

エグゼクティブサマリー

今後進展するデータ駆動型社会において、データ活用が自律的・継続的に発展する「データエコシステム」の形成が重要であるとの認識のもと、その実現には、実際に使えるデータを生み出すこと、またデータへのアクセス性が高まることが必要と捉え、「クオリティデータ」と「アクセシビリティ」をメインテーマに本年度の調査研究を行った。

医療ヘルスケアの「データエコシステム」で主役となるのは「デジタル化された患者・生活者のデータ」である。電子カルテや画像データ等の医療情報に始まり、ゲノム情報、健診・検診データや生活・行動データなどの PHR (Personal Health Record) も含めた多様な情報が対象となる。これらのデータについて、利用目的に応じた質の確保や、連結・相互運用の可能性を高めるとともに、データへのアクセス性を高め、データから創出された価値や成果が速やかにフィードバックされる環境を整備することが必要となる。

健康医療に関する情報は、特に機微性の高い情報が多く含まれるため、利活用促進と個人情報保護の両立を図ることが求められる。個人が自分のデータにアクセスでき、また適切にコントロールできる環境整備と、公益性の高い研究や事業におけるデータ利活用を広げていく枠組みの構築といった、情報の提供から活用に係るそれぞれの立場でのアクセシビリティを高める方策が必要である。

また、データの二次利用を進める視点では、データベースやレジストリ等の構築に際し、二次利用も前提としたデザインを進めることが必要である。さらには、データの連結や AI による解析を可能とする構造化、標準化や相互運用性を高める重要性は言うに及ばない。

これらに加え、データ提供者がデータを安全に・安心して提供できるシステムを構築すること、国民と一体となった研究を推進し、成果やメリットを実感しやすい仕組みの社会実装を進めるといった、国民の理解を深める方策も重要である。

上記の調査研究から、本報告書では、下記の 3 つの観点で重要な方策をまとめた。

重要方策① 医療ヘルスケアデータのアクセシビリティの向上

- ✓ 医療健康分野の仮名加工情報の取り扱いを定める「医療情報特別法（仮称）」の創設
- ✓ 公益性の高い産業・事業のクライテリアの明確化（ガイドライン・ホワイトリスト等）
- ✓ 個人が自分のデータへのアクセスや活用をコントロールできる環境の整備

重要方策② 医療ヘルスケア分野のクオリティデータ収集の推進

- ✓ PHR や医療情報の標準化の推進と相互運用性の向上
- ✓ レジストリ等における医産学官連携の推進と 2 次利用に向けたデータ品質の確保
- ✓ IoT やクラウドの活用による ready for AI なライフコースヘルスデータの収集推進

重要方策③ 信頼に基づくデータ循環と社会への成果・価値の還元

- ✓ セキュリティやプライバシーも含めた、データ利用者の信頼性を評価する仕組みづくり
- ✓ 国民と一体となった研究の推進と成果やメリットをタイムリーに還元する仕組みの実装
- ✓ 実効性と納得感の高い「データガバナンスモデル」の構築と国民啓発

