



医療健康分野のビッグデータ活用研究会
報告書 vol.5

2020年6月

日本製薬工業協会
医薬産業政策研究所

無断転載を禁ずる

医療健康分野のビッグデータ活用研究会 報告書 vol.5

森田 正実

(医薬産業政策研究所 統括研究員)

佐々木 隆之

(医薬産業政策研究所 主任研究員)

中塚 靖彦

(医薬産業政策研究所 主任研究員)

本報告書は研究上の討論のために配布するものであり、著者の承諾なしに転載、複写・複製することを禁ずる。

本報告書に記された意見や考えは著者の個人的なものであり、日本製薬工業協会および医薬産業政策研究所の公式な見解ではない。

内容照会先：

日本製薬工業協会 医薬産業政策研究所

〒103-0023 東京都中央区日本橋本町 2-3-11 日本橋ライフサイエンスビルディング 7F

TEL: 03-5200-2681; FAX: 03-5200-2684

E-mail: opir-sp@jpma.or.jp

URL: <http://www.jpma.or.jp/opir/>

謝辞

本研究を実施するにあたり、エキスパート勉強会講師として、ご指導・ご協力をいただいた吉澤尚先生（弁護士法人漆間総合法律事務所）、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）、鈴木正朝先生（新潟大学法学部 教授）に謝意を表します。

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至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求められる。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や、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二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

目次

| | |
|--|----|
| メンバーリスト | 2 |
| 勉強会講師 | 3 |
| 1. はじめに..... | 4 |
| 1-1. これまでの研究会の活動について..... | 4 |
| 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..... | 5 |
| 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..... | 8 |
| 2-1. 医療ヘルスケアデータを取りまく法制度の現状と課題..... | 10 |
| 2-2. セキュリティ・プライバシー規制の現状と課題..... | 14 |
| 2-3. 新たなデータガバナンスの視点と概念..... | 19 |
| 2-4. 医療情報の二次利用に焦点を当てた法整備..... | 21 |
| 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..... | 22 |
| 3. クオリティデータの収集と活用..... | 23 |
| 3-1. データ駆動型社会でデータに求められる新たな「質」..... | 24 |
| 3-2. PHR、ライフコースヘルスデータの標準化..... | 29 |
| 3-3. 電子カルテ情報の質の向上に関する取り組み..... | 34 |
| 3-4. クオリティデータに関する期待と課題..... | 36 |
| 4. まとめ..... | 38 |

メンバーリスト (2019年4月～2020年3月)

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)

渉外部

白神昇平 アステラス製薬(株)

渉外部

鈴木 雅 田辺三菱製薬(株)

デジタルトランスフォーメーション部

小林典弘 塩野義製薬(株)

デジタルインテリジェンス部

(以上 4名)

勉強会講師

○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

京都大学 医学部附属病院 医療情報企画部

黒田 知宏 先生

世界経済フォーラム 第四次産業革命日本センター

藤田 卓仙 先生

山本 精一郎 先生

岡本 雅子 先生

上記の先生方には、本報告書内に用いた図表についても提供いただいた。

1. はじめに

1-1. これまでの研究会の活動について

2015年7月に立ち上げた「医療健康分野のビッグデータ活用研究会」も、2019年度で5年目の活動を終えた。本研究会では、政策研を研究主体として、製薬協関係委員会に協力を頂き、また必要に応じて業界外の企業やコンサルタントなどの有識者の参画も募り、それぞれの年度の研究会テーマの調査研究を進めてきた。研究に際しては医療健康分野のビッグデータの最新の状況を踏まえつつ、高い専門性と知識を有するエキスパートやオピニオンの方々を講師として招聘し、勉強会や意見交換の場を創ることにより、現実的な活用を進める観点から調査研究を行ってきた。

初年度と2年目は、レセプトや電子カルテといった臨床情報とゲノムを中心とするオミックスデータを連結させた、診療や創薬研究を含む医学研究に活用できるビッグデータや、EHR (Electronic Health Record) および PHR (Personal Health Record) の整備状況、また急激に進展する IoT による医療健康情報の拡大について調査を行い、我が国において活用を進めるための課題や方策を「報告書 vol.1」「報告書 vol.2」としてまとめた^{1,2)}。

3、4年目には、製薬産業としての具体的なデータ活用と実用化に向けた課題を抽出することを方針に掲げ、「医療データの活用」、「PHR の活用」、「AI 活用」、「人材育成」、「知財の取り扱い (戦略)」の5点を主要な検討課題と設定して調査研究を進めた。このうち、短期的に重要と考えられた「医療データの活用」については、製薬産業のバリューチェーンにおける医療データの活用およびその実現のための環境整備における課題と対策を「報告書 vol.3」としてまとめ、5つの提言を発信した³⁾。さらに「報告書 vol.4」では、上記5つの主要課題を並行して検討しつつ、中長期的に重要な課題であり、医療データの活用と切り離せられない「PHR の活用」をメインテーマとして検討を行い、PHR 活用を進めるための5つの提案を発信した⁴⁾。

このように、ビッグデータの活用について4年間調査研究を進めてきたが、この間にデータ駆動型社会に対する関心や理解が進み、我が国でも「Society 5.0」の取り組みやデジタルトランスフォーメーションなど、データ駆動型社会の実現に向けた動きも活発化してきた。そこで5年目(最終年度)となる2019年度は、価値を生み出す源泉である「データの要素」に着目した。そもそもデータ駆動型社会では、実際に使えるデータを生み出すことが重要であり、データへのアクセス性が高まり、活用が自律的・継続的に発展するためにどのような課題解決が必要かという視点が大切になる。よって、特にデータの「クオリティ」と

1 医療健康分野のビッグデータ活用研究会 報告書 vol.1 (2016年7月)

2 医療健康分野のビッグデータ活用研究会 報告書 vol.2 (2017年5月)

3 医療健康分野のビッグデータ活用研究会 報告書 vol.3 (2018年5月)

4 医療健康分野のビッグデータ活用研究会 報告書 vol.4 (2019年5月)

「アクセシビリティ」の要素を検討の中心テーマとして設定し、「データエコシステム⁵」の創出をいかに実現していくかという視点から調査研究を行った（図1）。

図1 医療健康分野のビッグデータ活用研究会のキーテーマ

| |
|---|
| <p>1年目のキーテーマ（2015年度） ⇒「研究会報告書vol.1（2016年7月）」</p> <p>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</p> |
| <p>2年目のキーテーマ（2016年度） ⇒「研究会報告書vol.2（2017年4月）」</p> <p>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</p> |
| <p>3年目のキーテーマ（2017年度） ⇒「研究会報告書vol.3（2018年5月）」</p> <p>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</p> |
| <p>4年目のキーテーマ（2018年度） ⇒「研究会報告書vol.4（2019年5月）」</p> <p>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</p> |
| <p>5年目のキーテーマ（2019年度）</p> <p>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</p> |

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

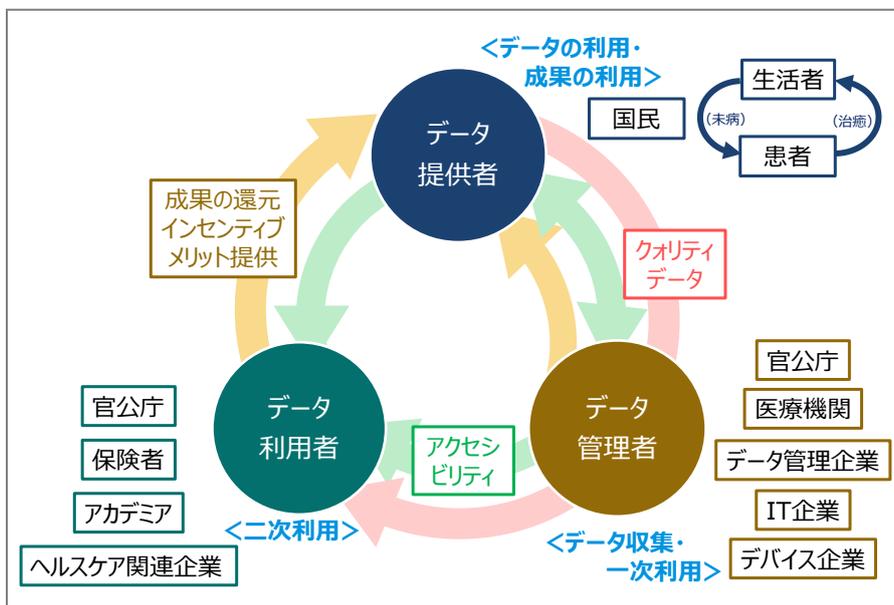
データ駆動型社会において、医療ヘルスケアの「データエコシステム」を形成していく際、今後主役となるのは「デジタル化された患者・生活者のデータ」である。特に、健康維持や予防、先制医療の視点からは、生活者が疾病に罹患するに至るまでの遷移データ（経年的な健診・検診データ、健康管理データ等）が、また医療の質を高める取り組みの観点からは、治療を行った後のデータ（アウトカム、予後、QOLの改善など）が重要なものとなる。そしてこれらのデータとともに、ゲノム等の生体分子情報や行動・生活情報、社会・環境情報といった”Determinant of Health”（健康の決定要因）に関する情報を付加することも求められる。まさに個人の一生涯のデータとしての「ライフコースヘルスデータ」を蓄積することが、今後のデータ駆動型ヘルスケアでは重要である。

こうしたデータを流通、連結させて社会や国民に対して新たな価値を提供していくためには、実際に使えるデータが生み出され、かつそれに対するアクセス性が高まったうえで、この仕組みが正しく循環し、価値を増大し続けることが必要である。これがデータの活用を源泉として自律的、継続的に発展する「データエコシステム」の概念であり、データ駆動型ヘルスケアを構築していくためには、このエコシステムに関わるステークホルダー（データ提供者、データ管理者、データ利用者）や社会そのものがメリットを享受しつづけられる仕

⁵ 医薬産業政策研究所、政策研ニュース No.59（2020年3月発行）
http://www.jpma.or.jp/opir/news/059/no059_10.html

組みを目指す必要がある（図2）。

図2 医療ヘルスケア分野におけるデータエコシステムの構成要素



ここで位置づけた「データ提供者」は、一般には患者も含めた生活者であり、国民全体がデータ提供者とも言える。「データ管理者」は、医療データであれば医療機関、国や保険者であり、PHRであれば企業や民間事業者も含まれることとなる。そして、収集されたデータを活用して新しい価値を創出するのが「データ利用者」であり、民間事業者、公的機関など多様なプレーヤーが考えられる。この図はデータの流れとそれぞれの役割を簡略化したモデルだが、「データ管理者」のなかにもデータを活用して価値を創出する「データ利用者」を兼ねるプレーヤーが考えられるなど、実際はより複雑である。

こうしたプレーヤーをつなぐ源泉は、質の確保されたデータ、すなわち「クオリティデータ」であり、そのデータを活用できる環境として「データアクセシビリティ」を確保することが「データエコシステム」実現の前提となる。そして、二次データ活用により得られた成果やデータ提供・管理に関するインセンティブをタイムリーにデータ提供者に還元するといった仕組みを社会実装し、かつそのサステナビリティを確保することが大切である。

図3には、データ駆動型ヘルスケアにおいてデータに求められる要素を整理した。上述のような多様なデータが収集されることはもちろん、①長期にわたる大量のデータを蓄積・管理する基盤があるか、②それらが連結・相互運用可能か、③一次・二次利用の目的に沿った質が確保されたデータであるか、④データ提供者がメリットを得られ、かつ安心してデータを提供できるか、⑤データ利用者や管理者へ新たな価値やソリューションを正しく、効果的かつ速やかにフィードバックできるか、といった「仕組み」の要素もまた重要となる。そして、データエコシステムが維持されるには、これらの「データの要素」が個々に成立するだ

けでなく、総体として循環フローのサイクルが回ることが大切となる。

図3 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」

| |
|--|
| <p>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</p> <p>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</p> <p>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）</p> <p>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</p> <p>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</p> <p>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</p> |
|--|

①と②は、これまでの研究会でも中心的なテーマとして検討を行ってきた点だが、③の「利用目的に応じたデータの質（Fit for Purpose）」とも切り離せない関係にある。加えて今後は、個々のデータに関するデータサイエンスの観点からの質だけでなく、産業をまたいだデータ活用の視点からの相互運用性の確保や、ブラックボックス性・アノテーションの質といったAIの特性を踏まえたデータの収集、さらにはデータ利用者がそのデータの質を評価できるような仕組みといった視点も求められる。

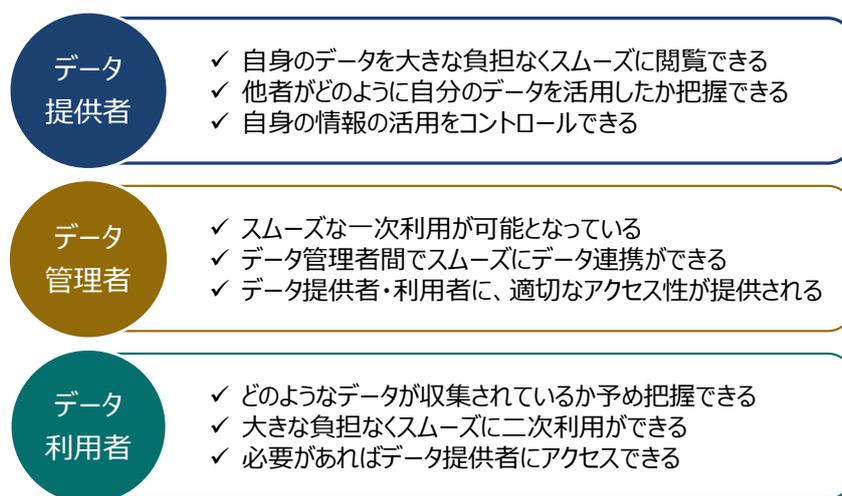
また、医療ヘルスケア分野のデータの多くは機微性の高い情報を含むことから、産業利用するにはデータエコシステムの「信頼性」が重要となる。すなわち、④に掲げたとおり、個人情報保護やセキュリティ確保に関する取り組みや法制度の整備はもちろんのこと、データ提供者がデータ活用に対して安心感を抱けるよう、データの利用過程や利用目的に関する透明性を確保し、データ利用者の適格性を継続的に評価していくことで、社会全体の信頼性を高める視点が欠かせない。そして、産業としてデータ駆動型ヘルスケアを育成していくためには、これらの仕組みがガラパゴス化しないよう、グローバルなハーモナイゼーションを念頭に議論を進める必要があり、⑤のようにデータの整備と活用を支えるインセンティブやメリット、情報や成果をデータ提供者が享受できる環境づくりも併せて必要である。

