

科学技術未来戦略ワークショップ報告書

**ライフサイエンス・臨床医学分野における
データベースの統合的活用戦略**



独立行政法人科学技術振興機構 研究開発戦略センター
Center for Research and Development Strategy, Japan Science and Technology Agency

エグゼクティブサマリー

独立行政法人科学技術振興機構（JST）研究戦略開発センター（CRDS）は、社会ニーズを探索し、ニーズを充足し、社会ビジョンを実現するための科学技術の有効な発展に貢献することを目的として、科学技術の全分野を対象にして、政府への戦略提言を行っている。平成24年度、提言作成対象となる重要領域として、「ライフサイエンス・臨床医学データベース」を選定し、平成24年4月に調査・提言活動を開始した。

データベースの統合的活用とそこで展開されるデータサイエンスは、ライフサイエンス・臨床医学分野でのイノベーション創出の最も重要な基盤の1つである。米国では、ゲノムデータベースと電子カルテシステムを組み合わせた遺伝疫学研究を進め、個別化医療の実現に向けて、データベースの統合的活用研究が進められている。しかしながら、我が国のライフサイエンス・臨床医学領域におけるデータベース活用は、世界と比べると動きが遅く、各領域で別々に作成されたデータベースがほとんど連携せずに動いている。JST-CRDSでは、「データベースの統合的活用」にフォーカスした調査と提言の検討を進め、その一環として、平成24年10月31日に、関連分野の第一線で活躍されている有識者を招聘したワークショップ「ライフサイエンス・臨床医学分野におけるデータベースの統合的活用戦略」を実施したので、その内容を本書にて報告する。

ワークショップでは、複数のデータベースの統合的活用から、データサイエンス活性化、イノベーション実現までを対象とした議論を進め、主に以下の課題と対策について共通認識が得られた。

- データベースを統合的に活用することで、国民、医療、行政などと幅広い対象にインパクトを与えることができる
- 個人のプライバシーを保護しつつ、研究開発成果を国民生活に還元するという理念が必要
- 統合すべきデータベースは、臨床系、基礎科学系、ゲノム情報にまたがる
- 医療関連の個人情報利用には、個人情報保護法の個別法が必要
- セキュアな情報ネットワーク（規格化、標準化、共有化を含む）が必要
- 機械学習などの最新情報処理技術の導入が必要
- 医師と情報処理研究者とデータマネージャーの協働が必須
- 我が国のデータベース戦略を国家戦略として立案し、国際競争に備えることが急務

JST-CRDSでは、本ワークショップでの議論を受け、データベース活用を促進するための研究開発提言作成を行う予定である（平成24年度末に発刊し、JST-CRDSホームページにも掲載予定）。

目 次

エグゼクティブサマリー

1. ワークショップ趣旨説明	1
2. 発表・討論概要（セッション1、セッション2）.....	2
セッション1：「データベースの活用目的」.....	2
セッション2：「統合的データベース活用に向けた技術課題と制度課題」.....	30
3. 総合討論（セッション3）.....	57
オリエンテーション.....	57
総合討論.....	61
4. まとめ.....	71
参考資料	73
①ワークショッププログラム	
②ワークショップ参加者リスト	

1. ワークショップ趣旨説明

独立行政法人科学技術振興機構（JST）研究開発センター（CRDS）は、「社会ニーズを充足し、社会ビジョンを実現する科学技術の有効な発展に貢献する」ことを目的として、科学技術の全分野を対象にして、政府への戦略提言を行っている。平成24年度、提言作成対象となる重要領域として、「ライフサイエンス・臨床医学データベース」を選定し、平成24年4月に調査・提言活動を実施するチームを発足させた（統括責任者：浅島誠上席フェロー、チームリーダー：森英郎フェロー、他5名）。

ライフサイエンスや臨床医学は多様性を扱い、ゲノム、代謝、病態、種と言った膨大なデータの解析から真理を追究する科学を基本としている。日々生み出される膨大なデータは、データベース化され、次世代の知識を生み出す源泉として活用が進んでいる。例えば、2012年のノーベル医学生理学賞を受賞した山中教授の研究では、我が国のマウス遺伝子公共データベースからの候補遺伝子抽出に始まり、最終的には細胞の初期化に必要な本質的な4遺伝子の特定に成功している。このように、個々のデータベースの単独利用から様々な研究成果が生み出されている。しかしながら、今後は複数のデータベースを統合的に解析し、より複雑な事象を解明する必要性が高まるだろう。例えば、米国ではゲノムデータベースと電子カルテシステムを組み合わせた遺伝疫学研究を進め、個別医療の実現に向けて、個人ゲノム情報を診断に取り込む研究を進めている。米国での例を見るまでもなく、データベースの統合的活用とそこで展開されるデータサイエンスは、ライフサイエンス・臨床医学分野でのイノベーション創出の基盤となる。

