

PHR の標準化に向けて

—クオリティデータ収集の視点から—

医薬産業政策研究所 統括研究員 森田 正実
医薬産業政策研究所 主任研究員 佐々木隆之
医薬産業政策研究所 主任研究員 中塚 靖彦

【はじめに】

IoTをはじめとしたデジタル機器の進化や通信環境の発展に伴い、健康に関する個人のデータ（PHR¹⁾）の蓄積環境が構築されつつある。今後の医療のパラダイムシフト、すなわち「集団」から「個人」へ、「治療」から「予防」へ、「モノ」から「コト」へといった変化を見据えた場合、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る。このような「データ駆動型ヘルスケア」において、データはいわば「第二の石油」としてその主役を担うわけだが、多種多様なステークホルダーが利用するというその性質上、二次利用を前提としたデータの「質」や「セキュリティ」、そしてそのデータを利用できるという「アクセシビリティ」が重要となる。

政策研ニュース No.57で述べたとおり、AIの活

用を想定したデータ駆動型ヘルスケアにおけるデータの「質」は、これまでのデータサイエンスの観点からの質（「閉じたデータ」の観点）だけでなく、データ利用者のニーズに合致した「適正品質」、産業や国家をまたいだ利活用の前提となる「相互運用性（Interoperability）の向上」、AIの特性を踏まえた「Fit for AI」といった「開かれたデータ」の観点が必要である。特に相互運用性の向上にあっては、例えば医療情報ではHL7 FHIR²⁾といった標準規格の導入が世界的に進みつつあるように、標準化は相互運用性向上の大きなカギとなっている。

本稿では、今後利活用が広がるであろうPHRについても、そ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、考察する。

※) 医薬産業政策研究所ではビッグデータの医薬産業に関する課題を研究するために、所内に『医療健康分野のビッグデータ活用・研究会』を2015年7月に発足させた。今回の報告は、国立保険医療科学院 研究情報支援研究センター 水島洋先生の講演など、『研究会』の調査研究を踏まえてまとめたものである。

1) Personal Health Record

2) Fast Healthcare Interoperability Resources

【PHRの標準化】

標準化とは

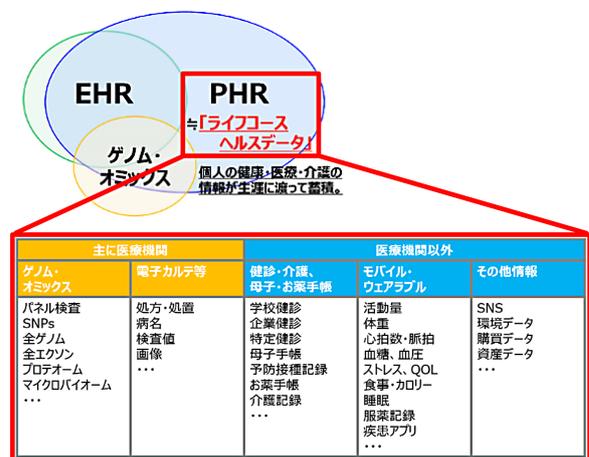
日本規格協会によれば、標準化とは「自由に放置すれば、多様化、複雑化、無秩序化してしまうような『もの』や『事柄』を少数化、単純化、秩序化すること」とされている³⁾。標準化の役割(メリット)としては、互換性の確保や品質の確保、生産効率の向上、相互理解の促進や技術の普及といったものが挙げられている。ビジネスのグローバル化や異業種連携が進む現代では、標準化の役割はますます高まっている。

PHRの範囲

PHRの範囲については、おおまかには「個人の健康医療情報」を指し示すものとしてコンセンサスが得られているものの、明確な定義は定まっていないのが現状である。一例として、厚生労働省は「国民の健康づくりに向けたPHRの推進に関する検討会」において「個人の健康診断結果や服薬履歴等の健康情報を電子記録として本人や家族が正確に把握するための仕組み」と定義する案を提示している⁴⁾。また、これに関連するものとして、未来投資戦略2018では、PHRの構築として、予防接種履歴に加え、特定健診や乳幼児健診等のデータの提供を開始することが目指されている⁵⁾。

しかしながら、効果的な健康増進や行動変容の促進、更にはヘルスケア関連産業全体の振興に向けたデータ流通の観点も見据えた場合、健診情報や服薬情報だけでなく、医療機関で取得されたデータや、ゲノムやオミックス等のより「深い」生体データ、更にはモバイルアプリやウェアラブルデバイス等のデジタルツールにより医療機関以外で取得されるセンサーデータや生活データも含め、生涯にわたって蓄積される個人の「ライフコースデータ」としてPHRをとらえていく必要があるだろう(図1)。

図1 本稿におけるPHRの範囲



3) 「標準化とは」 https://webdesk.jsa.or.jp/common/W10K0500/index/dev/glossary_1/ (2019.9.18閲覧)

4) 「第1回国民の健康づくりに向けたPHRの推進に関する検討会」 https://www.mhlw.go.jp/stf/newpage_06700.html (2019.9.18閲覧)