医療健康分野のビッグデータ活用研究会
報告書 vol.5
概要版

2020年6月

医薬産業政策研究所

森田 正実

佐々木 隆之

中塚 靖彦

2019年度活動（2019年4月～2020年3月）

OPIR
Office of Pharmaceutical Industry Research

2

研究員・事務局

医薬産業政策研究所

森田正実（統括研究員）、**佐々木隆之**（主任研究員）、**中塚 靖彦**（主任研究員）

協力メンバー

安中良輔（第一三共(株)）、**白神昇平**（アステラス製薬(株)）、**鈴木雅**（田辺三菱製薬(株)）

小林典弘（塩野義製薬(株)）

エキスパート勉強会

第14回 吉澤尚 先生 弁護士法人漆間総合法律事務所

第15回 水島洋 先生 国立保健医療科学院研究情報支援センター センター長

第16回 鈴木正朝 先生 新潟大学法学部 教授

メンバー情報交換会

黒田知宏 先生 京都大学 医学部附属病院 医療情報企画部

藤田卓仙 先生、山本精一郎 先生、岡本雅子 先生

世界経済フォーラム 第四次産業革命日本センター

1. はじめに
 - 1 – 1. これまでの研究会の活動について
 - 1 – 2. 「クオリティデータ」「アクセシビリティ」をメインテーマとした背景

2. 医療ヘルスケアデータのアクセシビリティ向上に向けて
 - 2 – 1. 医療ヘルスケアデータをとりまく法制度の現状と課題
 - 2 – 2. セキュリティ・プライバシー規制の現状と課題
 - 2 – 3. 新たなデータガバナンスの視点と概念
 - 2 – 4. 医療情報の二次利用に焦点を当てた法整備
 - 2 – 5. アクセシビリティに関する法規制等の在り方に対して期待と課題

3. クオリティデータの収集と活用
 - 3 – 1. データ駆動型社会でデータに求められる新たな「質」
 - 3 – 2. PHR、ライフコースヘルスデータの標準化
 - 3 – 3. 電子カルテ情報の質の向上に関する取り組み
 - 3 – 4. クオリティデータに関する期待と課題

4. まとめ

ビッグデータ活用研究会のキーテーマ

4

1年目のキーテーマ（2015年度） ⇒「研究会報告書vol.1（2016年7月）」
Precision Medicineの進展と疾患分類細分化、疾患層別化
Learning Healthcare、人工知能（AI）の活用、創薬手法の変化 等

2年目のキーテーマ（2016年度） ⇒「研究会報告書vol.2（2017年4月）」
EHRやPHRの現状と課題、2次活用のための基盤整備
IoT（モバイル、センサー、ウェアラブルデバイス等）による医療変革 等

3年目のキーテーマ（2017年度） ⇒「研究会報告書vol.3（2018年5月）」
製薬産業における「医療データ」の活用目的と必要データ
「医療データ」の活用を進める上で優先度の高い課題・取り組み 等

4年目のキーテーマ（2018年度） ⇒「研究会報告書vol.4（2019年5月）」
モバイルアプリやウェアラブルデバイスなど製薬産業における「活用データの広がり」
データポータビリティ推進に伴う「個人を軸としたデータ流通プラットフォーム」の活用 等