我が国では、高速情報通信インフラ、セキュリティ管理技術、マイナンバーによる個人認証、医療データ取扱いのポリシーガイドが整備されつつある。しかしながら、世界と比べるとその動きは遅い。当チームで関連分野の情報収集を進めると、様々な阻害要因が見出された。法制度、基盤整備、人材確保など様々な局面で問題があることが分かった。ライフサイエンス・臨床医学分野でのデータベース活用から、データサイエンス活性化、イノベーション実現までを推進する方策を検討するため、当チームにて問題点の分析を進めた結果、関連法案整備も含んだ制度・体制面での提案も含有する複合的な研究開発推進戦略の立案が必要との結論に至った。そこで、関連分野の第一線で活躍されている有識者を招聘し、「ライフサイエンス・臨床医学分野におけるデータベースの統合的活用戦略」を議論する本ワークショップを開催した。

ワークショップは3部構成とした。まず、セッション1でデータベース活用目的の明確化を行い、イノベーションにつながる領域の絞り込みを進めた。次に、セッション2で技術課題と制度課題の洗い出しを進め、優先課題の明確化を行った。そして、セッション3の総合討論において、法制度整備を含めた複合的視野から、データベースの統合的活用に向けた研究開発戦略骨子を議論した。ワークショップの討議内容をセッションごとにまとめ、本書にて報告する。

2. 発表・討論概要（セッション1、セッション2）

セッション1：「データベース活用目的」

医療・医学の視点から

永井 良三 先生（自治医科大学学長、CRDS 特任フェロー）

- ▶ 不確実性な世界をコントロールしようとする、理論では必ずしも上手くいかず、データベースを活用した確率論が非常に重要となる。医療・医学の世界は個のばらつきが極めて大きいので、実践結果のデータベースを構築し、それを科学的背景として活用していかなければならない。データベースからの科学で、個のばらつきに対応することは、医療的には「（個ではなく）サブグループに分けて医療をする」と表現するのが正しいだろう（図2-1-2）。
- ▶ 家電製品の失敗は、理論的アプローチの失敗であろう。よい製品は売れるはずだという理論で前に進み続けた結果である。市場は予測が困難なものであり、さまざまな要因が絡み合った複雑系の中で、よければ売れるのかということを考えるべきである（図2-1-3）。
- ▶ 医療・医学の世界は理論と実践の両方を踏まえないといけない。そこにデータベースが必要。例えば、合併症を予防する目的の薬物投与後の500日間を追って、脳卒中とか心筋梗塞のような合併症を追跡すると、偽の薬（プラセボ）を使った場合には合併症が15%ぐらい発生し、一方、予防薬の投与患者では、合併症の発生率が10%程度に減ったとする。この結果を簡単に解釈すると、「薬を飲んでいてよかったですね」というのは5%になる。しかしながら、この解釈は個については全く語っていない。薬を飲んでいない人（＝プラセボを飲んだ人）たちのほとんどは（85%）は発症しない（図2-1-4）。薬が効く人、効かない人、薬を必要としない人をデータベースで区別して、個別というよりもサブグループ化して、A群の人たちには薬Aを投与する。B群の人たちには別の薬Bを投与する、そういうことをするのが医療的な対応になる。
- ▶ メカニズム論のみで進むと、かえって悪い例がある。不整脈の多い人は突然死などの重大な事象（イベント）を起こしやすいので、積極的に不整脈抑制薬を投与した時期がある。しかしながら、結果を見ると、不整脈を抑えることで、かえって死亡率が高くなってしまった（図2-1-5）。こういう教訓が我々の現場にはある。
- ▶ 大きなばらつきを示す個人の集合に対して、有効性を示すものが薬であるが、それゆえに副作用を抑えることが重要となる。有効性、副作用という基準でのわずかな差をビジネスに持っていったのが欧米の製薬資本である。有効性が同じでも、副作用が少ないことがわかればそれだけで巨大市場を席卷できる。このような開発には大規模な臨床データ集めが重要であり、医療では、多くの場合ビッグデータこそがイノベーションに結びついている（図2-1-6）。
- ▶ データベース構築に関しては、医療・医学関係者は明確な姿勢を示してこなかったと思う。がんで亡くなった戸塚洋二先生（カミオカンデの物理学者）は、医療界はがんのデータベースをなぜつくらないのかと指摘している。物理学では、理論的に解明できないとき真っ先にデータベースを作り、詳しいデータを集めて、現象の全体像、バ