このような状況認識のもと、本研究会は5年目のテーマとして、医療ヘルスケア領域の「データエコシステム」を念頭におきつつ、「クオリティデータ」と「アクセシビリティ」の要素に特に注目し、調査研究を行った。

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

アクセシビリティの向上とは、まさに「データを活用できる環境を整備する」ということにほかならない。ステークホルダーごとのアクセシビリティについて考えた場合、大きく分けて3つの視点があると言えるだろう（図4）。すなわち、①データ提供者にとっては「自分のデータ」を閲覧・コントロールし、どのように活用されたか把握できること②データ管理者にとっては、容易なデータ収集や一次利用が可能で、データ管理者間（例えば医療施設間、医療機関と民間企業との間など）でスムーズなデータ連携が可能であること、またデータ提供者や利用者に、データに関する適切なアクセス性を提供すること、③データ利用者（二次利用）にとっては、どのようなデータが収集されているか予め把握でき、スムーズに二次利用が可能となり、さらに必要に応じてデータ提供者に直接・間接的にアクセスできること、がアクセシビリティ向上の主なポイントである。

図4 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



これらのポイントを実現するための方策は多岐にわたり、検討中のものやすでに実装されているものなどそのレベルは多様だが、大きく分けて下記の4点がアクセシビリティの向上に向け重要な点としてまとめられるだろう。

1) データ提供者への情報還元の仕組み

当研究会の報告書 Vol.4 で「PeOPLE (Person-centered Open Platform for Wellbeing)」や「情報銀行」について取り上げたとおり、データ管理者が収集したデータを、ポータルサイト等によりデータ提供者に還元できる仕組みが必要である。こうした情報還元が実現されるには、データ提供者の立場を踏まえてデータポータビリティの制度化を推進するとともに、例えばクラウド型 EHR (Electronic Health Record ; 電子化された医療情報) の普及や EHR と PHR との連結などにより、情報共有の観点からも整備が進まなければならない。

また、データ提供者がより使いやすいインターフェースの更なる充実も必要だろう。

2) データ利用者のアクセシビリティを高める法制度や基盤整備

日本では、医療情報に対するデータ利用者のアクセシビリティ向上が進まない現状がある。産業界での医療情報の研究活用を進める機運は高まっており、2018年には医療分野の研究開発に資するための「匿名加工医療情報に関する法律(次世代医療基盤法)」⁶も施行されたが、匿名加工により他の情報との連結が現実的に困難になる等の課題もあり、研究活用は準備段階である³⁾。また個人情報保護法の観点からは、公益目的による個人情報の取り扱いが不明瞭な点や、自治体の定める条例の不整合等の背景から、具体的な医療健康情報の利活用につながっていないという指摘もある。こうした法制度上の課題が解決されることも、アクセシビリティを高めるには重要であり、さらには二次活用を考慮したプラットフォームやデータベース構築といったアクセシビリティを高める基盤整備がキーとなる。

3) データ管理者・利用者の信頼性向上

データ提供者が安心してデータを提供できることが、データエコシステムの持続的な成長には不可欠である。そのためには、データ管理者・利用者が十分なセキュリティ対策を講じていること、プライバシーが保護されることがまず大前提にあり、さらにそれらをデータ提供者が確認できること、説明可能性や透明性が確保され、データ管理者・利用者の信頼性が外部評価されることが大切である。また、今後データ利用者が国外からアクセスすることや、データを用いて国外へソリューションを提供することなどを考慮すると、その信頼性は国際標準に則ったものとなるべきである。

4) データステークホルダーが共存できるガバナンスモデルの確立

1) から 3) に掲げた要素は、それぞれを個々に考えるのではなく、いかにしてデータ提供者・管理者・利用者の便益のバランスを確保するかという視点から、複合的に捉える必要がある。特に医療は、純然たるサービス産業と異なり、本人の意志にかかわらず誰にでも起こりうる「疾病」の管理に重点が置かれ、個人のプライバシーと生命に係る倫理性の高い領域であり、またそれ自体がインフラ的要素を有し、社会性、公益性の高い領域である。よって、データ提供者・管理者・利用者のいずれかに偏ったものではエコシステムを構成できない。バランスに配慮した利活用モデルを構築する必要がある。

このようにアクセシビリティの検討の切り口は多岐にわたるが、ここでは特に、個人情報保護やデータガバナンスに関連する動きについてレポートする。まず、セキュリティ・プラ

⁶ 医療分野の研究開発に資するための匿名加工医療情報に関する法律
https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=429AC0000000028 (2020/05/19 参照)

イバシー規制の欧米の状況について、日本と比較しながら概観し、次いで個人情報保護法の改正の動向や次世代医療基盤法の取組みについてまとめる。そして、これからの日本の制度を考える上で参考になるであろう、Authorized Public Purpose Access (APPA) の概念やデータ二次利用に焦点を当てた法整備などを紹介する。

2-1. 医療ヘルスケアデータを取りまく法制度の現状と課題

昨年 2019 年 1 月にスイス・ジュネーブで開催された世界経済フォーラム年次総会（ダボス会議）⁷⁾にて、安倍晋三首相がスピーチで「Data Free Flow with Trust（信頼性のある自由なデータ流通）」という言葉が提唱したことも記憶に新しいところである。また、首相は昨年 6 月に大阪で開催された G20 首脳会議の場でも「世界的なデータガバナンスが始まった機会として長く記憶される場としたい」と述べ、個人情報の安全性確保やデータ管理の高度化に向けた国際ルール作りを主導する考えを表明している。

実際、国際社会では情報通信技術の進歩とともにデータガバナンスのためのルール整備が進んでおり、そのルールには欧州、米国、中国といった国・地域によって特徴がある。例えば、GAF A などの巨大 IT 企業があるアメリカの場合は、個人データの自己管理を前提として、データを集積管理する企業などがプライバシーを配慮しつつ活用を進める対応が認められている。その一方、プライバシーや個人情報の取り扱いに厳正なヨーロッパでは「個人情報保護」の観点で重要視されており、「General Data Protection Regulation (GDPR)」⁸⁾によってより厳格なルール化が行われている。中国においては国家としてのデータ管理の重要性が強く打ち出され、中国政府は 2017 年に施行したサイバーセキュリティ法で、国家安全の目的から、国内で集めたデータの国外への持ち出しを制限し、国がデータを管理しデータを囲い込む政策をとっている。一方わが国では個人情報保護法の改正を中心に「GDPR」との整合など国際的なハーモナイゼーションを進める観点で検討が進められている。

個人情報保護法の改正等の動き

「個人情報保護に関する法律（以下、個人情報保護法）」は、2003 年に公布、2005 年 4 月に全面施行されたが、当初は住民基本台帳ネットワークシステムの導入に伴う住民基本台帳改正を受けて制定されたものであった。情報通信技術の発展や事業活動のグローバル化等の急速な環境変化により、制定された当初は想定されなかったようなビッグデータの利活用が可能となるなど個人情報の取り扱いや国際化の広がりにより大きな変化が出てきた。

そのような中、データ駆動型社会の形成に対する産業振興や個人の権利の保護を図ると

⁷⁾ 重要インフラにおける情報セキュリティ確保に係る安全基準等策定指針（第 5 版）
<https://www.nisc.go.jp/active/infra/pdf/shishin5.pdf>

⁸⁾ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC
(2020/5/19 参照)

いう目的や国際的なハーモナイズの動きに呼応して、2015年9月に改正個人情報保護法（2017年版）が公布され、2017年5月に全面施行となった。この際の議論において、国際的な動向も勘案しながら、個人情報保護法は3年ごとに見直しをすることになった。この3年ごとの見直しに対処すべく、昨年より次回改正の議論が本格化し、2019年12月13日に「個人情報保護法 いわゆる3年ごと見直し制度改正大綱」⁹⁾が発出され、2020年3月10日に「個人情報の保護に関する法律等の一部を改正する法律案」¹⁰⁾の閣議決定がなされ、通常国会に提出された。

医療データ活用における取り組み

国内における医療データの活用を考える上で、期待される取り組みの一つが、「次世代医療基盤法」に基づくインフラ整備である。2017年5月に施行された改正個人情報保護法によって、病歴等を含む個人情報が「要配慮個人情報」に位置付けられ、オプトアウトによる第三者提供が禁止されたことにより、医学研究等における医療情報の活用に支障が出るのではないかと懸念された。予てより診療行為のアウトカムを含んだ電子カルテを中心とした全国規模のデータベース構築の重要性も認識され、その構築に向けた取り組みについての議論もあった³⁾。

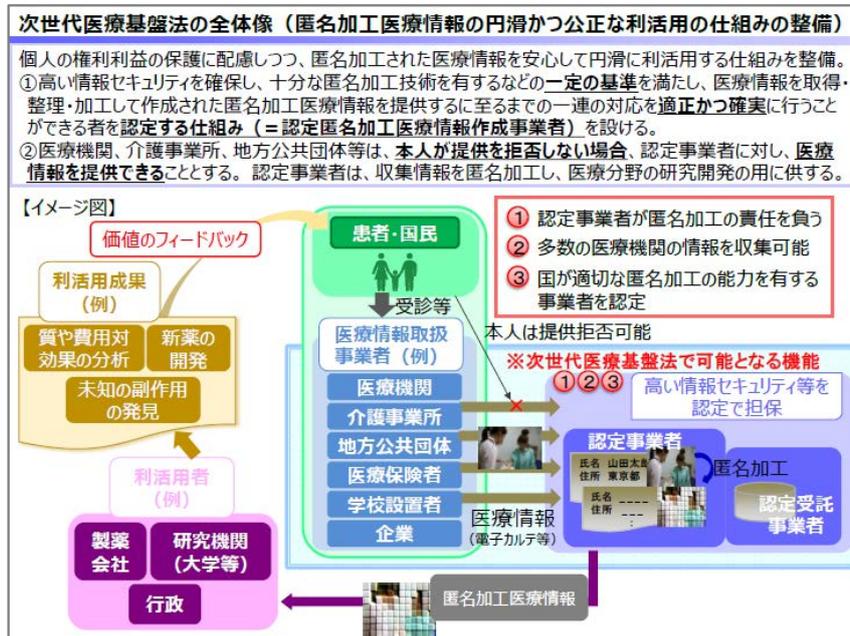
こうした背景のもと、医療分野の研究開発を推進するために2018年5月に施行されたのが、次世代医療基盤法である。この法律によって、高いセキュリティや匿名加工技術などの一定の基準を満たす事業者（認定匿名加工医療情報作成事業者（以下、認定事業者））を認定する仕組み¹¹⁾が定められ、さらにこの認定された事業者に対しては、医療機関等がオプトアウトで要配慮個人情報を含む医療情報を提供でき、匿名加工したうえで、その情報を医療分野の研究開発に提供できることが定められた（図5）。

⁹⁾ 個人情報保護法 いわゆる3年ごと見直し制度改正大綱 <https://www.ppc.go.jp/files/pdf/seidokaiseitaiko.pdf>

¹⁰⁾ 個人情報の保護に関する法律等の一部を改正する法律案 https://www.ppc.go.jp/files/pdf/200310_houritsuan.pdf

¹¹⁾ 「次世代医療基盤法に基づく事業者の認定について」、内閣府、<https://www8.cao.go.jp/iryoinintei/nintei/jigyousha/pdf/shiryoku.pdf>

図5 次世代医療基盤法の概要



出所：「次世代医療基盤法」とは；内閣府

2019年12月には次世代医療基盤法のいわゆる「認定事業者」が公表された。認定事業者の第1号は、京都大学を中心とした千年カルテプロジェクト²⁾を基盤とした「一般社団法人ライフデータイニシアティブ (LDI)」である。LDIは「健康・医療全般の情報の利活用者、参加施設へのサービス展開により、日本の医療発展に寄与できる事業を行うと共に医療の質・効率性や患者・国民の利便性向上、臨床研究等の研究開発、産業競争力の強化、持続可能な利用システム実現へ貢献すること」を事業の目的としている¹²⁾。

提携する医療施設ネットワークの拡大や、電子カルテ情報等の医療情報集積、それらデータのクレンジング作業などを進めているが、開始当初は「市場調査代行」ならびにデータベース研究による論文作成を支援する「学術支援」といったサービス業務を予定しており、体制が整い次第「製造販売後調査支援」、「匿名加工情報提供」¹³⁾などの本格的な研究開発支援サービスを提供していく予定である。製薬企業等の民間企業が利用者の中心となるが、現時点では実際にどのような項目、粒度のデータが利活用可能になるのか等の詳細が公になっていないため、データの内容やニーズなどを共有化するなど、利活用する民間事業者との事前の話し合いや連携が不可欠である。

¹²⁾ 2019年12月19日ニュースリリース、NTT DATA、<https://www.nttdata.com/jp/ja/news/release/2019/121900/>

¹³⁾ 特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報であって、もとの個人情報を復元できないようにしたもの。例えば、紙面に関する項目の全てを削除したり、年齢の表記を全て年代の表記に一般化して置き換えたり、特定の個人の情報を示す項目などを削除するなどの措置をしたもの。個人データに該当しない。

今後、上述の認定事業者のサービスにより医療情報の利活用が進んでいくことが期待されるが、この次世代医療基盤法の仕組みにはいくつか課題も指摘されている。

まず一点目として、特異性が高く個人を特定しうる記載内容（希少疾患、超高齢者の情報等を含む）や画像データ（脳のスキャンデータ、レントゲン、臓器画像等）等の匿名化の取り扱いについてである。このような情報では個人の特定を防ぐため、記述や画像を削除、もしくは丸め処理をするといった匿名化の処理が考えられ、提供データとしてオリジナルの情報が得られない恐れがある。特に、今後の希少疾患に対する研究開発や医療の層別化の進展、画像解析 AI など先進技術の活用を見越した場合、「狭くても深いデータ」の価値はますます増大していくと考えられ、特異性の高い医療情報の二次利用の活用（可能性）についてさらに議論していく必要があるだろう。