5年目のキーテーマ（2019年度）

データ駆動型社会で価値を生み出すデータの「クオリティ」とデータへの「アクセシビリティ」
「データエコシステム」の創出 等

医療ヘルスケアのデータエコシステム において考慮すべき「データの要素」

5

① データの多様性と連結（データの種類の観点）

- ・医療情報に加え、ゲノム等の生体分子情報、行動・生活情報など、
多種多様な幅広い情報の収集・統合が必要

② データの継続性、ライフコースデータの確保（データの時間軸の観点）

- ・発症前からの継続した時系列データを、個人を軸として収集・解析
長期にわたる「ライフコースヘルスデータ」の収集と連結が必要

③ 利用目的に応じたデータの質の確保（データの質の観点）

- ・標準化、構造化に加え、幅広いデータ利活用の視点からの相互運用性の確保
利用目的に応じたデータの質を確保しておくことが必要（Fit for Purpose）

④ データへのアクセス性と信頼性の確保（データのアクセシビリティの観点）

- ・安心・安全なデータ提供の仕組みづくりとデータ利用環境の改善
セキュリティ、プライバシー確保とアクセシビリティに関する法制度等の整備が必要

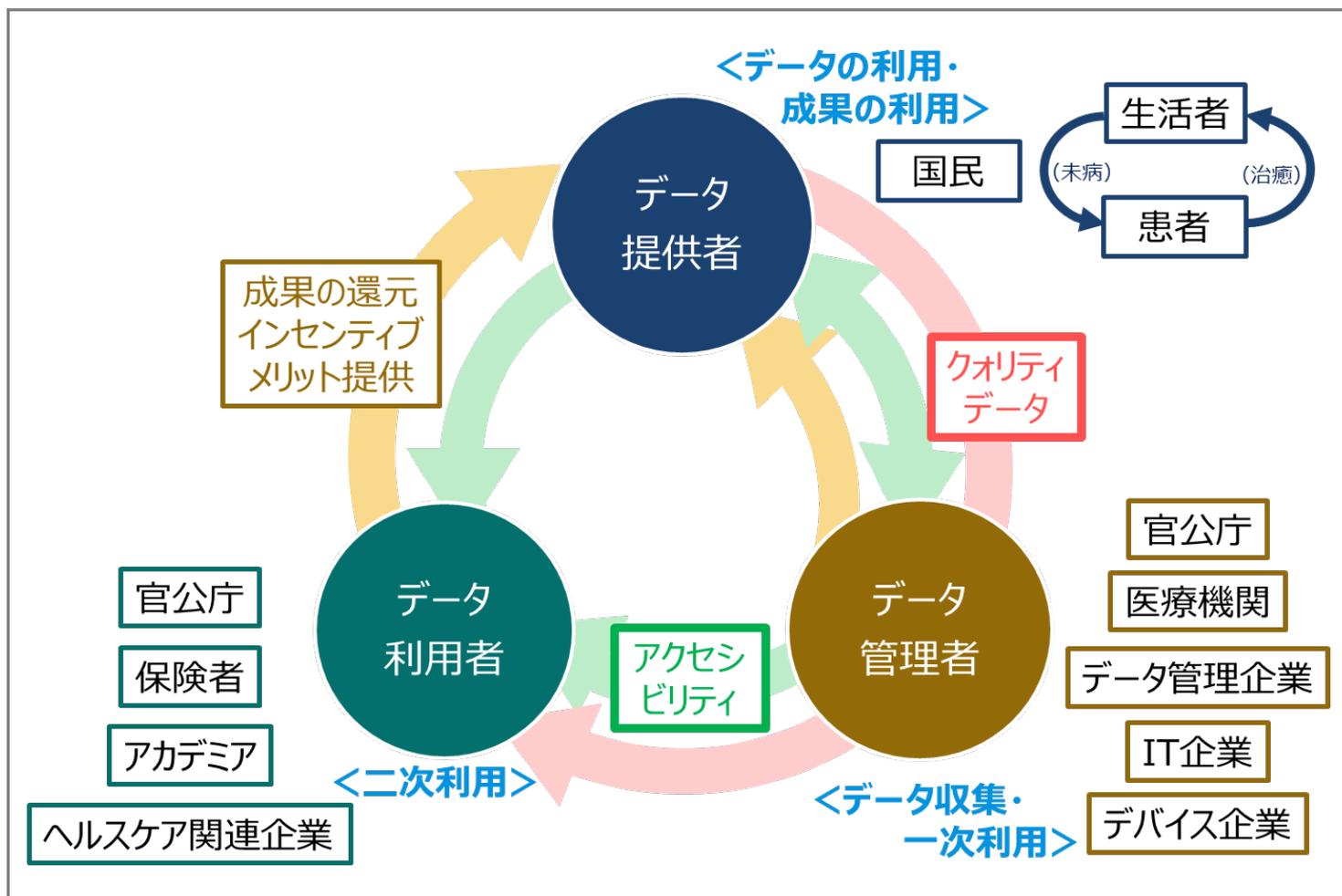
⑤ 成果や価値の還元（フィードバック、インセンティブの観点）

- ・データ提供者や社会がメリットを速やかに得られるよう、成果や価値を
タイムリーにフィードバックできる仕組みやデータ整備を支えるインセンティブが必要

これらが成立したうえで、それが**サイクルとして回ることが重要**

医療ヘルスケア分野におけるデータエコシステム

6



今後の医療ヘルスケアでは、データが価値創成の源泉であり、データと価値が循環する「データエコシステム」を構築していく必要があるという認識のもとに、データフローの重要要素である「クオリティデータ」「アクセシビリティ」をメインテーマに設定