リエーションをとらえて現象の背後にある本質を抽出している（図2-1-7）。天文学でもデータベースを作り、因果を推測するのにベイズ統計など統計学の様々な手法が活用されている（図2-1-8）。

- ▶ ゲノムの世界では、レアバリエント（rare variant）が非常に注目されており、「軽い遺伝性疾患が高齢者疾患になる」という概念が出てきた。ただ、その有効性については、もっともっとデータを集めて検証する必要がある。さまざまな別のパラメータを組み合わせた方が、サブグループ化にはよいということもある（図2-1-9）。同じパラメータでゲノムが違ったときにどのような意義があるかという議論が重要である。ゲノム科学の進歩のためにも、患者の背景因子を揃えての解析が必要となる。そのため、臨床情報に関するさまざまなデータベースが必要である。
- ▶ 集団の中でニーズを汲み出して、応用して、また集団の中で検証するということが重要である。そのためには倫理や社会との対話、協働が必要である。そこでは、メカニズムの解明とともに、統計解析が重要となる（図2-1-10、11）。社会的な事象の解析に統計学的方法を導入したのは、ケトラーである。ケトラーの取り組みを医療の言葉に代えると、ばらつきを制御することによって、臨床試験が可能となる（図2-1-12）。
- ▶ 医療にとって重要な統計学だが、実は我が国には根付いていない。フィッシャーによって推測統計学が進歩したが（図2-1-13）、日本は非常に遅れた。1940年代になって北川敏男先生や増山元三郎先生が推測統計学を導入したが、戦後にイデオロギー論争に巻き込まれ、資本家のための統計学と攻撃された。推測統計学を用いる薬剤の研究は産学連携反対の雰囲気もあり、我が国では育たなかった（図2-1-13～16）。統計学が根付いていない我が国では、臨床開発が進みにくく、ファイザーは日本の研究所を閉鎖し、韓国に臨床研究の拠点をつくるために300億円の投資を行った（図2-1-17）。
- ▶ 我が国の統計学の遅れの象徴的な事件としては、最近、武田薬品の薬がフランスで、膀胱がんが多いと指摘された話をあげることができる。それが事実かどうか、よくわからない。しかし、これに対して武田薬品からの反論は一切なかった。自前のデータベースに基づいて反論してもらいたいと考える（図2-1-18）。
- ▶ 臨床データはこれからの基盤だが、怪しげなデータも増えてくる。それに対して常に反論しないと何が本当かわからなくなる。欧米では、巨大なデータベースがいくつも作られている（図2-1-19）。
- ▶ 医薬品の開発で、臨床試験を「死の谷」として見ることもある（図2-1-20）。「死の谷」を突破するために、非常に狭い適用範囲で承認を得た後に、適応拡大する「育薬」と言われるフェーズに入る（図2-1-21）。そこで蓄積した資本で次の基礎研究、あるいは薬剤開発に向かう。医薬品開発は、リニアな単純な世界ではなく、循環型の複雑なシステムである（図2-1-22）。「死の谷（臨床試験）」、育薬、課題設定、課題抽出など、医薬品の開発過程すべてにおいて、臨床データベースが必要である。
- ▶ 日本人の研究は、個人的な悟りの世界を目指して一生懸命頑張るが、システムの中で課題をとらえ、検証や評価、仮説の設定という循環を描かない。システムのことになると、なりゆきと勢い任せという感じがする。患者のデータベースを用いた臨床科学で、そこを修正しなければならない（図2-1-22）。

(質疑応答ならびに討論)