5) 「未来投資戦略2018」 http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_zentai.pdf (2019.9.18閲覧)

PHRの標準化の現状

PHRをゲノムや行動データまで含めたデータとして扱う場合、その範囲は非常に広範となり、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。

一方で、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基盤が整っている。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2)。そこで現在は、データ収集基盤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できる、としている⁶⁾。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レ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が、項目は健診機関により独自設定されており、対応表の作成が個別に必要なため、対応表設定費用のみ有償(税抜5万円)となっている(2019年9月現在)。

またこれとは別に、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⁷⁾。この標準化ソフトを活用した場合、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、免疫検査といった検診・健診にはないデータをも標準化することが可能となっている(図3)。また、本フォーマットは医療データの国際標準に則ったXML形式で出力され、保険者間連携や診療での活用などが可能になるとされている。

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」において、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おり、事業者等の保存義務が有限であることや永続的保存が必要であることも踏まえ、健診情報の標準化に取り組んでいる⁸⁾。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。

図2 健診機会とその課題

一生の間の健診の機会

- ・ 胎児期の検診 (母子手帳)
- ・ 出生時検診 (母子手帳)
- ・ 幼児期健診
- ・ 学校健診
 - 幼稚園・保育園
 - 小学校
 - 中学校
 - 高等学校
 - 大学・短大・大学院・専門学校
- ・ 職場健診
- ・ 特定健診
- ・ がん検診
- ・ 地域健診
- ・ 人間ドック
- ・ 老人健診

健康状態での数値を把握することが重要

・ 検診結果はそれぞれで不連続に保管(本人でも収集困難)
・ 同一人物のデータが分散
・ 疾患時に活用されない

図3 特定健診に含まれる情報

特定健診結果ファイルデータに含まれる情報

- ・ 身体計測
- ・ 診察
- ・ 血圧等
- ・ 生化学検査
- ・ 血糖検査
- ・ 尿検査
- ・ 血液像検査
- ・ がん検診・生体検査等
 - 肺機能、視力、聴力、眼底
- ・ その他
 - CRP, ABO, 梅毒, HBs, HCV各種, その他の法定検査
- ・ 医師の判断
 - メタボ判定、保健指導レベル、診断、
- ・ 質問票
 - 服薬、既往症、貧血、喫煙、食習慣、飲酒(量)、睡眠、改善

特定健診項目のみならず、がん検診、肝炎ウイルス健診、事業主健診の一部にも対応

6) 「健診標準フォーマットについて」 <https://www.kenshin-hyojun.jp/> (2019.9.18閲覧)

7) 国立保険医療科学院 研究情報支援研究センター 水島洋先生提供資料より

8) 「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> (2019.9.18閲覧)

図4 PHR推奨設定(2018年10月版:抜粋)

| 図表2-1 基本シート | | 「PHR推奨設定」第1版 (基本シート) | | | | 「PHR推奨設定」第1版 2018年10月公開 Copyright © 2018, © 臨床学会社大会編, All Rights Reserved. | | | | | | | | | | | |
|------------------------|----------------|----------------------|---------------|-----|------|---|------------|----------------|-----------------------|-----------------|------------------------|----------------------------|------------------------|-------------------------|----|------|---------------------|
| 生活習慣病自己管理項目セット 基本ID | 項目名 | 単位 | 各病態の自己管理項目セット | | | | PHRで管理する項目 | | | | | | | | | | |
| | | | 入力タイミング設定 | | | | リスク階層毎の閾値 | | | | | | | | | | |
| | | | 糖尿病 | 高血圧 | 脂質異常 | 慢性腎臓病 | 正常範囲 | 軽度リスク | 中リスク | 高度リスク | 設定閾値によるアラート 下閾値 上閾値 | 前段階からの変更によるアラート 下閾値 上閾値 | 本人が訂正権 (あり/なし/数値など) | | | | |
| 1 | 身長 | cm | ○ | ○ | ○ | ○ | 初回のみ | BMI(計算項目(1))へ | BMI(計算項目(1))へ | BMI(計算項目(1))へ | BMI(計算項目(1))へ | - | - | なし | 10 | 300 | |
| 2 | 体重 | kg | ○ | ○ | ○ | ○ | 12ヶ月 | BMI(計算項目(1))へ | BMI(計算項目(1))へ | BMI(計算項目(1))へ | BMI(計算項目(1))へ | - | - | 10ヶ月以内の前値より3kg以上の増減 | 1 | 300 | |
| 3 | 空腹時血糖 | mg/dL | ○ | ○ | ○ | ○ | 12ヶ月 | 18.5以下 <25 | 25以下 <35 35以下 <100 | 35以下 <35 | 35以下 | - | - | なし | 1 | 100 | |
| 4 | 空腹時血圧 | mmHg | ○ | ○ | ○ | ○ | 12ヶ月 | <130mmHg | 130mmHg以下 <140 | 140mmHg以下 <160 | 160mmHg以下 | 90 | 160 | なし | 10 | 300 | 収縮期血圧は収縮期血圧本人が訂正権あり |
| 5 | LDLコレステロール(*2) | mg/dL | ○ | ○ | ○ | ○ | 12ヶ月 | <100mg/dL | 100mg/dL以下 <140 | 140mg/dL以下 <180 | 180mg/dL以下 | - | 160 | 10ヶ月以内の前値から50mg/dL以上の上昇 | 0 | 1000 | |
| 6 | HDLコレステロール(*2) | mg/dL | ○ | ○ | ○ | ○ | 12ヶ月 | ≥40mg/dL | 設定なし | 40mg/dL以下 ≥30 | 30mg/dL以下 | - | - | なし | 0 | 300 | |
| 7 | 総脂質 | mg/dL | ○ | ○ | ○ | ○ | 12ヶ月 | なし | 設定あり | あり | 設定なし | - | - | なし | - | - | |
| 8 | 血清クレアチニン | mg/dL | ○ | ○ | ○ | ○ | 12ヶ月 | eGFR(計算項目(2))へ | eGFR(計算項目(2))へ | eGFR(計算項目(2))へ | eGFR(計算項目(2))へ | - | - | 10ヶ月以内の前値から50mg/dL以上の上昇 | 0 | 30 | |