医療ヘルスケアデータのアクセシビリティ向上に向けて

医療ヘルスケアデータの 「アクセシビリティ」に関する主なポイント

8

データ 提供者

- ✓ 自身のデータを大きな負担なくスムーズに閲覧できる
- ✓ 他者がどのように自分のデータを活用したか把握できる
- ✓ 自身の情報の活用をコントロールできる

データ 管理者

- ✓ スムーズな一次利用が可能となっている
- ✓ データ管理者間でスムーズにデータ連携ができる
- ✓ データ提供者・利用者に、適切なアクセス性が提供される

データ 利用者

- ✓ どのようなデータが収集されているか予め把握できる
- ✓ 大きな負担なくスムーズに二次利用ができる
- ✓ 必要があればデータ提供者にアクセスできる

<次世代医療基盤法の課題>

- 特異性が高く個人を特定しうる記載内容（希少疾患、超高齢者情報等）や画像データ（レントゲン、臓器画像等）等の匿名化によりデータ活用に制約がかかる可能性
- ゲノムデータは一般的に個人識別符号と認められ、個人の同意を得なければ、次世代医療基盤法のスキームの中では活用することが困難
- 認定事業者の取り扱うデータの項目、粒度等の詳細が未確定。医療機関のRWDの質をコントロールすることが難しい。

<「仮名加工情報」利活用のための基盤整備>

- 2020年の個人情報保護法改正で「仮名加工情報^{*1}」が新たに定義された
- 適切な利用目的の範囲内で利活用できるようにするための検討（特別法など）が必要

<セキュリティ・プライバシーに関する法制度の整備>

- 欧米でのセキュリティ・プライバシー関連法制度（GDPR^{*2}、CSF^{*3}）の制定

- 
- 次世代医療基盤法の認定事業者によるデータ項目等の開示と事業者との事前連携
 - 個人情報保護法における「適切な利用目的」の明確化（ガイドラインやQ&Aの充実）
 - 仮名加工情報をより活用できる法整備、取り扱う対象や利用目的等の検討
 - セキュリティ・プライバシーの国際規制との整合性確保、データポータビリティの推進

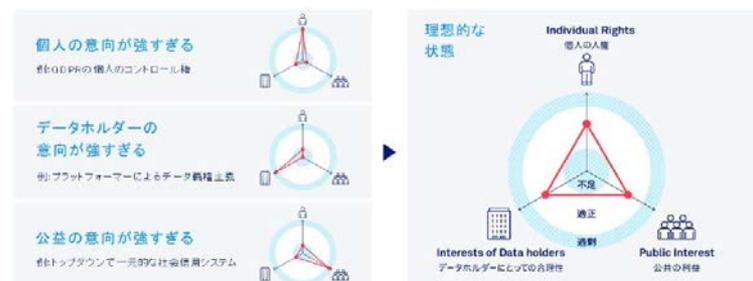
*1他の情報と照合しない限り特定の個人を識別することができないよう個人情報を加工して得られる個人に関する情報

*2General Data Protection Regulation（EU一般データ保護規則）：EUで策定された新しい個人情報保護の枠組み

*3Cyber Security Framework：サイバーセキュリティ対策を向上させるための指針

＜新たなデータガバナンスの視点と概念＞

- データ駆動型社会では、Data Free Flow with Trust (DFFT ; 信頼性のある自由なデータ流通) の考え方が重要
- Authorized Public Purpose Access (APPA) による新たなデータガバナンスの概念の提案
- 個人、データホルダー、公益のバランスにより、データ活用を進める
- 公共性の高い産業の定義、ホワイトリストの作成
- 「新しい同意取得」について国民の理解が必要



出所："APPA: Authorized Public Purpose Access
WEF抄訳資料（世界経済フォーラム第四次産業革命日本センター提供）”