- ▶ 諸外国における活用可能な主要データベースの性質は少しずつ異なる。米国では保険会社を中心となったデータベースが多い。イギリスとか北欧系のデータベースは、医療制度の中に組み込まれており、開業医たちが国の制度の下に電子カルテを作成し、それが自動的にデータベースに蓄積されている。我が国でも、カテーテルだとかデバイスを使った患者さんの10万人とか50万人単位のデータベースがあり、保険償還制度の下に、入力に義務付けられている。保険制度下でデータベースを作れば、立ち上げのときは1億円ぐらいの投資が必要だが、維持は保険システムをうまく使うことになる。そうすれば、上手くデータを集めて、それをサイエンスの発展に結びつけることが可能となる。ただし、海外のデータベース利用は有償である。
- ▶ 我が国の場合、大規模なデータが集まっておらず、基本的には大規模データベースのない国と考えてよいだろう。我が国でも臨床系データベースを大型化する試みは続けられており、入院患者の診断群分類別包括制度（DPC）については、共通で使えるデータベースになりつつある。よりデータを集めるためには、共通番号（医療IDなど）の導入、データの形式の規格化が必要であるが、何よりも忙しすぎる我が国の臨床医にデータ入力の時間を作ってあげる制度改革が必要である。

医学・医療・イノベーションにおけるデータベースの必要性

自治医科大学
永井良三

1

図2-1-1

実証主義の限界

実証できない世界

実証できる世界

■科学

実証できないことには沈黙
実証できる世界では、普遍性に関心
個については無関心

■医学の進歩

決定論よりも確率論の世界へ
個別医療→正しくはサブグループ医療
実証できないことは、統計と価値観に依存

図2-1-2

実証主義の限界

実証できない世界

実証できる世界

■家電製品

よい製品は売れるはずだ

■現場の状況

市場（予測困難、複雑系、確率に支配）
個別の嗜好（価値観）

図2-1-3

集団に対する薬剤の効果は判明しても、個人に対する効果は予測できない

イベント発生率(%)

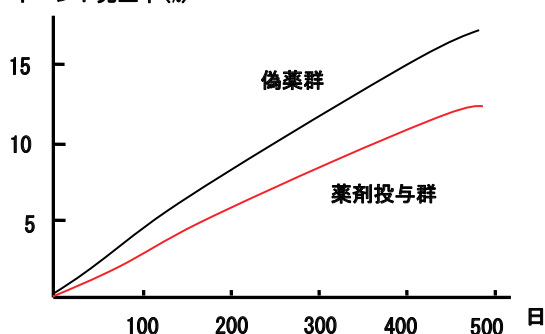


図2-1-4

不整脈を抑える薬によって、心不全患者の死亡率が増加することがある

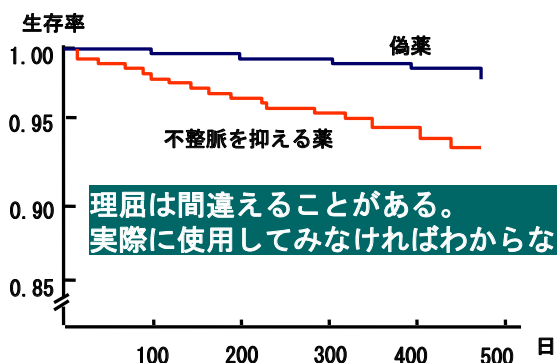


図2-1-5

どの治療が発作の頻度を減らせるか。 30人／千人／年間→24人／千人／年間

=600人→480人／千人／20年間

=6800万人→4800万人／1億人／20年間

■ばらつきを制御することで、わずかな差を示す
(高齢社会では大きな意義)

■有効性は同じであっても副作用が少なければ、
巨大市場を獲得できる。

図2-1-6



戸塚 医療界にお願いがあります。がんのデータベースが是非必要だと思うんです。

未知の現象に対して理論的に解明できていないとき、私のような物理学者なら、まっ先に何をするか。それはデータベースの作成です。詳しいデータを集め、解析し、現象の全体像、ヴァリエーションを捉える。こうして現象の背後にある本質を抽出していくわけです。

図2-1-7

データベース天文学の世界 天文学データ解析計算センター 水本好彦
異なった波長、装置、時間により観測された異なったデータの膨大な塊(数値宇宙)の中からあるパターンを探す。多数の天体について様々なパラメータ(n種類)を測定し、そのn次元パラメータ空間の中の天体の分布を調べる。分布図の中でデータ点が密集したクラスターの形や種類、塊にならないデータ点などを調べ、天体の分類を行う。