二点目としては、ゲノム¹⁴データの情報提供についてである。一般にゲノム情報はそれ自身が個人識別符号¹⁵に該当するため、匿名加工そのものできない。つまり、個人情報そのものであり、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できないこととなる。今後の医学研究、創薬研究においては、ゲノムデータと医療データ、及び日々の生活データ等とを連結し、疾患の発症、悪化のメカニズムや予後の予測、あらたなバイオマーカーの発掘等へとつなげることは重要な研究テーマであり、ゲノムデータの提供ができる仕組みについて更なる検討が必要であろう。

三点目には、認定事業者が取り扱うデータは数百施設から収集された一般の診療情報（電子カルテ情報等）のため、データの正確性や測定項目の欠落、検査値の単位や各コードの不整合等が存在することが考えられることである。この点は次世代医療基盤法の問題以前の、電子カルテ、EHR の標準化や構造化といったそもそものデータの質の問題であるが、匿名加工医療情報の利活用範囲が製造販売後や市場調査など特定の領域に限られてしまうことが懸念される。次世代医療基盤法の運用スキームの中で、疾患レジストリのように、収集データのフォーマットを事前に決めてテンプレートのようなものを用いデータ収集を行うなど、データの質を高める工夫も必要かもしれない。

認定事業者の活動が自律的に進展することが望ましいという点を踏まえると、製造販売後や市場調査の領域では、既にデータベース活用事業を行っているメディカル・データ・ビジョン（MDV）や JMDC などの民間企業もあり、データの質や量、及び認定事業者のデータ利用料も含め、既存事業者との差別化や優位性を図っていかなければならない。そのためにも、医療のアウトカム情報をよりタイムリーに提供できる体制や、ゲノム情報、PHR（特

¹⁴ 個人識別符号に該当するゲノム情報：細胞から採取されたデオキシリボ核酸（別名 DNA）を構成する塩基の配列ゲノムデータ（細胞から採取されたデオキシリボ核酸（別名 DNA）を構成する塩基の配列を文字列で表記したもの）のうち、全核ゲノムシーケンスデータ、全エクソームシーケンスデータ、全ゲノム一塩基多型（singlenucleotide polymorphism：SNP）データ、互いに独立な 40 箇所以上の SNP から構成されるシーケンスデータ、9 座位以上の 4 塩基単位の繰り返し配列（short tandem repeat：STR）等の遺伝型情報により本人を認証することができるようにしたもの

¹⁵ 特定個人の身体の一部の特徴を電子計算機の用に供するために変換した文字、番号、記号その他の符号であって、当該特定の個人を識別することができるもの

に健診・検診情報等)と連結した情報の提供を視野に入れるなど、疾患研究、創薬研究などにも活用できる医療情報提供のあり方を検討していくことも必要かもしれない。

「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の改正「個人情報の保護に関する法律等の一部を改正する法律案」で、最も注目する点としては、法2条9項に、「仮名加工情報」が新たに定義されたことである。

「仮名加工情報」は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に個人情報を加工して得られる個人に関する情報である。「仮名加工情報」は、その加工の程度に応じて、「個人情報である仮名加工情報」と「個人情報でない仮名加工情報」に分かれる。すなわち、「仮名加工情報」は、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ものであるが、他の情報と容易に照合でき特定の個人が識別できるものは個人情報に該当することになる。

しかし、「仮名加工情報」が個人情報を含むことはその取扱いを明確にすることで、前述した匿名加工情報の活用では取り扱えない、「特異性の高いデータ」や「ゲノムデータ」も取り扱う検討を行うことが可能となる。したがって、次世代医療基盤法が、個人情報保護法(一般法)に対する医療分野での特別法として制定されたように、「仮名加工情報」を取り扱えるような新たな法の制定も一つの方策であろう。

もちろん、「仮名加工情報」は個人情報を含むため、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が必要となる。この「適切な利用目的」は、個人情報保護法改正に伴って提示されるガイドラインやQ&A等の細目通知なども考慮しながら明確化していくことが期待される。その際、仮名加工情報を取扱いできる「対象をどうするか」と「どのような利用」に限定するかといった検討がポイントとなるだろう。すなわち、仮名加工を担うデータ取扱い主体である「対象」をあらかじめ指定された資格者(例えば医師や指定された研究者など)に制限し、社会性、公益性の高い「明確化された利用目的」に限定するなど種々の法定の制限の下に、「仮名加工情報」という個人情報提供のルール化を検討することで、医療情報の活用が推進できるのではないかと考える。

2-2. セキュリティ・プライバシー規制の現状と課題

デジタル化された医療ヘルスケアデータ(ゲノム、医療情報、生活情報などを含む)を最大限に活用し、Precision Medicine¹⁶⁾(精密医療)や先制医療を実現していくためには、それぞれをバラバラに活用するのではなく、統合されたデジタルヘルスの情報インフラとして構築していく必要がある。デジタルヘルスへの注力を拡大しているWHOにあっても、

¹⁶⁾ 医薬産業政策研究所、政策研ニュース No.45 (2015年7月発行)
<http://www.jpma.or.jp/opir/news/pdf/news-45.pdf>

2019年4月に発行されたデジタルヘルスガイドライン¹⁷⁾において、統合されたデジタルインフラが必要であることが言及されている。また、デジタル化された保健サービスの効果的な実施を支援するために組織され、世界23か国が参加する“Digital Health Partnership”においては、更に踏み込んで、デジタルヘルスインフラの重要な要素として「サイバーセキュリティ」「相互運用性 (Interoperability)」「エビデンスと評価」「ポリシーの環境 (Policy Environments)」「臨床・患者エンゲージメント」が提示され¹⁸⁾、それぞれについてホワイトペーパーも作成されている¹⁹⁾。

米国での動き

米国ではGAF A (Google、Apple、Facebook、Amazon)等の民間ITプラットフォームを中心として医療ヘルスケアデータのプラットフォーム化が加速している。例えばAmazonはMERCK、Accentureと連携してPrecision Medicineのためのクラウドプラットフォーム構築²⁰⁾をし、医薬品、バイオテクノロジーの研究開発の増大・多様化する研究データをプラットフォーム化により、迅速かつ効率的に研究を推進させ、新しい治療法の発見を促進させることを目指している。

一方Googleは、Deloitteと提携して人工知能ソリューション、臨床データウェアハウス、ゲノミクス、及び画像といった様々なデータを医療機関とライフサイエンス企業に提供するためのヘルスケアクラウドプラットフォームの構築を進め²¹⁾、早期診断等の推進を目指している。

また、IBMは豊富な医療データを医療AIへと拡大しヘルスケア関連企業IoT、ビッグデータを利活用可能なプラットフォーム構築を目指している。例えば、Watson for Oncologyでは医療データ、医療文献、専門家からのガイドライン等を分析し、エビデンスに基づく治療方針作成を支援する取り組みを行っている²²⁾。

以上のように民間の取り組みが中心ではあるが、国としての動きもある。例えば、米国がん研究所 (National Cancer Institute : NCI) は2017年にクラウドベースのインフラを整備したNCI Cancer Research Data Commons (CRDC)を発表している。本構想は医師、研究者などがゲノム・オミックスデータ等にアクセスし、がん研究とPrecision Medicineの加速化を目指すもので、関連するデータプラットフォームの整備を目標として掲げている²³⁾。

米国ではPrecision Medicine Initiative等の医療ヘルスケア関連の政府戦略をNIH等の

¹⁷ WHO Guideline : recommendations on digital interventions for health system strengthening

¹⁸ <https://www.gdhp.org/> (2020/5/19 参照)

¹⁹ https://www.gdhp.org/media-hub/news_feed/gdhp-reports (2020/5/19 参照)

²⁰ <https://newsroom.accenture.com/news/accenture-and-merck-collaborate-with-amazon-web-services-to-launch-a-research-platform-to-drive-innovation-in-drug-discovery-and-scientific-research.htm> (2020/5/19 参照)

²¹ <https://cloud.google.com/blog/topics/partners/deloitte-and-google-cloud-expand-their-alliance-to-bring-the-cloud-tomorrow-industries> (2020/5/19 参照)

²² <https://www.ibm.com/watson/jp-ja/health/> (2020/5/19 参照)

²³ <https://datascience.cancer.gov/data-commons> (2020/5/19 参照)

国家機関が支えつつ、GAF A 等の巨大プラットフォーマーやアカデミア、先端医療機関、製薬企業、IT 企業などが連携して、プラットフォームを構築している動きが顕著である。

医療ヘルスケアデータは機微性の高いデータとなるため、プラットフォーマーはプラットフォーム構築にあたりセキュリティ対策を施し、医療・医薬品産業における各種規制、法律を遵守している。また、各国のプライバシーポリシーや個人情報保護法に準拠し対応することが必要となる。

例えば Google においては「ユーザーとそのプライバシーの尊重」、「収集データ内容と目的の明確化」、「ユーザーの個人情報を販売しないこと」、「ユーザー自身の自己管理」、「ユーザーのデータ所有権」、「最高水準のセキュリティ技術導入」、「セキュリティ強化の模範化」の 7 つを「プライバシーとセキュリティ原則」に定めている。さらには医療に係るプライバシー、及びセキュリティ標準を一元的に規定している「HIPAA」の要件に準拠したシステムを構築している²⁴⁾。

欧州での動き

欧州では国の戦略として医療ヘルスケア分野のプラットフォームを構築する動きがある。2000 年ごろから世界に先駆けて EHR ネットワークの構築が取り込まれ、医療情報のプラットフォーム化が進められている。

フィンランドでは、新しい医療プラットフォームとビジネス創出を目的とした Personalized Health Finland プログラムという取り組みを 2018 年より行っており、個別化医療・予防の促進を目的として、バイオバンク、健康データ、診療データ、処方データ等様々なデータのプラットフォーム化が盛り込まれている²⁵⁾。このように国が中心となって医療情報を始めとする医療ヘルスケアのデジタル化を進めている事例はエストニアや英国、北欧諸国等、欧州には散見される。

EU におけるゲノム情報について、2018 年 4 月には EU 加盟 13 カ国（イタリア、エストニア、キプロス、スロバニア、スペイン、スウェーデン、チェコ、フィンランド、ポルトガル、マルタ、リトアニア、ルクセンブルク、英国）は「2022 年までに EU 域内で 100 万人ゲノムにアクセス可能とする」という目標を掲げたゲノム情報の越境アクセスに関する宣言書を採択²⁶⁾した。英国ではこれまでに英国全域の約 50 万人の遺伝子治療データを収集しているが、シーケンシングやバイオバンクの構築等において、既に実施された加盟国の施策を、この 13 カ国で活用・最大化することとしている。

併せて欧州では Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) という欧州各国の 250 以上のバイオバンクを統合した組織もあり、研究者が国

²⁴⁾ <https://safety.google/principles/?hl=ja> (2020/5/19 参照)

²⁵⁾ <https://www.businessfinland.fi/en/for-finnish-customers/services/programs/personalized-health-finland/> (2020/5/19 参照)

²⁶⁾ <https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders> (2020/5/19 参照)

境を越えて、信頼できる環境で膨大な量の研究データを保存、管理、分析、再利用し、臨床試験や臨床研究にリアルワールドデータを活用できるような環境整備が進められている。さらに、EUでは、「GDPR」や、標準化などの施策を進めるとともに、ゲノム情報の越境利用を含む European Open Science Cloud (EOSC) を展開²⁷⁾ し、ゲノムデータと医療データの統合を目指している。

医療ヘルスケアデータ利活用に関係する欧米の法規制

データエコシステムにおいては、データセキュリティやプライバシーに関する整備を始めとした法制度の対応なども併せて重要となる。ここでは米国と欧州のセキュリティやプライバシーに関する主な法規制・ガイドラインについて纏めた。日本の制度も参考として表1に掲載する。

表1 セキュリティ/プライバシーに関する主な法規制・原則 P: Privacy S: Security

| | 名称 | 施行時期 /改正時期 (直近) | 概要 | 罰則 | 分類 |
|----|---|-----------------------|---|----|-----|
| 米国 | HIPAA (The Health Insurance Portability and Accountability Act) 医療保険の相互運用性と説明責任に関する法律 | 1996年 /2006年 | 個人の医療データのプライバシーを守りつつ、データを医療の進歩に役立てることを目的とした法律 | 有 | P/S |
| | HITECH (The Healthcare Information Technology for Economic and Clinical Health Act) 経済的及び臨床的健全性のための医療情報技術に関する法律 | 2009年 | 医療ITの導入に際して想定されるようなセキュリティ・プライバシー条項の適用拡大や罰則強化を盛り込んだ法律 | 有 | P/S |
| | CSF (Cybersecurity Framework) 重要インフラのサイバーセキュリティを向上させるためのフレームワーク | 2014年 | IT世界のセキュリティリスクに特化したフレームワークであり、セキュリティ対策の強化を目的としたガイドライン | — | S |
| 欧州 | NIS Directive (Network and Information Security Directive) ネットワークおよび情報セキュリティ指令 | 2013年 | 各国での法制化及び、デジタルサービス提供者と交通、水、金融、医療等、重要なインフラサービス提供者にセキュリティ対策の強化と義務を求める指令 | — | S |
| | GDPR (General Data Protection Regulation) 一般データ保護規則 | 2018年 | 個人の権利を強化（情報への権利、アクセスの権利、処理の制限権、忘れられる権利、データポータビリティ権等）し、個人データ保護の調和と統一性の強化を図った規則 | 有 | P |
| 日本 | 重要インフラにおける情報セキュリティ確保に係る安全基準等策定 指針 | 2006年 /2018年 | 情報セキュリティに係るリスクへの必要な備えや、有事の際の適切な対処等を実現する指針 | — | S |
| | サイバーセキュリティ基本法 | 2014年 /2018年 | サイバーセキュリティに関する施策を総合的かつ効率的に推進するため、基本理念を定め、サイバーセキュリティ戦略の策定、及びその他当該施策の基本となる事項等を規定した基本法 | — | S |
| | 個人情報の保護に関する法律 | 2003年 /2017年 | 個人情報の適正かつ効果的な活用を目指し、かつ、個人の権利利益を保護することを目的とした法律 | 有 | P |
| | 次世代医療基盤法 | 2018年 | 医療分野の研究開発に資するための匿名加工医療情報に関し、健康・医療に関する先端的研究開発及び新産業創出を促進し、健康長寿社会の形成に資することを目的とした法律 | — | P |

