？



- ①アウトカムバリデーションスタディ
- ②希少疾病、難病対象としたケースシリーズ
- ③EHRの手法論開発を目的とした研究

既存データベースとは使い勝手が異なる。数多くの発見が眠っていると考える。