明るさや位置の時間変化→高赤方偏移のHypernovaやガンマ線バースト天体、クエーサーの時間変化、マイクロレンジングや系内惑星、太陽系近傍天体などの探索。

天体の2つの色の分布図→高赤方偏移のクエーサーや、冷たい白色矮星、褐色矮星など稀な天体の探索。

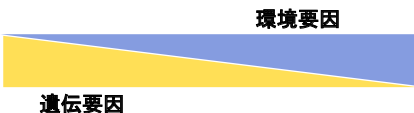
複数の色での天体位置の相関→cosmic string(宇宙紐)探索

大量データの統計処理を駆使、n次元パラメータ空間を2次元または3次元に投影して見せる可視化技術が威力を発揮するだろう。データベース天文学によって、宇宙年齢のような宇宙論パラメータを膨大なデータの統計処理で正確に決定するような**天文学の精密化**と、**未知天体の発見ラッシュ**が期待される。

図2-1-8

ゲノムは高齢者疾患をどこまで制御するか

- 1 軽症の遺伝性疾患が高齢者疾患として発症するか
- 2 それは通常疾患か、希少疾患か
- 3 生活習慣病において、ゲノムの果たす役割は疾患により大きく異なり、様々なパラメーターを組み合わせた判断が必要

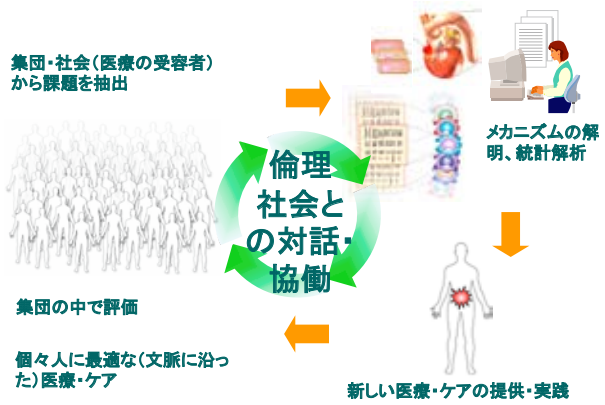


環境要因

遺伝要因

図2-1-9

臨床医学研究のあり方



集団・社会(医療の受容者)から課題を抽出

倫理・社会との対話・協働

メカニズムの解明、統計解析

集団の中で評価

個々人に最適な(文脈に沿った)医療・ケア

新しい医療・ケアの提供・実践

図2-1-10

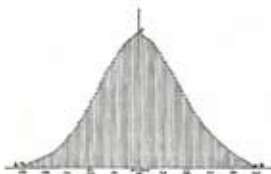
人間集団のデータの重要性

- 1 人間を理解する。
- 2 人間の営みの中に法則性を見出す。
- 3 課題の抽出
- 4 人間の営みの検証
- 5 アンメットニーズに基づく開発

図2-1-11



アドルフ・ケトレー
1796-1874



個人の行動や社会の活動における法則性

**ばらつきを制御
→臨床試験**
偶然にしてはどの位、稀な現象かを評価する

図2-1-12



推測統計学

p値
帰無仮説が正しいと仮定したときに、観察された事象が起こる確率。

稀に発生する現象や、群間でみられるわずかな差がばらつきや偶然によるのではないと推測するための理論的基礎を構築

フィッシャー
R. A. Fisher
1890-1962

図2-1-13

独占資本の要求に応じて育成させられた推測統計学が、唯物弁証法に立つと主張するからには、何時、何処で性格転換を行なって労働者階級の要求に応じるようになったか、また、唯物弁証法を利用すると云うなら、独占資本の立場から労働者階級を利用すると解してよいのか、この辺は明確にしてほしいところである。推測統計学それ自体の構造や性格から必然的にこういったものが出たのでないとすれば、唯物弁証法のかわりに戦時中のように皇道主義をかつぎ出しても事態は変わりはないさそうである。そうだとすると危機の断化における進歩的で粉飾した有産者思维の一形態ではないか、と推定される「危険率」がきわめて大となってくる。

大橋隆憲 現代統計思想論

図2-1-14

数理統計学の日本への導入は、社会統計学よりやや遅れた。藤沢利喜太郎1861-1933は、1896年に東京法科大学で確率論を基礎とする統計学を講義し、1919年から東京帝国大学理学部で数理統計学を開講した。それ以後しばらくの間、日本でもK. Pearson流の数理統計学が盛んであった。母集団と標本の概念を強調するFisher流の理論が導入されたのは1930年代で、佐藤良一郎、北川敏男、増山元三郎、河田竜夫、等により宣伝された。1941年に北川敏男を委員長とする「統計科学研究会」が生れた。この会の幹部は軍部への協力を約し、軍部の支持の下に1944年6月、文部省に「統計数理研究所」を設置せしめた。…好戦的数理統計家は戦争の提出する応用数学的問題に得々と取り組んだ。