米国の法規制

米国では医療情報の機密性を踏まえ、1996年に「The Health Insurance Portability and Accountability Act (HIPAA)」²⁸⁾を制定し、さらに医療IT化を目的に2009年に「The Healthcare Information Technology for Economic and Clinical Health Act (HITECH)」²⁹⁾を制定した。これらの法律では医療に係るプライバシー、及びセキュリティ標準を一元的に

²⁷⁾ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud> (2020/5/19 参照)

²⁸⁾ <https://www.hhs.gov/hipaa/index.html> (2020/5/19 参照)

²⁹⁾ <https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html> (2020/5/19 参照)

規定しており、医療情報の保護責任をもつ事業者がプライバシーとセキュリティを確保するための対策を講じることが義務づけられている。「HIPAA」は医療情報の電子化の推進と関連するプライバシー保護とセキュリティ確保について定めた法律であり、医療データを扱う事業者はこれに準拠する必要がある。

「HITECH 法」により促進されると想定されていた医療 IT の導入およびデータの活用の際に求められるセキュリティ・プライバシーへの対応策が、従来の「HIPAA 法」の内容では不十分であったため、HITECH 法に合わせ HIPAA も修正され、プライバシー条項の適用拡大や罰則強化が盛り込まれた。

米国国立標準技術研究所（National Institute of Standards and Technology : NIST）が 2014 年 2 月 12 日に公表した「Cyber Security Framework(CSF)」³⁰⁾ は、IT 世界のセキュリティリスクに特化したフレームワークであり、セキュリティ対策の強化を目的としたガイドラインである。米国のセキュリティについては、NIST によって様々規格化されているが、クラウドインフラを前提としたマルチステークホルダーエコシステムの構築と、重要なシステムに対しての直接攻撃や情報漏洩を避けるために多層の防御を行なう分野横断型の多層防御の考え方がとられている。この点は、集中クラウド層、フォグノード層、エンドデバイス層³¹⁾ の IT インフラ基盤を前提として多くの IT インフラ企業が関わってセキュリティについて適切な管理を行うこととなる。また、米国には脅威情報の収集、共有や警告などの業界別の情報共有分析組織（Information Sharing and Analysis Center : ISAC）が数多く存在しており、現在はさらに広範なカテゴリを網羅する情報共有分析機関（Information Sharing and Analysis Organization : ISAO）によって補完されつつある。なお、米国では国土安全保障省が業界横断的にセキュリティの管轄を行っている。

欧州の法規制

2018 年 5 月 25 日に「GDPR」が施行され、企業による個人データの取得利用が規制された。個人の権利を強化（情報への権利、アクセスの権利、処理の制限権、忘れられる権利、データポータビリティ権等）し、個人データ保護の調和と統一性の強化がなされた。「GDPR」が適用されない地域の企業に対しても、「GDPR」と同レベルのデータ保護義務が課されることとなる。「GDPR」は個人データの処理に適用され、対象者としては EU 域内に拠点をもちデータ管理者と処理者に適用される。また、EU 域内の個人に対し商品、サービスを提供し、個人データを処理、または監視するデータ管理者、処理者にも適用される。「GDPR」は EU 域内の統一ルールであり、加盟国全体への拘束力を持っている。各国はこの統一ルールに則り各々法規制を定める必要がある。

³⁰⁾ <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf> (2020/5/19 参照)

³¹⁾ 集中クラウド層、フォグノード層、エンドデバイス層：加工されたデータの活用や他サービスとの連携を担う集中クラウド層、デバイスから収集したデータの 1 次処理を担うフォグノード層、製品で構成されるエンドデバイス層

「Network and Information Security (NIS) Directive」³²⁾ は 2016 年 5 月 17 日に EU 理事会で採択され、同年 8 月 8 日に発効している。EU の全体的なサイバーセキュリティを高めるための指令で、EU 加盟国の重要インフラ（エネルギー、交通、金融等、）事業者など、社会基盤となるサービスを提供する組織にセキュリティ対策の強化と義務を求めるものである。ただし、本指令は法令そのものではなく、EU 加盟各国それぞれに対し国内法を定めるように指示するものとなる。そのため、各国における法制度については異なることに留意する必要がある。NIS 指令は欧州内のミニマムスタンダードであり、実際の法令設定や運用は各国当局に委ねられている。

例えば、EU 加盟国のドイツでは NIS 指令に先立ち 2015 年に「IT セキュリティ法」を制定³³⁾ し、サイバーセキュリティに関する最低限の基準を満たしている証明を情報セキュリティ庁から得ること、セキュリティ監査を定期的（隔年毎）にうけること、サイバー攻撃を受けた際には情報セキュリティ庁へ報告することを主要・重要インフラ事業者に対して要求している。違反時の罰則規定もある。英国も「ネットワーク・情報システム法」を 2018 年に定め³⁴⁾、重要インフラ事業者に対し効果的なサイバーセキュリティ対策を講じることを求めており、「GDPR」だけではなくセキュリティにも厳しい課徴金制度があるので注意が必要である。

2-3. 新たなデータガバナンスの視点と概念

医療ヘルスケア分野のデータ利活用についてさらに掘り下げた考え方として、世界経済フォーラム(WEF: World Economic Forum)から Authorized Public Purpose Access (APPA) という新たなデータガバナンスの概念が報告された³⁵⁾。本報告では、適切なデータガバナンスモデルの具体的な構成要素を示し、不適切なデータ利用を避けると同時に、データ利活用のあるべき姿の例として APPA が提示されている。

APPA は、個人の人権を尊重する方法として、必ずしも「個人の」オプトインでの同意だけに拠らない新たなガバナンスモデルを示しており、医学医療の発展や公衆衛生の向上等の、社会的な合意がなされた特定の公的な目的のためであれば、必ずしも明示的な個人同意によることなく個人の人権を別の形で保障し、データへのアクセスを許可することによって目的とするデータの活用やそれによる価値創出を実現するモデルである。

本報告では、利活用におけるガバナンスモデルの要素として、個人、データホルダー、公

³²⁾ JETRO 「EU 一般データ保護規則 (GDPR) について」 <https://www.jetro.go.jp/world/europe/eu/gdpr/> (2020/5/19 参照)

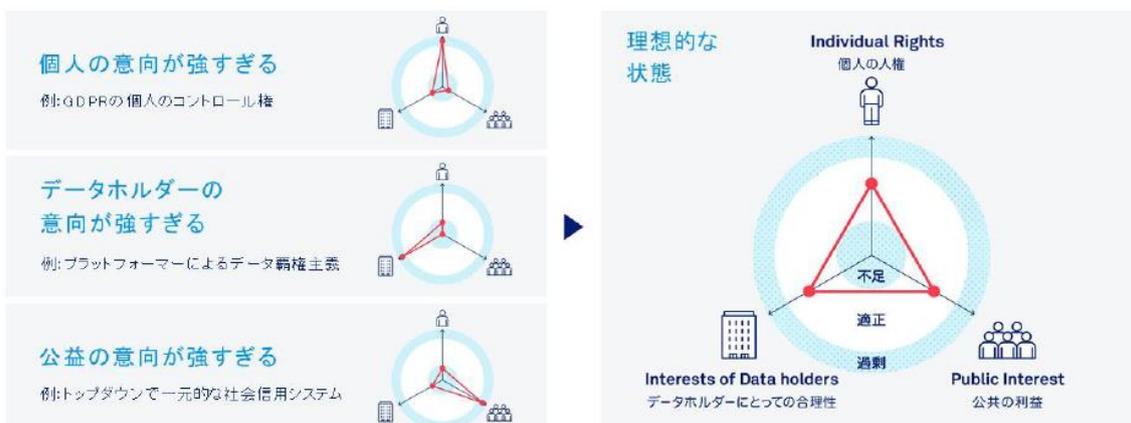
³³⁾ <https://eur-lex.europa.eu/eli/dir/2016/1148/oj> (2020/5/19 参照)

³⁴⁾ 諸外国のサイバーセキュリティ、及び個人情報保護に関する法律まとめ <https://www.j-cic.com/pdf/report/Cybersecurity-Privacy-Law.pdf>

³⁵⁾ APPA : Building Trust into Data Flows for Well-being and Innovation ホワイトペーパー (2020 年 1 月 17 日)、World Economic Forum <https://www.weforum.org/whitepapers/appa-authorized-public-purpose-access-building-trust-into-data-flows-for-wellbeing-and-innovation>

益の3要素が整理されており、APPAはそれぞれが過剰に偏る問題ケース（個人偏重、データホルダーの利益偏重、公益偏重）への対応法のひとつとして位置づけられている（図6）。

図6 Authorized Public Purpose Access



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）”

また、利用するデータや個人への影響に応じ、然るべきプロセスを踏む必要があるとし、個人同意に代わって、合意や第三者機関の活用も一つの手段として挙げられているが、今後、3要素の扱いに関する議論に基づき、より具体的な方法を示すことでAPPAモデルの信頼性が高まるものと考えられている。さらには、図7に示したようなデータを利用できる目的や事業者を予め定め、医療情報だけでなく健康情報も含めて利用されることを前提としてホワイトリストを作成し、公共性の高い産業の定義をしていくことも医療ヘルスケアデータを今後取り扱う上では重要な方策の一つであろう。

一方、経済的に持続可能なモデルとして構築されること、効率的な同意の取得、プライバシーリスクを最小化するテクノロジーの活用などの諸課題も挙げられており、今後更なる議論が望まれるところである。

図7 APPAを活用したデータ活用のフローチャート案



出所：世界経済フォーラム第四次産業革命日本センター提供

2-4. 医療情報の二次利用に焦点を当てた法整備

ここまで、医療健康情報を含めた個人情報の活用に向け、アクセシビリティを高めていく際に重要な視点として、セキュリティやプライバシーに関する近年の動向や、個人情報保護法改正に向けた動き、さらに公益性の観点を踏まえた新たなデータガバナンスモデルについて紹介してきた。一方で、データエコシステムの構成者はデータ利用者だけでなく、データ提供者、データ管理者を併せた三者であり、この三者が win-win の関係でメリットを持ちつつ、システムの自立と継続的活用ができることが、エコシステムとしては重要となる。ここでは、この三者にメリットがもたらされるよう、法律面からのバックアップを図った事例として、フィンランドを挙げたい。

フィンランドでは、2019年4月に、「健康と社会データの二次利用に関する法律」が制定された³⁶。フィンランドは早くから医療情報の電子化に取り組んできた国であり、社会福祉と医療ヘルスケアの分野において包括的かつ高品質なデータリソースを有している。民間企業も交えた AI 等による解析も試みられており、2017年には「CleverHealth Network」が立ち上げられ、IBM や Google といった民間企業も一体となり、画像データからの脳出血の検出、早産児の重度細菌感染の早期発見といったプロジェクトが実施されている。

他方、「GDPR」の施行により個人情報保護の取り扱いがより厳しくなったことから、それまでのフィンランド独自のデータ保護法に代わり、「GDPR」を補完する新たな国内法の制定が望まれていた。そこで、研究開発やイノベーション、教育、統計、ナレッジマネジメントといった目的での二次利用に対し、健康および社会福祉データを使用するための統一的なルールを策定して処理を合理化すること、およびデータへの迅速なアクセスを可能にし、セキュリティを向上させることなどを目的として制定されたのが、この「健康と社会データの二次利用に関する法律」である。本法律が施行されて以降、まずデータ利用の許諾は国家福祉研究所 (THL) が一括して行うこととなった。また、ナレッジマネジメントの法的根拠が明確になり、サービスプロバイダが登録データを将来独自に使用して新規な解析を行ったり、あるいは企業が国家の管理データを活用したりすることにも、法的根拠が与えられた。

この取り組みは、システムのステークホルダーである国民、研究者、サービスプロバイダ三者にとって、メリットが大きい。まず国民（データ提供者の立場）については、効果的で科学に裏打ちされた、より良い医療ヘルスケアサービスをタイムリーに享受できるメリットがある。また研究者（民間での研究も含む；データ利用者の立場）にとっては、既に統合されたデータにアクセスできるというメリットとともに、データ活用の許認可が短縮され、迅速な研究の開始が可能となるうえ、認可機関のレベル差を気にしなくても済むという利点がある。サービスプロバイダ（データ管理者の立場）にとっても、国家のリアルタイムな社会福祉データを利活用できる法的根拠が与えられ、新たなサービスの開発や提供が容易

³⁶ https://stm.fi/en/article/-/asset_publisher/uusi-laki-mahdollistaa-sosiaali-ja-terveys-tietojen-tehokkaan-ja-tietoturvallisen-kayton?_101_INSTANCE_yr7QpNmlJmSj_languageId=en_US (2020/5/19 参照)

になるというメリットがある。

ただし、この法律が広く理解され、また活用されるには、「信用 (trust)」が必須であると Findata のディレクターである Seppanen 氏は述べている³⁷⁾。すなわち、個人のデータセットがセキュアにリモートユーズされること、データセットが個人を直接同定する識別符号を除いた状態で構築されること、第三者機関により法律のコンプライアンスが監督されること、データ運用環境のセキュリティが査察体によりコントロールされることが求められるとされている。日本でも、データガバナンスをどのように推進していくか議論が必要とされるなか、透明性や信頼をどう定義し評価するか、また評価機関の適格性をどう担保するかといった課題も、今後討議していく必要があるだろう。