＜二次利用に焦点を当てた法整備＞

- フィンランド「健康と社会データの二次利用に関する法律」制定
- 企業が国家の管理データを活用することに法的根拠を付与
- 国家福祉研究所（THL）がデータ利用を一括許諾
- 処理の合理化、データへの迅速なアクセス、セキュリティ向上が目的
- 研究開発やイノベーション、教育、統計等の分野が対象



- 医療ヘルスケア分野の仮名加工情報をスムーズに取り扱うための新たな法の制定
- 個人情報を活用した公益性の高い研究・事業の定義や公益性の内容の明確化（国家としての「データガバナンスモデル」の明示と、Q&Aやガイドラインによる周知）
- 国民が自分のデータがどのように使われたか把握できるデータフローやデータ活用の透明性を高める施策の推進（データ管理者・利用者に対する信頼性の向上）

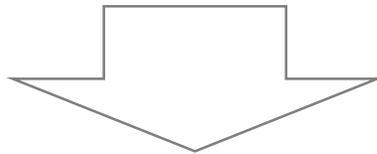


諸外国のアーキテクチャ（法制度やネットワークも含めた社会の仕組み・構造）に学びつつ、日本の法制度や社会とのマッチングや効果の想定などにより具体的な方策を実装

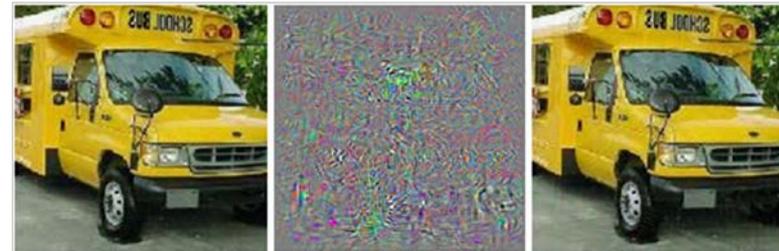
クオリティデータの収集と活用

<AI時代における多種多様なデータの利活用>

- 医療ヘルスケアデータは多種多様であり、それらは異なる基準、異なる環境、異なる考え方で収集されている
- 多様な産業での二次利用が期待されるが、データは二次利用を目的に収集されているわけではなく、二次利用者の目的にかなった質の確保は困難
- 教師データ等のデータソースに偏りがなく、どのようなデータが使われたかという透明性や、判定を導くに至った過程の説明可能性を求める声の高まり
- AIは人間には判別できないレベルのノイズデータが入っただけで判定を誤ることがある
- アノテーションやラベル付けの質も重要



深層学習 (Deep Neural Network) による誤認識の例



(左) 元のイメージ (中) 誤認識するように加えられた歪みイメージ (右) ダチョウと認識されたイメージ

出所 : Christian S. et al. "Intriguing properties of neural networks." (2014)

<クオリティデータを収集・利活用するうえでの新たな課題>

- 相互運用性の向上 (保健医療情報に関する国際標準規格FHIR*の普及推進)
- FHIRの他分野 (ゲノム、PHR等) への展開
- 二次利用者からの「求める質」に関する情報の発信
- 意図しないデータの混入を防ぐ対策の充実 (セキュリティ確保含む)
- アノテーションの質の確保、公平性、説明可能性、透明性への配慮

* Fast Healthcare Interoperability Resources

<PHRの標準化の現状>

- PHRの一部である検診・健診は、すでに自治体や健保等によって広く実施されており、継続的なデータを収集する体制がある
- しかし現状では不連続かつ分散して保管されており、本人でも収集困難なうえ、疾病罹患時にも活用されていない