大橋隆憲 現代統計思想論

図2-1-15

労働者階級の立場に立つ統計学の建設の動向に対して、労働者階級の立場を本質的に認めようとせぬ日本の反動的な資本家階級とそのイデオロギーは、いろいろな形態で反撃を加えるであろう。彼等は「経営」に直接役立つ学問を大学に要求し、大学もまたその要求に応じて実用主義化の傾向を強めつつある。かかる状態にあればあるほど、高野岩三郎以来の、日本の社会統計学の伝統をかついで、労働者階級の立場に立つ確たる統計理論の創造的建設が緊要な課題となる。

大橋隆憲 現代統計思想論

図2-1-16



図2-1-17

例えば、糖尿病薬で膀胱癌？

- 医薬品：ピオグリタゾン塩酸塩
- 販売名：アクトス
- 欧米の医療情報データベースを活用した薬剤疫学研究の結果、ピオグリタゾン塩酸塩使用者において、膀胱癌発症リスクが高まる可能性を指摘。
- KPNC研究（米国）
会員制医療保険組織KPNC登録患者の10年間コホート研究
コホート 193,099例
うち本剤使用者での膀胱癌 90例/ 30,173例
非使用者での膀胱癌 791例/162,926例
- CNAMTS研究（仏国）
保険データベースSNIRAMに登録された糖尿病患者の後向きコホート研究
コホート 1,491,060例
うち本剤使用者での膀胱癌 175例/ 155,535例
非使用者での膀胱癌 1841例/1,335,525例



図2-1-18

DB/ DB運用組織名	国	規模	含まれるデータ	備考
GPRD	英	657万人	診療情報、処方、患者情報、検査結果等	MHRAが管理運営するDB 一般診療所488施設より情報収集
THIN	英	500万人	診療情報、処方、患者情報 等	EPICがGPRDの代替として構築したDB GP300人より情報収集
PHARMO	蘭	200万人以上	診療情報、処方、検査結果 等	ユトレヒト大学、ロッテルダム大学が構築したDB
IMS Disease Analyzer	英独仏豪	1570万人	診療情報、処方、患者情報、医師情報等	IMS Health社が構築したDB GP3600人より情報収集
i3 Aperio	米	3900万人以上	診療・処方レセプト、患者情報、検査結果等	保険会社コナティッドヘルス・グループの1部門であるi3のDB
Kaiser Permanente	米	860万人以上	診療・処方レセプト、患者情報、検査結果等	米国最大の非営利総合医療団体であるKaiser PermanenteのDB。7地域にリサーチセンターがあり、それぞれ独自のDBを所有する
HMO research network	米	4000万人以上	診療・処方レセプト、患者情報等	カイザーを含む14の保険会社のコンソーシアムが収集したレセプトデータのDB