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

ここまで述べてきたとおり、グローバルには個人のデータをいかに医療やヘルスケアに活かしていくかという視点から、データ利活用に向けた取り組みが進んでいる。我が国においても医療情報を匿名加工によって研究開発に活用する取組みを進めようとしているが、加工情報の運用や目的の限定化といったアクセシビリティの課題が提起されており、仮名加工情報を取り扱えるような新たな法の制定も考えうる方策のひとつではないだろうか。

また、個人情報保護法改正の動向や APPA の概念からもわかるとおり、健康医療に関する個人情報を二次利用する場合は、利用する目的が公益性の高い研究や事業かどうか、公益性と個人情報のバランスという点が焦点となっている。データへのアクセシビリティを高める上で、データガバナンスと国民の権利及びコンセンサスが重要なことは、前述したとおりである。例えば個人情報を活用した公益性の高い研究・事業の定義や社会利益の内容を明確にし、具体的に Q&A やガイドラインとして周知していくこと、また国民が自分のデータが実際にどのように使われたか把握できる仕組みを構築するとともに、データ管理者・利用者に対する信頼を国民に高めていくことなどが、個人情報の更なる活用には必要であろう。そのためには、先行する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、日本の法制度とのマッチングや効果の想定などにより、日本で展開していく具体的な方策を実装していくことが求められるのではないだろうか。

³⁷⁾ HIMSS Europe 2019 (2019/6/11-13 開催)

3. クオリティデータの収集と活用

データ駆動多型社会の実現には、多様なデータを収集、連結することが必要となるが、データ提供者が多様であると、生まれるデータの質もまた多種多様となる。また、このような多種多様なデータの収集において、二次利用者は一次利用者の収集したデータを利用する立場にあり、二次利用者の目的に適った質の確保は困難である。基本的にはデータ提供者やデータ管理者の現場での活用（一次利用）の質を上げることを優先に、二次利用の目的に合致する質が整備されたデータを入手するという流れが必要となる。

たとえば、医療データの場合、入手するデータの中心は、疾患治療や費用の償還を目的とした電子カルテ情報やレセプト情報、画像データなどとなる。PMDAのMID-NETプロジェクトはこのような医療現場の診療に資する情報を収集し、安全性評価という目的に対して二次活用のプラットフォームを構築しているが、収集データの質の点で多くの課題が指摘され、データ統合やデータクレンジング等に多大な労力が必要であったとされる。このような状況を少しでも改善していくためにも、診療で発生するデータの質そのものを向上させる取り組みも必要である。医療（介護・予防も含む）の情報共有化、診断診療等のAI支援、Learning Healthcare System (LHS)³⁸⁾による医療改善など広く医療現場でのデータ利用を進めることもその一つの方策である。

しかしながら、データ利用者としての製薬企業の立場では、創薬ターゲット探索などの研究開発のクオリティを確保するため、一般的な診療データだけでなく、さらに品質の高いデータセットを取得する必要がある。⁴⁾ 疾患研究や医学研究などで前向き研究への参加をする等により、二次利用を考慮したデザインで集積されたデータの活用ができるよう、基盤整備を進めることが望まれる。Clinical Innovation Network(CIN)³⁹⁾プロジェクトに代表される疾患レジストリの活用やバイオバンクの活用、また健常者コホート研究である東北メディカルメガバンク (ToMMo) のデータ活用など、二次利用の視点からもデータのクオリティが期待できる基盤整備も進みつつある。

さらに、これからのヘルスケア分野では個人の情報を連結したPHRの活用の重要性が高まっている。近年、「クオリティデータ」という語が注目されるようになってきた。デジタル化を背景に大量のデータが取得できるようになってきたが、AIの利活用や質の高い医療ヘルスケアの実現を目指す際、単にデータを集めてきただけでは質の高い解析やソリューションの創出には結びつかず、「クオリティデータ」の利用基盤を創らなければかえって利用者に不利益を生じさせる懸念がある、という点も指摘されている。

本項では、今までの報告書と内容の重複を避け、まずデータ駆動型ヘルスケアデータに求められる新たな「質」の観点にどのようなものがあるか、簡単にまとめた。さらにこれからのデータ活用で重要性を増すライフコースヘルスデータ、特にPHRに関する標準化の現状

³⁸⁾ Institute of Medicine, “The Learning Healthcare System: Workshop Summary” (2007), <https://www.ncbi.nlm.nih.gov/books/NBK53494/>

³⁹⁾ Clinical Innovation Network, <https://cinc.ncgm.go.jp/>

や課題を整理した。そして医療ヘルスケアデータの中核をなす電子カルテ情報の質の向上に関する取り組みを紹介する。

3-1. データ駆動型社会でデータに求められる新たな「質」

これからのデータ駆動型ヘルスケアにおいて、ライフコースヘルスデータの重要性が増していくことは間違いないが、そのソースはまぎれもなく患者・生活者であり、データの収集者・管理者は、医療機関をはじめ、保険者、ヘルスケア企業（DTC ゲノム企業、フィットネス企業、デバイス企業、アプリ提供企業等）など極めて広範となる。「医療ヘルスケアデータ」とひとくくりに表現されるデータであっても、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現状がある。つまりデータ利用者の立場から考えると、広範なデータを統合利用する場合には「質や精度が異なること」を念頭に置く必要があり、また統合して解析を進めて行く上での課題となる可能性がある。また、今後活用が期待されるAI、特に深層学習では、判定に至る過程がブラックボックスであるという課題も提起されており、その計算過程を可視化する技術なども研究が行われている。さらにはAIで多量のデータを解析する際、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⁴⁰⁾、またその誤りを判断するのが困難な例があることも知られており、AI利活用においては新たなデータ品質の担保の観点が必要である。総務省から2018年に発表された「AI利活用の10原則（案）」においても、AIに学習させるデータの質に留意する「適正学習の原則」が盛り込まれており、データの「質」の確保はデータ駆動型社会における重要課題と認識されているといえる⁴¹⁾。

データの質に関する評価軸と「程度」

データの質に関する考え方には様々なものがあり、実世界の実体を正しく反映しているデータが高品質なデータだという考えもあれば（データ管理者（収集者）や現場でのデータ活用の視点）、データの利用目的・用途に適しているものが高品質なデータだという考えもある（主にデータ二次利用者の視点）。ここでは、収集したデータを解析・分析を行う立場からデータの質について整理しているものとして、国際データマネジメント協会英国支部の定める6つの主要評価軸を紹介する⁴²⁾（表2）。これらの項目自体はいずれも、収集・管理されるデータが実世界の実体をどのように反映しているか、ならびにそれが正確に表現されているかという「データ管理者（収集者）側」の視点に立って設定されているが、各評価軸の項目をどの程度許容するかはデータを利用する側の要求に依存するとされている。例えば適時性に関して言えば、“The degree to which data represent reality from the

40 Christian S. et al. “Intriguing properties of neural networks.” (2014) <https://arxiv.org/abs/1312.6199>

41 「AI ネットワーク社会推進会議 報告書 2018」情報通信政策研究所

42 “THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT” (DAMA UK, 2013)

required point in time.” すなわちデータの現実性に関する「程度」をどのようにデータ利用者が求めるかが重要となる。データがタイムリーに収集されることは、データ管理者（収集者）側にとって、データの質を担保するうえで重要な点ではあるが、それがどの程度重要か（今この時点でのデータが重要なのか、昨日のデータで良いのか、3か月前のデータで良いのか、あるいはバラツキがあっても良いのか等）は、データ利用者側が自身のデータ利用目的に応じて考える性質のもの、となる。ここからわかるとおり、データ駆動型社会で幅広くデータを活用していく際には、目的に応じた「適正な品質」がどのようなものかをデータ利用者側が提示していくこと、またデータ管理者（収集者）もある程度データ利用者のニーズを把握してそれに応えられるようにすることが、より重要となってくると考えられる。

表2 データの品質に関する6つの評価軸

| 評価軸 | 定義 |
|-------------------|---------------------------------------|
| Completeness(完全性) | 保管されたデータが、潜在的な全データに対して100%網羅されていること |
| Uniqueness (一意性) | 特定の対象が一度以上記録されないこと |
| Timeliness (適時性) | データが要求された時点で、そのデータが現実を代表していること |
| Validity (妥当性) | データが定義された構文規則（フォーマット、タイプ、範囲）に準拠していること |
| Accuracy (正確性) | データが記述されている実世界の対象またはイベントを正しく記述していること |
| Consistency (一貫性) | 一つの定義に対して複数の表現の差異がないこと |

ISO8000 にみる、データの質の新たな側面

データを最大限に活用し、従来の産業領域が連合化したデータ駆動型社会にあっては、多様なステークホルダーがデータ二次利用者となり得る。この場合、データ収集者にとっては、「自分ごと」としてのデータマネジメントの観点からの質の確保だけでなく、「他人ごと」すなわちデータが二次利用者にとって利用されることを念頭にデータの質等を確保していく必要が発生する。

近年はこのような考え方に基づいた情報の構造化に関する国際標準も整備されつつある。例えば、データの質に関する国際標準として、これまでもビジネスプロセスに関する品質・マネジメントを対象とした ISO 9000 が広く国内外で普及してきたが、新たに整備されている ISO 8000 では、組織間、システム間で情報交換する際のデータ品質要件、評価の方法やプロセスといった標準が定められ、相互運用性の向上が試みられている。ISO 8000 シリーズの各規格（図8）は、Part 1（概要）が2011年に発行されたのを皮切りに、2018年現在 18 のパートが発行済または開発中とされている⁴³⁾。欧州を中心に推進されている Industry 4.0 も、掛け声や概念だけではなく、国際標準群を整備することによりこれを実現しようとしており、標準群にあわせてデータの品質が考えられている。すでに ドイツ・鉄道ネットワーク、韓国・政府機関、サウジアラビアにおけるサプライチェーンなど、海外で

⁴³⁾ 「ISO 8000 データ・クオリティの国際標準化 - Society5.0の実現に向けて（蒲生秀典）」STI Horizon（2018）Vol.4 No.4

は ISO 8000 に準拠したデータ運用は広がりを見せている。

図8 ISO 8000シリーズの構成と主な規格（開発中含む）



出所：ISO 8000 データ・クオリティの国際標準化 – Society5.0の実現に向けて」

医療ヘルスケアデータの相互運用性

医療ヘルスケアデータを相互運用可能にする取り組みには、どのようなものがあるのだろうか。現在、医療情報標準規格としてもっとも注目されているのが保険医療情報の相互運用性に関する標準規格である“Fast Healthcare Interoperability Resources (FHIR)”である。FHIR の特徴として、データの共有性向上を目的としている点、API の利用によってサンプルで効率的にシステム間の情報共有が可能である点、短期間で開発・導入が可能である点、既存システムの接続システムとして運用することで、既存システムを 存続させたままデータ交換が可能である点などがある。

米国では医師会が“Integrated Health Model Initiative (IHMI)”を推進しており、データ交換のための標準の開発等を目的とし、クラウドベンダーや Digital Therapeutics 企業、liquid biopsy 提供企業、医療保険サービス提供企業などあらゆる業界から参加する 28 の協力組織とともに、医療ヘルスケア分野のデータ統合を目指して活動している⁴⁴⁾。これらの活動には政府も関与しており、例えば 2018 年 8 月には、ホワイトハウスにおいて、FHIR をベースとした API に関するカンファレンスが開催され、大手テック企業 6 社 (Amazon、Microsoft、Google、IBM、Oracle、Salesforce.com) が、クラウドと人工知能技術に重点を置いて医療データの相互運用性を実現する共同声明に署名した⁴⁵⁾。この他にも、HIMSS (Healthcare Information Management Systems Society) においては“Nordic Interoperability Project”が 2018 年に発足し、国境を越えたデータ統合が推進されているなど、国をまたいだ健康医療データを関連産業全体で活かすべく、相互運用性向上の取り組

⁴⁴ <https://www.ama-assn.org/amaone/integrated-health-model-initiative-ihmi> (2020/5/19 参照)

⁴⁵ <https://www.mobihealthnews.com/content/amazon-google-ibm-microsoft-oracle-and-salesforce-pledge-remove-interoperability-barriers> (2020/5/19 参照)

みが各所で進められている⁴⁶⁾。

FHIR は医療以外の分野における医療ヘルスケアデータの利活用においても、標準規格としての運用が推進されている。例えばゲノムデータの臨床情報との統合利用に関しては、Genomic Alliance for Global Health (GA4GH) により、PHR 分野におけるヘルスケアデバイス等の相互接続を推進するためのガイドライン展開については NPO 法人 Personal Connected Health Alliance (PCHA) により、それぞれ FHIR の活用が推進されている^{4,47)}。

GA4GH は、「ゲノムデータと臨床情報の統合・活用によるヒトの健康増進」をミッションとする国際的な非営利アライアンスであり、2013 年の活動開始以降、71 カ国から 500 以上の団体が参加して活動している。表 3 のとおり、データ利用者の利便性を高めるために標準化をはじめ様々な取り組みがロードマップとして策定されており、臨床データの交換には FHIR が採用されている。

PCHA は、医療機器やヘルスケアデバイスのデータ通信方法に関するガイドライン「Continua Design Guidelines」を作成し、ガイドラインに準拠したデバイス認証を行うなど、相互運用性の推進活動を行っている。糖尿病、心不全、高血圧、COPD、ヘルス&フィットネスなど、分野ごとにデバイスの種類や相互接続に関するガイドラインが設定されているのが特徴であり、FHIR を作成している HL7 とも連携しながら活動を推進している。

表 3 GA4GHの掲げる主なタスク

| タスク | 概要 |
|-------------------------------|---|
| データリポジトリサービス(DRS)API | クラウドやプラットフォームに関係なくデータを参照してアクセスするための共通の方法を作成 |
| データ利用オントロジー | データセットにオントロジータグをつけることで、目的の使用方法に基づいてデータセットの自動的な検出を可能とする |
| 臨床データ交換 (FHIR / Phenopackets) | オントロジーおよび用語が臨床情報を取り込むための標準的なデータ概念定義を提供 高度な臨床表現型情報および深い臨床表現型情報を交換できるように情報モデルを提供 |
| 変異の表現 | 変異の表現のための標準化された拡張可能なデータモデルとメッセージスキーマ仕様 |