検診・健診情報や日常生活における情報を
連結可能なかたちで蓄積し、個人の行動変容や医療に役立てる視点が必要

<2つの視点の標準化>

- フォーマットや項目に関する「データを格納するハコ」の標準化
- データ本体の「品質」に関する「データの品質確保のため」の標準化

測定者間差、機器間差、
デバイス間差、施設間差、消耗
品のロット間差、日間差、日内
変動、記録の正確性・・・



すべてのPHRに対して前処理や測定法等
の標準化を求めることは困難

<メタデータ(サンプリング・デバイスの種類・測定法などの付帯情報)の充実>

- データユーザーの求める質に対する適合性⇒2次利用者が判断できる情報の充実
- メタデータの記述様式やフォーマット、簡便化、機械可読化などの検討が必要

次世代健康医療記録システム共通プラットフォーム課題研究会（通称：NeXEHRs研究会）による、新たな電子カルテシステムの共通プラットフォームの設計が開始

NeXEHRs

3つの基本コンセプト Patient-centered, Sharable, Co-welfare

1. **本人主体管理**: 個人に基づく健康医療情報は医療提供機関単位ではなく、本人（患者等）単位で1記録とし、そのバックアップコピーを恒常的に預かる組織が運用されることを前提とする。
2. **本人・医療提供者間での情報共有**: 本人と医療提供者は、医療時に医療情報を共有する（明示的に拒否する場合を除く）。
3. **自他共栄**: より良い医療を開発して他の患者への診療にも将来貢献するために、仮名化した医療情報を安全に二次利用することを前提とする。

5つの実現方針 AI-BiCS

1. **P**: 患者・市民参画 (PPI: Patient and Public Involvement)
2. **AI**: AI、Automation、多様なヒューマンインタフェース (HI) の活用 (IoT、ウェアラブルデバイス、タッチデバイス、AI (DL・KP)、自然言語処理 NLP、音声認識、手書き認識、4K-8K画像、5Gネットワーク、ブロックチェーン技術、匿名化処理など)
3. **Bi**: BigData に対応 (多施設データ管理とゲノムデータ対応)
4. **C**: Cloud環境の積極的採用
5. **S**: 使える標準化 Standard の徹底した採用

出所：次世代健康医療記録システムのための共通プラットフォームの構築へ

- ✓ 普及率を高めるには、国の定める標準として展開することと導入インセンティブが必要
- ✓ 海外製IoT機器や国をまたいだ医療データの連結には国際標準を取り入れる必要

- 多岐多様の活用が求められるため、「相互運用性」の観点からの標準化や構造化の推進は不可避
- ウェアラブル機器やセンサー等、IoTからの機械可読かつ連続的なデータを医療ヘルスケアデータの中心に据える「発想の転換」も検討の必要あり
- 二次利用者が対象データの質を予め評価できるよう、データ収集法や測定法に関するメタデータの充実や、データ説明インデックスの標準化の推進が求められる
- アノテーションの質の確保や標準化、公平性、説明可能性、透明性に配慮したデータ収集など、AIの特性を踏まえた利活用の視点が重要
- 国際的に信頼される仕組みとなるよう、国際標準との整合性を確保する必要

医療ヘルスケア分野におけるデータエコシステム 構築に向けた重要方策

医療ヘルスケア分野におけるデータエコシステム 構築に向けた重要方策

重要方策① 医療ヘルスケアデータのアクセシビリティの向上

- ✓ 医療健康分野の仮名加工情報の取り扱いを定める「医療情報特別法（仮称）」の創設
- ✓ 公益性の高い産業・事業のクライテリアの明確化（ガイドライン・ホワイトリスト等）
- ✓ 個人が自分のデータへのアクセスや活用をコントロールできる環境の整備

重要方策② 医療ヘルスケア分野のクオリティデータ収集の推進

- ✓ PHRや医療情報の標準化の推進と相互運用性の向上
- ✓ レジストリ等における医産学官連携の推進と2次利用に向けたデータ品質の確保
- ✓ IoTやクラウドの活用によるready for AIなライフコースヘルスデータの収集推進

重要方策③ 信頼に基づくデータ循環と社会への成果・価値の還元

- ✓ セキュリティやプライバシーも含めた、データフローの信頼性を評価する仕組みづくり
- ✓ 国民と一体となった研究の推進と成果やメリットをタイムリーに還元する仕組みの実装
- ✓ 実効性と納得感の高い「データガバナンスモデル」の構築と国民啓発