図2-1-19

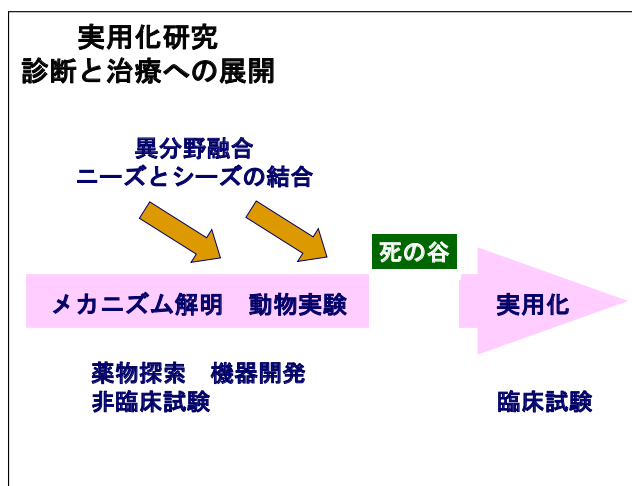


図2-1-20

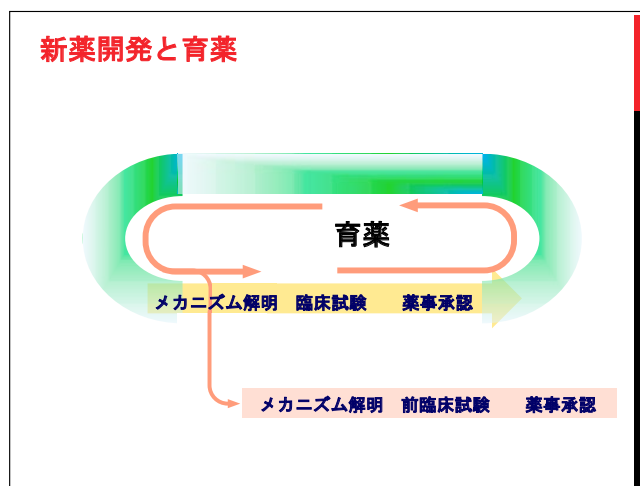


図2-1-21

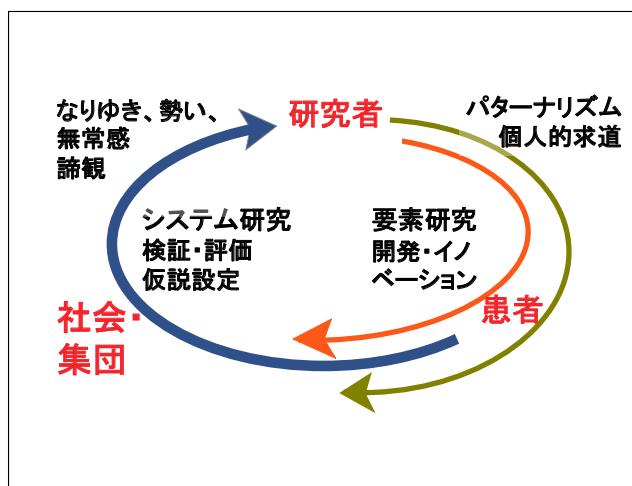


図2-1-22

基礎生物学の視点から

高木 利久 先生（東京大学大学院新領域創成科学研究科教授、NBDC 副センター長）

- ▶ 基礎、ゲノムを中心としたデータの共有・統合という観点でお話しする（図2-2-2）。
- ▶ 新型センサーによりデータの爆発が起きており、コンピューターの性能の伸びを超える状況。ペタ、10の15乗ぐらいのオーダー。知識の爆発も起きている。論文は約2,000万件。毎年数10万件の論文が出ている。データベースも爆発している（図2-2-3）。
- ▶ ミレニアムプロジェクト以降、大規模なデータ産出のプロジェクトが進められている。仮説駆動型からデータ駆動型（データインテンシブなど）になってきている。一方でデータを囲い込んでしまう傾向も強まっている。そのような中でイノベーションにつなげていくためには、狭い意味での公開、単にデータがブラウズできるだけではなくて、丸ごと活用できるような仕組みをつくっていかないといけない。データ生産者以外の不特定多数のイノベーターがいる、という状況である（図2-2-4）。
- ▶ 外国では、米国NCBI、欧州EBI、データベースセンターとしての中国BGIなど、いずれも数百人規模のバイオインフォマティクス研究者のセンター。プロジェクトとしても米国でビッグデータのイニシアチブ、ニューロインフォマティクス関係のBIRN、がん関係のインフラストラクチャ、グリッドコンピューティング関係のようなプロジェクト、欧州のELIXIRなどがある。共有ルール、データのセキュリティをどう保つかといったガイドラインがさまざま出ている。税金で行われたプロジェクトの成果はだれに帰属するかといった観点での法律等も出ている（図2-2-5）。
- ▶ 我が国では、平成18年頃からCSTPで議論が始まり、文科省、経産省、農水省、厚労省を中心にライフサイエンスの統合データベースプロジェクトがスタートした。文科省のプロジェクトの目的は、データを公共財化、コモンズ化し、その価値を最大化することである。平成19年に情報・システム研究機構（ROIS）に設置されたライフサイエンス統合データベースセンター（DBCLS）、平成23年にそれを拡張する形でJSTに設立されたバイオサイエンスデータベースセンター（NBDC）を中心に活動が展開されている。クリエイティブ・コモンズの採用、統一的なフォーマット、辞書などを整備している。分野ごとのデータベースを統合化する目的で、NBDCプロジェクト採用時には、データ提供を義務化している。文科省ライフ課のプロジェクト、CREST、科研費、厚労科研費などにもNBDCへのデータ提供が記載されている。文科省プロジェクトとして、データの共有ガイドライン、特にヒトのデータの扱いに関するガイドラインが作成されつつある（図2-2-6）。
- ▶ DBCLSとNBDCのホームページには、月に10万人弱、230万程度のアクセスがある。我が国のデータベースをほぼカバーするポータルサイトとなっている（図2-2-7、8）。
- ▶ 次に、医療・医学についてデータベースの観点からお話する。大規模なコホート研究はもう数十年前から始まっている。一番有名なフラミンガムは1万人規模のGWASで、さまざまな成果がでている。英国でも大規模コホートが実施されている。我が国では久山町、長浜、今年スタートの東北メディカルメガバンクなどのコホート研究がある。東北メディカルメガバンクとNBDCは、データ公開、データ解析で連携している（図2-2-9）。