出所：GA4GH戦略的ロードマップより作成

AI からみたデータの質

AI の利用からみたデータの質の観点も重要である。機械学習の第一人者であるイアン・グッドフェローらのグループは、2014 年、深層学習 (Deep Neural Network) に知覚できない特定の摂動を適用することにより、ネットワークに画像を誤分類させることが可能であることを示した⁴⁰⁾。例えば、図 9 の左右の写真は人間にはバスと認識されるが、中央の画像を学習データとして加えられた右の画像については、ネットワークはダチョウと認識

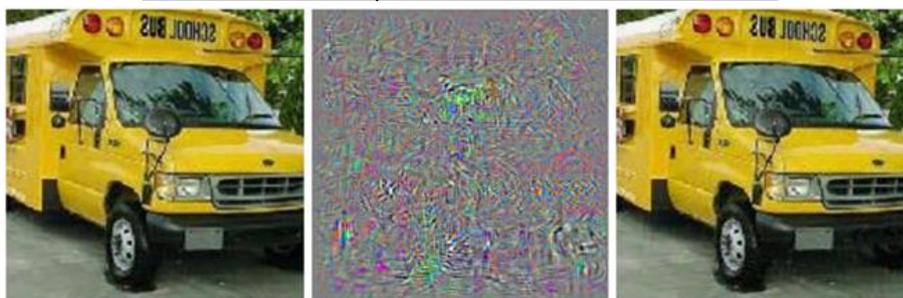
⁴⁶⁾ <https://www.himss.eu/communities/himss-nordic-community> (2020/5/19 参照)

⁴⁷⁾ <https://www.ga4gh.org/> (2020/5/19 参照)

する。この事例のように、AI には AI の性質を考慮した「データの質」が必要であり、また、誤ったデータをもとに為された判定を人間が「誤り」と正しく判断できるような仕組みも必要となっている。総務省の「AI ネットワーク社会推進会議 報告書 2018」においても、AI の学習等に用いるデータの質の観点として、AI の特性や用途を踏まえて学習データの質に留意する必要があること、AI が不正確または不適切なデータを学習することにより、AI (のセキュリティ) に脆弱性が生じるリスクがあることが指摘されている⁴¹⁾。

またこれに加え、AI のトレーニングに用いる教師データのラベル付けの質も、AI の質に影響を及ぼす。これには、間違っただラベルを付与してしまうかどうかという点にとどまらず、例えば医療画像であればどの範囲をがん組織とするか、正常組織の範囲も含まれてしまっているか、という「アノテーションの質」の観点もあれば、訓練データに適応しすぎ、評価データに対する正答率が低くなり汎用性が失われる、いわゆる「過学習 (over fitting)」の観点もある。すなわち、AI を使用する場合には AI に適したデータを確保する必要があること、またそのデータを用いて得られた結果が妥当であるか、また解析手法が妥当であったかを、現時点では人間が確認する必要があるといえるだろう。

図9 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)

さらに近年 AI の「FAT (Fairness (公平性), Accountability (説明可能性), Transparency (透明性))」も、データの質の向上に寄与するものとして、またデータの質を見える化するものとして注目されている^{48,49)}。例えばデータの Fairness の観点では、メラノーマの診断 AI では、教師データの人種が偏っていたため、患者の肌の色によっては判定精度が向上しないという課題を抱えていた⁵⁰⁾。また、Accountability については、「XAI (Explainable AI)」という概念を米国の DARPA (Defense Advanced Research Projects Agency、アメリカ国防高等研究計画局) が提示⁵¹⁾ して以降、急速に研究が進んでいる。例えば Google は同名の

48 "ACM Conference on Fairness, Accountability, and Transparency"

49 『「人間中心の AI 社会原則」及び「AI 戦略 2019 (有識者提案)」について』内閣府 (2019/4/17)

50 "Machine Learning and Health Care Disparities in Dermatology" *JAMA Dermatol.* 2018;154(11):1247-1248

51 <https://www.darpa.mil/program/explainable-artificial-intelligence> (2020/5/19 参照)

クラウドツールを発売し、どの特徴がモデルの予測結果にどれだけ貢献しているかを確認することが出来る⁵²⁾。Transparencyについても、どのようなデータセットが学習に使われたかというデータの透明性の観点に始まり、機械学習モデルの透明性（アルゴリズムやそのアップデート）や、データ提供者がそれを確認できるかという観点も存在する。

以上のように、AI が私たちの世界で普及していくには、それが「実世界で信頼できる AI」である必要があり、特に医療ヘルスケアでは高い信頼性が求められる。それゆえ、「データの質」を確保する取り組みも、AI の信頼性に直結するものとして捉えることが、今後は不可欠である。

3-2. PHR、ライフコースヘルスデータの標準化

ウェアラブル機器やセンサーをはじめとした IoT の進化や通信環境の発展に伴い、PHR の蓄積環境が整備されつつあり、医療ヘルスケアに関する多種多様なステークホルダーが、PHR を活用してソリューションを提供することが想定されている。本項では、今後利活用が広がるであろう PHR の標準化について着目し、特にデータの「質」からどのような対処が必要となるか検討した。

PHR の標準化の現状

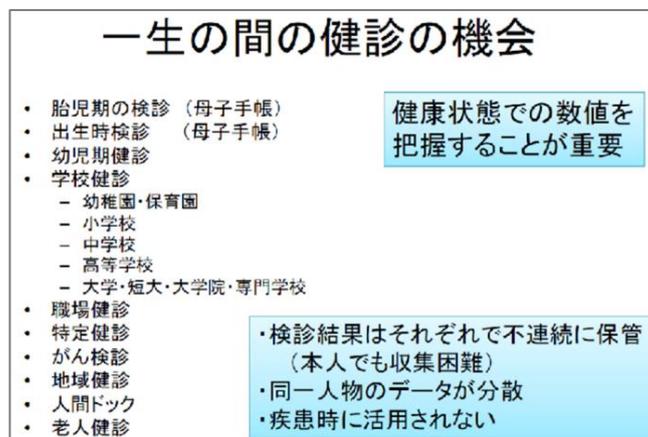
PHR の範囲は非常に広範であり、すべての PHR の標準化を同時並行的に進めることは困難である。標準化の取り組みも分野により大きな乖離があるが、比較的標準化が進んでいる分野として、検診・健診データがある。

PHR の情報の一部である検診や健康診断は、すでに国や自治体、企業等によって広く実施されており、ライフロングなデータを収集する体制が整っている。ただし個別の健診・検診結果はそれぞれが不連続かつ分散して保管されており、デジタル化されていない情報があるなど、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という現状がある（図 10）。そこで現在は、データ収集体制がインフラとして整っているこの検診・健診について、標準化の取り組みが進められている。例えば健診関係 10 団体で構成される日本医学健康管理評価協議会は、健診標準フォーマットを配布、幼少期から老年期に至る健診（検診）データの一元管理を実現していこうとしている⁵³⁾。この健診標準フォーマットの特徴として、すべての健診への対応が可能であること、受診者毎に独立したコードで表現されること、画像所見なども用語の標準化が目指されていること、などがある。標準フォーマットへの変換ツールは、健診事業者が作成した CSV ファイルを標準フォーマットへと変換可能であり、日医総研より無償配布されている。

⁵²⁾ <https://ai.google/> (2020/5/19 参照)

⁵³⁾ 「健診標準フォーマットについて」 <https://www.kenshin-hyojun.jp/>

図10 健診機会とその課題



(国立保健医療科学院研究情報支援センター 水島先生提供)

これとは別に、国立保健医療科学院は、特定健診保健指導の標準化ソフトウェアを無償提供している⁵⁴⁾。このソフトは特定健診や保健指導の実施に伴う支払い情報や健診データを健診機関から保険者・支払基金などで伝達する際に用いられる標準のXML書式を作ることが主眼であり、身体計測や診察、血圧データといった基本要素だけでなく、がん検診、生体検査、免疫検査といった一般的な検診・健診にはないデータ項目も標準XML書式へ変換することが可能となっている。

また、一般社団法人PHR協会は、報告書「健康管理情報連携の標準化レポート」を発刊しているが、その報告書の中で最も大量かつ全国的に蓄積されている個人健康記録として健診情報を挙げている。本来であればライフコースヘルスデータとして経年的にデータ連結が必要なところであるが、事業者等の保存義務が有限であることなどによりデータが散在、消失している現状を課題ととらえ、永続的保存を前提とした「健診情報の標準化」の検討に取り組んでいる⁵⁵⁾。同会からは、健診結果の詳細を記録するための仕様書や、データ形式変換ツールなどが提供されている。

医学系の学会関係では、2018年、生活習慣病（糖尿病、高血圧、脂質異常症、慢性腎臓病）の診療ガイドラインを策定している4学会（日本糖尿病学会、日本高血圧学会、日本動脈硬化学会、日本腎臓学会）、および検体検査の測定法やデータの標準化に関連する日本臨床検査医学会、医療情報全体の標準化や活用を推進する日本医療情報学会の計6学会によって、「PHR推奨設定」が策定されている⁵⁶⁾。この推奨設定では、生活習慣病の未発症者および発症者のそれぞれについて、生活習慣病自己管理項目セットが策定されているのが特徴となっている（図11）。

⁵⁴⁾ 国立保健医療科学院 研究情報支援センター 水島洋先生提供資料より

⁵⁵⁾ 「健康管理情報連携の標準化レポート 2017-2018 第2.0版」 <http://www.phrj.org/archives/332> (2020/5/19 参照)

⁵⁶⁾ 日本医療情報学会 HP <http://jami.jp/medicalFields/2018Oct23> (2020/5/19 参照)

図11 PHR推奨設定 (抜粋)

| 生活習慣病自己管理センター開設状況 | 項目名 | 単位 | 各種データの自己管理項目セット | | | | 医療従事者 人工知能支援 設定 | リアルタイムデータの連携 | | | | | PHRで連携する設定 | | 医師連携によるアラート | | 医師からの差違によるアラート | | 本人が設定可能な項目(任意) | |
|-------------------|--------|-------|-----------------|------|------|------|-----------------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|-------------------|--------|----------------|-----------|----------------|--|
| | | | 連携項目 | 連携項目 | 連携項目 | 連携項目 | | 企業業務 | 検査リクエスト | 中リクエスト | 高リクエスト | 下リクエスト | 上リクエスト | 下リクエスト | 上リクエスト | 下リクエスト | 上リクエスト | その他の連携ルール | | |
| | | | 連携項目 | 連携項目 | 連携項目 | 連携項目 | | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | 連携項目 | | |
| 1 | 身長 | cm | ○ | ○ | ○ | ○ | 初期のみ | 初期(計測項目(1))へ | 初期(計測項目(1))へ | 初期(計測項目(1))へ | 初期(計測項目(1))へ | - | - | - | - | なし | 10 | 300 | | |
| 2 | 体重 | kg | ○ | ○ | ○ | ○ | 12ヶ月 | 初期(計測項目(1))へ | 初期(計測項目(1))へ | 初期(計測項目(1))へ | 初期(計測項目(1))へ | - | - | - | 20日以内の超過より5%以上の増減 | 1 | 300 | | | |
| 3 | 肺活量(1) | litre | ○ | ○ | ○ | ○ | 12ヶ月 | 100% < 20% | 100% < 20% | 100% < 20% | 100% < 20% | 100% < 20% | 100% < 20% | 100% < 20% | なし | 1 | 100 | | | |
| 4 | 収縮期血圧 | mmHg | ○ | ○ | ○ | ○ | 12ヶ月 | < 130mmHg | 130mmHg < 140% | 140mmHg < 150% | 150mmHg < 160% | 160mmHg < 180% | 180mmHg < 200% | 200mmHg < 240% | なし | 10 | 300 | | | |

出所：日本医療情報学会 PHR推奨設定 (第1版)

以上のように、PHRの標準化は、出生時等の健診や特定健診など、国や自治体、企業等によって広く実施されている検診・健診に対する取り組みが始まっているほか、生活習慣病など一部の医療現場においても、標準化が開始されている。しかしながら今後は、広く普及し始めている健康管理のバイタルデータや行動データを始めとしたPHRの測定データに対しても、標準化の検討が求められている。データの信頼性や互換性の観点、また国際的標準との整合性などの検討を進めて行くべきである。事業や団体ごとにPHRの標準化を進めるのではなく、例えば厚生労働省が主催している「健康・医療・介護情報利活用検討会」などの場を活用し、諸外国の医療ヘルスケアデータの標準化状況も参照しつつ、多様なPHRを連結可能とするような標準化を進める具体的な取り組みの検討が必要ではないだろうか。

また、これらの標準化はデータをつなぐための、フォーマットや項目に関する標準化、すなわちデータを格納する「ハコ」の標準化が主である。日本規格協会の考え方に当てはめれば「互換性の確保」の視点を中心に据えた取り組みと言えるであろう。一方で、いざデータを統合して解析するとなった場合には、そのデータそのものの「中身」の品質も肝心である。データの精度を保てるか、という観点からは、「データの品質の確保のための標準化」の視点も必要となる。

測定手法の標準化の必要性

ビッグデータの利活用が叫ばれるようになって数年が経過し、質の悪いビッグデータはビッグなミステイクを引き起こし、あるいはディープラーニングはディープに間違える、といった笑えない指摘もされるようになってきた⁵⁷⁾。

これは、医療ヘルスケアのデータに関しても同じように当てはまる観点である。例えば、健康診断などで肺活量の測定を経験された方も多だろう。呼吸機能検査の精度は、患者や検査技師に依存して変化することが知られており、EV (Extrapolated Volume: 外挿気量)、PEF-time (Peak-Flow-Time: 最大呼気流量時間) といったいくつかの指標については検査技師の経験年数により差が生じることなどが明らかとなっている⁵⁸⁾。これは測定者間差としてデータマネジメント上考えなければならぬ点だが、他にも機器間差、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、記録の正確性など、意味のある解析を