医療の領域では、昨年、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」が策定された⁹⁾。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4)。

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・健診や、生活習慣病など一部の医療現場において、その取り組みが始まっている。一方で、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」の標準化が主であり、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「中身」も肝心である。よって、その精度を保てるか、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¹⁰⁾。

これは、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¹¹⁾。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、記録の正確性など、意味のある解析を行うためにデータの質の観点から考えなければならぬ点は多数ある。

9) 日本医療学会 HP <http://jami.jp/medicalFields/2018Oct23> (2019.9.18閲覧)
 10) 「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
 11) 「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

別の例では、2種類のがんに関し、米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¹²⁾。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⁷⁾。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「測定の本質」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、そしてその許容できる質をデータ利用者が示し、データ収集者と話し合っ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の多くはFDAによってレビューされておらず、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、といったケースも指摘されている¹³⁾。

一方、比較的「深い」情報としてのバイオマーカーは生体内部の情報であり、どちらかというバイオテクノロジーの要素が強い。それゆえ、生体試料の採取や分析法の堅牢性確保といった、ウェットな要素が重要となる。生体試料取扱い方や、前処理も含めた分析法の標準化は、まず「どのような条件で実施されたか、記載すること」から進められている。国際的な取り組みとしては、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が進んでおり、我が国においても導入が議論されている¹⁴⁾。

表1 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 Blobank 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 |

12) 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

13) 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/>

14) 第4回バイオバンク連絡会（2018.6.9開催）

翻って我が国では、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¹⁵⁾。厚生労働省やAMEDの推進しているクリニカル・イノベーション・ネットワーク推進事業においても、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疾患レジストリ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）、東北メディカル・メガバンクの支援も実施しており¹⁶⁾、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。ここで蓄積された「質の高い生体データを得るための手法」を、例えばレジストリ管理基準の制定などを通じて個別の疾患レジストリに展開していくことも、質の高い研究のためには今後必要な要素となろう。

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。一方で、PHR全般を考慮すると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、まずは重要となろう。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。

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はセマンティックウェブ¹⁷⁾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。我が国では、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¹⁸⁾。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り、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のではないかと。

-
- 15) 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。
- 16) 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）
- 17) 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。
- 18) 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>（2019.9.18閲覧）

【まとめ】

医療のパラダイムシフトが進展していくなか、健康に関する個人のデータとしての PHR についても、その質が重要となってくる。

PHR をクオリティの視点から見た場合、フォーマットや構造、用語といったいわば「ハコ：情報学的観点に基づく互換性の確保」の観点と、そのハコに詰め込む「中身：データそのもの」の観点が存在し、コストや時間、煩雑さといった実装面との折り合いをつけていくためには、まずはデータ収集者（管理者）の収集の利便化と、データユーザーがデータの使用を判断するに足る「データに付随するデータ（メタデータ）」の充実の視点が大切になる。

こうしたデータ収集基盤の構築において、クオ

リティデータの考え方を浸透させて異なる規格やフォーマット等が乱立するのを防ぐには、相互運用性を有するプラットフォームの構築が必要であり、特に医療や検診・健診と異なり既存プラットフォームが少なく、かつ他産業での幅広い利活用も想定される PHR については、さらにその点を意識する必要があるだろう。

加えて、ヘルスケア産業の発展可能性が全世界的に広がる今後にあっては、国際展開や国際連携を視野に入れて国際標準とのハーモナイズも進めつつ、サイバーセキュリティやアーキテクチャ（制度・法規制等の設計）への対応も同時に必要となってくる¹⁹⁾。我が国においてもこのような議論が進展することを期待したい。

19) 政策研ニュース No.58 「欧米のヘルスデータプラットフォームにおける「セキュリティ・プライバシー」規制の概況について」