57 「第3回 デジタルヘルスシンポジウム」、「IoT/AI 時代におけるオープンイノベーション推進協議会 第1回シンポジウム」経済産業省江崎禎英氏講演より

58 「努力性肺活量測定の妥当性指標と患者・検査技師との関連」 医学検査 Vol.63、No.6、2014

行うためにデータの質の観点から考えなければならない点は多数ある。

別の例では、2種類のがんに関し、米国国立がん研究所のコホートデータ（がん画像アーカイブ TCIA ; The Cancer Imaging Archive）のデータと我が国の国立がん研究所のデータをクラスタリングした場合、がん種でなく施設で分類されてしまうケースや、複数施設にまたがるデータを使ってAIが学習した場合、特徴量として「施設」が抽出されてしまう、というケースも報告されている⁵⁹⁾。この他にも、3種類のリストバンド型ウェアラブル機器を同時に腕にはめてバイタルデータを測定したところ、3種類とも異なる数値を示した、との例もある⁴⁷⁾。すなわち、特定施設や特定機器の使用といった限られた条件下というコントロール可能な範囲でデータを収集してフィードバックしていくならまだしも、社会実装して国民から広くデータを収集、解析し、そのデータにより個別介入やアウトカム評価を実施する場合には、こうした「測定の質」を許容できる範囲に収められるよう、可能な範囲で手技や機器等の標準化を進めておくこと、データの質を利用者が正しく理解しておくこと、そしてデータ利用者がデータの質を理解できるよう、データ収集者がデータの質に関する情報を開示したり、あるいは試用できる環境を整えたりしておくことが必要と言えよう。

測定手法標準化の現状

デジタルツールを用いて取得されるPHRの測定手法については、そのテクノロジーの性質上、ウェアラブル機器やセンサー等を提供する企業を中心に確立されている。測定の堅牢性確保が個別になされるとともに、既存のデジタルに依らない測定手法との整合性を検証することで、バイオマーカーとしての有用性が確認されている。しかしながら、例えばバイタルデータを取得するためのウェアラブル機器は、必ずしもFDAの認可を得ているわけではなく、消費者が自分の日常的な健康状態を管理する目的で収集したそれらのバイタルデータが、医師からみて信用に足るか判断できない、といった指摘もされている⁶⁰⁾。

一方、遺伝子解析で得られるゲノム情報やバイオマーカーといった生体内部の情報については、バイオテクノロジーの要素が強く、生体試料の採取や分析法の堅牢性確保といった、検体の取り扱いや測定方法などのウェットな要素が重要となる。これら生体試料の取り扱い法や前処理、分析法の標準化は、すべてのサンプルや取扱い機関において統一することは現実的に困難であるため、まず「どのような条件で実施されたか、記載すること」から標準化が進めようとしている。例えば、バイオバンク試料の付随情報（メタデータ）の標準化手法として、SPREC（試料採取から凍結保管するまでの時間、遠心条件、処理温度等の情報を階層化して共通コードにしたもの）、BRISQ（研究報告の質を高めるための枠組み）、MIABIS（バイオバンク情報を共通化するための最小データ項目）といった国際標準の整備

⁵⁹⁾ HiDEP 特別シンポジウム（2019/9/13 開催）浜本隆二氏、瀬々潤氏講演内容より

⁶⁰⁾ HIMSS19“Doctors don't know what to do with data from wearables”<https://www.medtechdive.com/news/himss19-doctors-dont-know-what-to-do-with-data-from-wearables/548575/>

が進んでおり、我が国においても導入が議論されている⁶¹⁾ (表4)。

表4 バイオバンク試料の付随情報(メタデータ) 標準化スキームの例

| 略称 | 名称 | 概要 |
|--------|--|--|
| SPREC | Standard PREanalytical Code | 試料採取から凍結保管するまでの時間、遠心条件、処理温度等の情報を階層化して共通のコードにしたもの |
| BRISQ | Biospecimen Reporting for Improved Study Quality | 研究報告の質を高めるための枠組み |
| MIABIS | Minimum Information About BIoBank data Sharing | バイオバンク情報を共通化するための最小データ項目 |

日本では、生体試料情報の質の確保は、まだ不十分であるとされる。例えば日本医療研究開発機構(AMED)が実施したアンケートによれば、生体試料利用のボトルネックとして、品質管理情報の不足・開示は第2位に挙げられている⁶²⁾。医薬品等の研究開発を行う企業による利活用を推進していくためには、生体試料の質は重要な観点であり、生体試料のバンキングを行っている場合には、その収集法などの付随情報の充実・開示が望まれる。厚生労働省はバイオバンク・ジャパン(BBJ)、ナショナルセンター・バイオバンクネットワーク(NCBN)等バイオバンクを取り扱う機関の支援も実施しており⁶³⁾、バイオバンク連絡会等でも上述のように質の高い生体試料を得るための取り組みを共同検討している。こうして蓄積される「質の高い生体データを得るためのノウハウ」を、医療機関やアカデミアに水平展開していくことも、バイオ分野の質の高い研究のためには必要な要素である。

メタデータの充実

上項では、特に生体試料の取り扱いについて、前処理も含めた測定手法の標準化の観点から現状を紹介した。一方、生体試料の取り扱いの課題と同様の状況がPHRでも考えられる。しかしながら、PHRはユーザー範囲が格段に広く、目的によってデータユーザーの求める質が著しく異なるため、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、かえってユーザビリティを悪化させることになると予想される。よって、現実的には、サンプリング、前処理、デバイスの種類、測定法、測定環境といった付帯情報、すなわち「データのためのデータ(メタデータ)」を本体データに常に関連付けすること、ならびにメタデータの構造や説明手法(共通言語: Ontology)などの標準化を図ることが、まずは重要となる。すなわち、すべてを標準化するのではなく、データユーザーがメタデータを参照して、使えるクオリティかどうかを自分の目的に応じ判断する、という仕組みである。

⁶¹⁾ 第4回バイオバンク連絡会(2018/6/9開催)

⁶²⁾ 第7回バイオバンク連絡会(2019/4/20開催) 日本医療研究開発機構資料より。なお、第1位は「倫理申請手続きの困難さ」であった。

⁶³⁾ 厚生労働省 第9回厚生科学審議会(臨床研究部会)資料より(2019/1/23開催)

これらを可能にする代表的な取り組みとして、ゲノム科学の領域では、RDF (Resource Description Framework) の活用が進められている。RDF は Semantic Web⁶⁴⁾ の一種であり、元来はインターネット上の情報をコンピュータで処理しやすいものにする国際的な標準形式としてワールド・ワイド・ウェブ・コンソーシアム (W3C) が提案しているものだが、近年はバイオサイエンス分野での活用が進んでいる。日本では、ライフサイエンス統合データベースセンター (DBCLS) とバイオサイエンスデータベースセンター (NBDC) が RDF によるデータ統合を推進しており、2018 年に運用が開始された日本人ゲノム多様性統合データベース (TogoVar) にも取り入れられている⁶⁵⁾。ライフサイエンスの知識と臨床情報を共有していくためには、データの保有するメタデータまでも含めて機械可読な形に整えて解析する必要がある、メタデータの記述様式はクオリティデータの確保において今後重要な要素となるであろう。

3-3. 電子カルテ情報の質の向上に関する取り組み

また、医療ヘルスケアデータの標準化を考える点で忘れてはならない点が、医療情報 (EHR : Electric Health Record) に含まれる電子カルテの標準化である。日本では 1990 年代から取り組みが始まっているが、現状は医療先進国中で最も普及が遅れているといわれている⁶⁶⁾。電子カルテなどに記載されている医療データが標準化されることは、元より医療機関をまたいだ情報を統合整理できる素地が出来ることになり、医療現場での情報共有化や医療連携などの一次利用が広がることに繋がる。また二次利用が進み、医療ヘルスケアの質の向上につながるとして、国はデータの利活用基盤整備を進める中で電子カルテの規格の標準化・一元化に向けた取り組みを進めている⁶⁷⁾。

2018 年 6 月に閣議決定された「未来投資戦略 2018」⁶⁸⁾ では、医療機関等における健康・医療情報の連携・活用を具体的施策として掲げ、全国的にデータ共有を行うための標準規格等を策定する方針が盛り込まれており、個人の健康・診療・投薬情報を医療機関等の間で共有するための工程表を策定することが決定された。また、2019 年 4 月 24 日には、規制改革推進会議の医療・介護ワーキング・グループより、医療データに関する「標準規格」「標準仕様」を早急に検討・公表し、各種のデータを連結するための環境整備に向けた検討を進めるための意見書として、「医療分野におけるデータ利活用促進に関する意見」が発出され

⁶⁴⁾ web ページに記述された内容について、メタデータを一定の規則に従い付加し、コンピュータによる自律的な情報の収集や加工を可能にするシステム。

⁶⁵⁾ 科学技術振興機構 (JST) プレスリリース <https://www.jst.go.jp/pr/announce/20180607/index.html> (2020/5/19 参照)

⁶⁶⁾ 日本経済再生本部、未来投資会議、https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_zentai.pdf

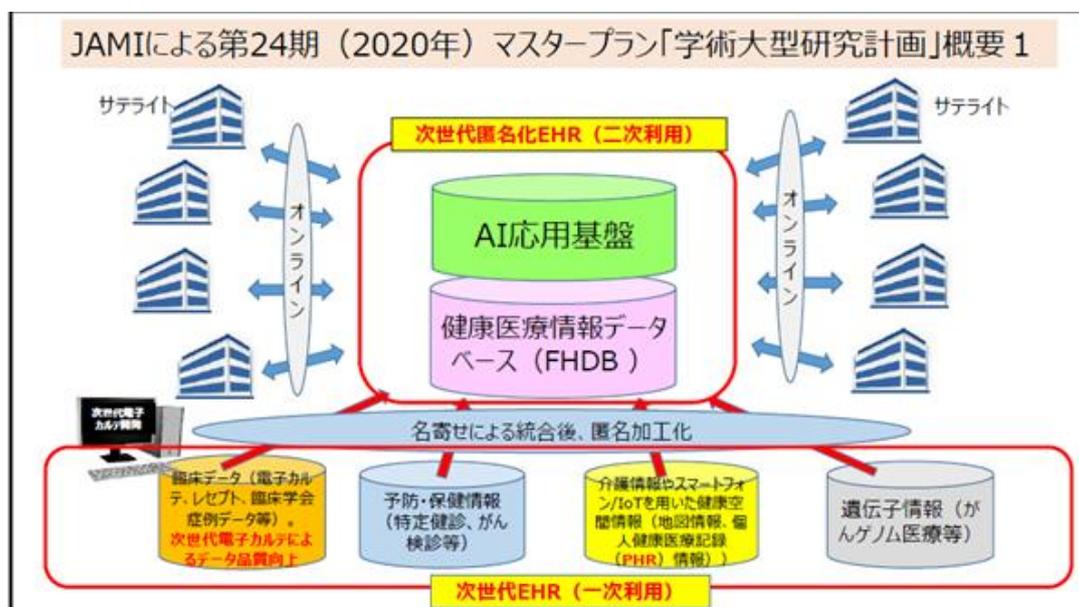
⁶⁷⁾ 厚生労働省、保健医療情報標準化会議、https://www.mhlw.go.jp/stf/shingi/other-isei_127298.html

⁶⁸⁾ 「未来投資戦略 2018」https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_zentai.pdf

た⁶⁹⁾。同様に、社会保障審議会の医療部会や、保健医療分野 AI 開発加速コンソーシアムなどからも「電子カルテの標準仕様」を検討すべき、との強い要望が出ており、厚生労働省は 2019 年度予算において「医療情報化支援基金」(国費ベースで 300 億円の計上) を創設し、電子カルテ標準化に向けた医療機関の電子カルテシステム等導入費を補助することとした⁷⁰⁾。

実際の活動として、医療の質の向上に貢献する医療情報化を推進する団体である日本医療情報学会 (JAMI : Japan Association for Medical Informatics) は、マスタープランにおいて次世代電子カルテによる臨床データ (電子カルテ、レセプト、臨床学会症例データ等) の品質向上が掲げられ (図 12)、この学会理事のメンバーが中心となって 2019 年 7 月には次世代医療電子カルテシステムを検討する、「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)」が立ち上がった⁷¹⁾。

図12 JAMIによるマスタープラン概要



出所：医療RWD共通プラットフォームに向けて – JAMIの方向性 –

NeXEHRs は、「新しい技術を柔軟に活用できる新たな健康医療記録のありかたを検討し、これまでの標準化基盤をベースにして、これらの技術にも対応していく新しい電子カルテシステムの共通プラットフォームを設計する」、「AI/IoT 時代の次世代標準健康医療記録システムの基本コンセプト、共通プラットフォームのあり方、医療制度と法制度の課題も含めて議論する」ことを目的として設立されている。新しい電子カルテシステムの共通プラットフォームの設計にあたり、図 13 に示す通り 3 つの基本理念、5 つの実現方針をコンセプト

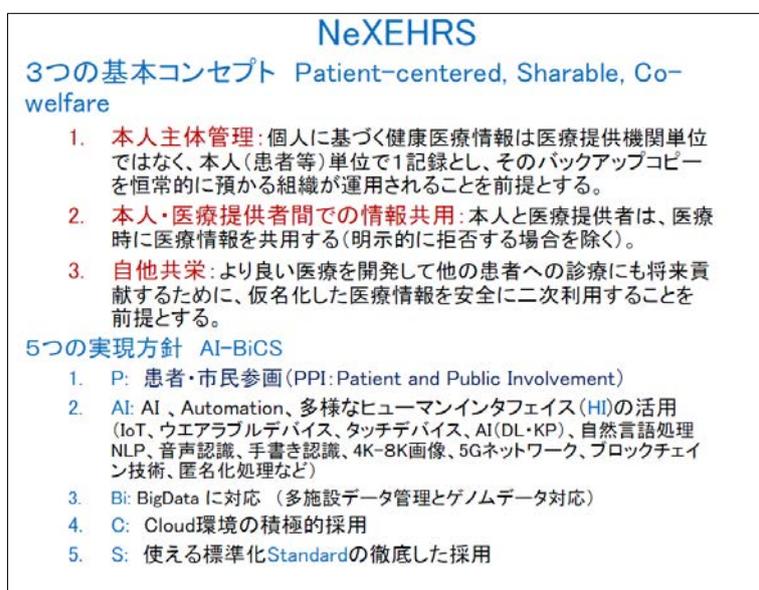
⁶⁹⁾ 医療分野におけるデータ利活用促進に関する意見(2019/5/10) <https://www8.cao.go.jp/kisei-kaikaku/suishin/meeting/committee/20190510/190510honkaigi02.pdf>

⁷⁰⁾ 厚生労働省、「保健医療情報を全国の医療機関等で確認できる仕組み及び標準的な医療情報システムについて」
<https://www.mhlw.go.jp/content/10801000/000555960.pdf>

⁷¹⁾ <https://nexehrs.jp/> (2020/5/19 参照)

として掲げている。

図13 次世代健康医療記録システム



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

このようなコンセプトが実現すれば、健康医療情報が医療機関単位ではなく患者単位で永続的に記録管理され、さらに患者自身も医療者もその記録情報を共有しつつ効果的に利用できることとなる。また、患者状態や健康履歴、診療履歴の記録と医療行為の記録とが相互に関連づけられ、ひいては医薬品の研究開発などの二次利用にも効率的に対応できるようになる。医療ヘルスケアデータの重要な一部である医療情報の標準化や相互運用性の向上は、データ駆動型ヘルスケアの実現に向け、大きな推進力となるだろう。

3-4. クオリティデータに関する期待と課題

本章では、データ駆動型ヘルスケアを実現するためには、目的に沿った質を有するデータが必要であることを大前提として、医療情報やPHRに求められる様々な視点を取り上げた。こうしたデータが広く産業で活用されるには、「相互運用性」の観点から標準化や構造化の推進は必須であろうし、機械可読性を高めて入力ミスやデータの揺らぎを抑える、主観的評価を排除して客観性を高めるといった視点からは、ウェアラブル機器やセンサー等のIoTからのデータを医療ヘルスケアデータの中心に据えることも検討されるべきであろう。

このような課題を検討しつつ、二次利用者が対象データの質を評価できるよう、例えば測定法に関するメタデータの充実、データ説明インデックスの標準化なども並行して進めることが必要であろうし、アノテーションの質の確保や標準化や、さらには公平性、説明可能

性、透明性に配慮したデータ収集といった、データ活用の実装や AI 利活用の観点からのきめ細やかな取り組みも大切になってくる。

これらに加え、今後日本でデータ駆動型ヘルスケアのシステムや関連産業を育成し国際展開を進める、あるいは諸外国で確立されたシステムを素早く国内展開を行うためには、これらの規格や法制度をガラパゴス化させず、国際標準に沿った形で普及させることが重要となる。第 2 章で取り上げたセキュリティやプライバシーと同様、国際的にも信頼される仕組みを構築していくことが、国の基盤整備を行う上で必要であることは間違いない。

4. まとめ

当研究会では、医薬品産業としての具体的なビッグデータの活用と成果取得や実用化に向けた課題抽出を行うため、特にここ2年間においては、「医療データの活用」「PHRの活用」といったテーマの検討を行ってきた。我が国におけるビッグデータの活用を概括的にみると、短期的には医療機関にある医療データの活用を進めることが重要であり、レジストリの構築・活用、テンプレートの普及、ゲノム等生体分子情報との連結、PPPsの構築やトライアル環境の整備といった課題解決が必要であること、また、中長期的にはPHRの活用を進めることが重視され、データポータビリティの推進、多様なデータを連結・統合したデータベースの構築、産学官民協働による評価指標・標準等の確立、デジタルソリューションの活用・普及、データ提供者等への成果フィードバックが必要であることを、2017、2018年度の各報告書にて具体的な提言、提案としてまとめてきた。

5年目を迎える本年度（2019年度）は、データの利活用には、データ提供者、データ管理者、データ利用者の3者すべてにメリットが生じるような「データエコシステム」の形成が必要であるとの認識のもと、特にデータの価値の観点からは「クオリティデータ」を、データ/価値の循環の観点からは「アクセシビリティ」を重点テーマとし、有識者の意見を仰ぎつつ、課題抽出を行った。本項では、これらの観点を下記の3点の重要方策にまとめた。

1) 医療ヘルスケアデータのアクセシビリティの向上

2018年5月に施行された次世代医療基盤法は、匿名化された医療情報の医学研究への活用が主目的であり、個人識別符号に該当するゲノム情報との連結ができないことや、個人に戻っての追加のデータ取得が困難といったことなど、データ利用者のアクセシビリティ上の課題が指摘されている。他方、2020年の個人情報保護法改正では、個人情報を適切に活用していくため、個人情報の新たな類型として仮名加工情報が創設されることとなる。この仮名加工情報を活用することによって、匿名加工情報の活用を企図した次世代医療基盤法の課題の多くを解決できるのではとの期待がある。

そこで、個人情報保護法を補完する特別法である次世代医療基盤法の普及を推進するとともに、医療健康分野の仮名加工情報の取り扱いを定め、研究活用を推進する「医療情報特別法（仮称）」の創設を検討してはどうか。

健康医療に関する情報は、機微性の高い情報が多く含まれるため、特に健康医療に関する仮名加工情報は利活用と本人情報保護の両立を図ることが肝要である。当該情報の取り扱い者は、医療健康分野の十分な知識と情報取扱い実績が求められると考えられ、個人情報保護法における仮名加工情報取扱者とは別に、この「特別法」の中で仮名加工情報取扱事業者の認定資格を設け、研究活用を進めることが望ましいと考える。その際、安全管理措置や利用目的の限定、明示、通知といった事業者要件や罰則を設ける必要があるだろう。さらに、

仮名加工情報が正しく取り扱われているかといった「取扱者のガバナンス」に焦点を当て、それを適切な監査機関が適時監査していくことも考慮されるべきである。信頼あるデータ利活用の実現を目指すことが、仮名加工情報活用のバックボーンになるといえるだろう。

一方、医療ヘルスケアに関する仮名加工情報を二次利用する場合は、昨今の個人情報保護法改正や APPA の議論にもあるように、二次利用の目的が公益性の高い研究や事業であることも大切なポイントである。そのため、例えば個人情報保護法における公益性の高い研究・事業を定義し、具体的に Q&A やガイドラインとして明記することや、APPA のフローチャートにおけるホワイトリストの作成や公益性の評価基準を充実させることが必要であり、個人情報の更なる活用につながると言えよう。

また、二次利用者の視点のみでなく、データを生み出す個人にとっても、医療健康データのアクセシビリティの向上は必要である。これまで医療機関に存在していた個人の医療データも対象に含め、データポータビリティや、情報銀行等のデータ流通プラットフォームに関する制度の整備を推進していくことなどにより、個人が自分のデータにスムーズにアクセスでき、また適切に活用をコントロールできる環境の整備が求められる。医療健康分野の仮名加工情報の概念が検討される場合には、二次利用の仕組みとデータ提供者による情報活用の両面から検討されるべきであり、情報銀行での活用や個人によるデータポータビリティの制度化といった点をあわせて議論することも重要となろう。

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

2) 医療ヘルスケア分野のクオリティデータ収集の推進

質の高いデータ駆動型ヘルスケアを実現するためには、その元となるデータの質を高めなければならない。一方で、いたずらにデータの質ばかりを追求しては、コストの上昇、クレンジングやクリーニング労力の増大、データ収集の手間や時間の浪費といったデメリットが生じ、データが広く活用される社会を早期に実現することにはつながらない。

ユーザーの目的に応じた質のデータが収集される「fit for purpose」がデータ収集の大前提にあるなかで、効果的な二次利用を実現するためには、まずは二次利用者がどのような目的（purpose）でデータを取得するのか、あるいはどのようなクオリティのデータが必要なのかを明確化することが重要であり、データやデータベースの品質に関しては、可能な場合はユーザーがプロアクティブに関与していくことが望まれる。一般診療や PHR の取得にお

いて、前向きなデータ収集を企図するのは困難ではあるが、疾患研究やレジストリの構想を持つデータ収集では、製薬企業等の二次利用者も交え、医産学官が連携してデータベースの品質をデザインすることが求められる。特に今後期待の大きい疾患レジストリについては、**Good Registry Practice** 等の基準を整備し、外部評価を実施しながら質の向上を図ることが、医学研究を社会に還元していくためには重要である。

一方、医療ヘルスケアデータのコアとなる電子カルテ等の医療情報についても、医療現場での活用だけでなく、患者のライフコースヘルスデータとしての活用や二次利用に向けた取り組みが進展しつつある。電子カルテの更なる普及にとどまらず、構造化、標準化がなされ、異なる医療機関で取得されたデータを経時的に連結できるなど、個人を 1 単位として記録されるような情報パッケージであること、更には他の情報と統合解析するための相互運用性が確保されていること（デジタルライゼーション）等の基本設計が、今後の医療情報については重要だろう。これを実現するためには、現在学会やベンダーを中心に進められている次世代電子カルテの取り組みに政府が連携し、インセンティブ付与などにより普及において力を合わせる必要がある。さらに施設を超えた医療情報共有を ID 等により可能とし、**patient centricity** を推進すること、そして厚生労働省の推進する PHR やゲノム情報等の多様な個人情報との連結、相互運用性のあり方について早期の実現を図っていくことが、データ駆動型ヘルスケアを進める上で不可欠だろう。

さらに今後は、レジストリや電子カルテ等の医療情報だけでなく、患者・生活者から連続的に取得されていくであろう IoT からの膨大な健康・生活情報や、モバイルからのデータ、画像・映像データ等の資源の比率が急速に高まっていくだろう。言わば次世代の医療においては、「社会全体がデータソース」であり、膨大なデータを AI で活用すること、データをクラウドで必要に応じて連結できることを前提に、広義の **Personal Health Record (PHR)** 収集の仕組みをデザインし、そのデザインに沿って各データベースを充実させていくことが重要となろう。

そして、こうした社会の実現に向け、予め二次利用の目的を織り込んでおくということは、すなわち二次利用が一次利用の性質を帯びてくるということに他ならない。もとより価値を還元される個人や社会にとっては、一次利用者も二次利用者も価値提供者であり、むしろ一次利用者と二次利用者が一体となって価値を提供する体制を構築していくことが望まれる。

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

3) 信頼に基づくデータ循環と社会（国民）への成果・価値の還元

医療ヘルスケアデータの提供者は、国民である。したがって、国民がデータ提供のメリットを素早くかつ明確に得られること、そしてデータの利用者やシステムが信頼されることが、継続的で円滑なデータの利活用には不可欠である。

そのためには、まずはベースとして自分のデータがいつでも容易に、自分の望む形で閲覧できることが大切であり、またこれまで述べてきたような標準化、相互運用性の確保やデータポータビリティの推進が必要である。さらに、スマートフォンなどを活用したユーザーフレンドリーなインターフェースの提供や、データのクラウドへの保管といったテクノロジーの活用も推進しつつ、サイバー/フィジカルの両面からデータへのアクセシビリティを高めることが重要である。

これに加え、利用者によって使われたデータが素早く価値（成果）として結実し、かつタイムリーに国民・社会に還元される仕組みをつくることが、データ駆動型ヘルスケアの社会認知を高めるためには必要であると考えている。例えば医薬品産業における創薬研究は民間企業の研究であるが、同時に、新薬を待っている患者や社会、医療への貢献が期待されるという点で、公益的な目的を有している。データ活用の自由度を高めることで、このような恩恵がより早く国民に届くということを、社会に伝え、理解してもらう必要がある。医療にとっても、AI 診療支援の高精度化や、薬剤師・介護者等も含めた患者中心のチーム医療や他施設連携におけるデータ利活用の推進、データを基にした LHS の実装などがデータ活用によって進み、データを提供することが自分たちのメリットになるということを、国民に実感してもらうことが必要である。

もちろんこれらの取り組みに加え、個人のデータがセキュリティやプライバシーに配慮されたかたちで利用されていることを実感できることも大切であり、どのデータに誰がいつどのようにアクセスしたか、またデータがどのようなアウトプットに使われたか、それが倫理面においても問題なかったか、といった「データの使われ方」を確認できるよう、透明性や説明可能性を高めることも大切である。特に今後は、データ利用者の信頼性をどう評価するか、といった視点も重要になる。データの取り扱いに対する国としての考え方や方向性を明確にし、国民のコンセンサスを得ることも求められる。諸外国の規制動向等も鑑みつつ、

国際的にもハーモナイズした、実効性と納得感の高いデータガバナンスの明確化（ガイドライン等の策定）を進める必要があるだろう。

そして、こうした概念を素早く確実に社会実装するためには、国民に近い場所で取り組みを進めることが有効であろう。昨今注目されている「リビング・ラボ」⁷²や東北メディカルメガバンクに代表されるコホート研究のように、国民の協力を得ながら、国民とともに健康や生活と疾患との関係を探る研究や、その成果の社会実装を進めることが、データ提供のメリットについて認識を高める方策の一つである。こうした生活者と一体となった研究を、例えば自治体を中心としたソーシャル・インパクト・ボンド（SIB）の仕組みでスピーディーに進めることも、今後日常生活からクオリティデータを収集していくためのモデルの1つとなるのではなかろうか。こうした取り組みにおいて、標準化や相互運用性、セキュリティ/プライバシーの国際基準との整合といった観点を盛り込むことで、二次利用可能なデータを収集していくためのデザインの検討の進展も併せて期待される。

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

ここまで述べた項目は、取り組みが既に始まっているもの、既に議論がされているものもあるが、データの活用とその先にあるデータエコシステムを構築していくために、いずれも必要な重要方策であると考えられる。その中でもわが国のデータ駆動型ヘルスケアに対する検討や取組み進展状況を考慮した場合、特に下記の4点が喫緊の対応として精力的に進めて行く必要のある項目として挙げられる。

- ① 仮名加工情報を利用する「医療情報特別法(仮称)」の創設
- ② クオリティデータの集積を目指した「レジストリの高品質化」
- ③ 「次世代電子カルテシステム」の国を挙げての実装
- ④ 「データガバナンス」の明示と国民への啓発的取り組み

こうした取り組みにより、データ駆動型ヘルスケアの実現がさらに進展していくことを期待したい。

⁷² リビング・ラボとは、市民・社会を中心に据えて、ものづくり・サービス・政策等を創り出す新しいイノベーション創出の考え方