- ▶ ケースコントロール研究からは、2型糖尿病、リウマチ、アルツハイマーなどを対象としてGWASによる疾患関連のSNP発見が続々と発表されている（図2-2-10）。
- ▶ 幾つかのケースコントロールのデータを合わせたメタ解析も進んでいる。16グループ、18グループのデータを集めて、規模を大きくして解析し、新たにSNPが見つかるというとも言われている（図2-2-11）。
- ▶ 東大の徳永先生の事例では、1施設の場合は高いオッズが出るが、多施設の場合はそれが大分下がってしまう。他施設のデータを足すと治療不応答群と治療応答群のところの差が小さくなってしまう、という話もある。ビッグデータを扱うときの品質管理は非常に重要で、診断基準、データの品質管理、それから集団の補正、そういうことを行わないとうまくいかない（図2-2-12）。
- ▶ 海外ではデータの共有化が進んでいる。NCBIではSNP、構造多型、あるいはジェノタイプ、フェノタイプ関係のGWASのデータ等が収集、共有化されている。欧州のEBIでも似たようなデータベースがつくられている。オーストラリアがサポートして、最近中国がファンドしている、ヒューマンゲノムバリエーションソサエティから地域特有のデータベースが多数出ている。これ以外にも、国際がんゲノムコンソーシアム、1000人ゲノム等で大量のデータが出て、公開、共有されている（図2-2-13）。
- ▶ 米国のアルツハイマー研究プロジェクト、それに引き続いたパーキンソン病のプロジェクトでは、即座にデータを公開し、データの所有を認めない、パテントも認めない、そのようなやり方をして非常に研究が進んだと言われている。このようなことがデータの権利関係をすっきりさせるために非常に重要だろうと思っている。
- ▶ NBDCでも分野ごとの統合化をしている。例えば、東大の岩坪先生のヒト脳疾患画像データベース、東大の徳永先生のヒトゲノムバリエーションデータベースの開発、京大の松田先生の長浜のコホート研究など、メディカルなテーマのデータベースの共有化、統合化を進めつつある。先行するヒトゲノムのバリエーションデータベースに関しては、データ統合化が進んでいる（図2-2-14）。
- ▶ バリエーションデータベースは、徳永先生を中心に行われているもので、GWASのデータ、それからヒューマンバリエーション、ヒトのリシーケンスのデータ、コピー数変化（CNV）のデータ、さまざまなものを公開している（図2-2-15）。公開しているものは10疾患13スタディ、公開準備中のものが36疾患50スタディある。50万から100万ぐらいのSNPを解析できるチップを利用した実験結果を掲載している。検体数は1スタディで200から1,500ぐらいである。このようなデータの共有化、活用が、今後の研究を進める上で重要と思っている（図2-2-16）。
- ▶ データベースの統合的活用に関する課題は、データを構造化しないといけない点にある。ゲノムの配列やタンパク質の構造というデータは統合化しやすいが、例えば遺伝子発現の場合、その文脈的なことを入れないといけない。その文脈はどうしても言葉で書く必要がある。世界的に標準化がそれほど進んでいない。パスウェイとか相互作用のデータベースも300以上ある。標準化の動きはあるが、まだまだ整備が進んでいない（図2-2-17）。
- ▶ 標準化されていない、メタデータがない、品質管理できていない、解析ツールによって2次データベースが異なったり、オントロジーが違う、などの原因があり、様々なデータベースがあっても、それらを統合的になかなか使えない状況である。このよう

なものをどう使いこなすかということで、NBDCも取り組んでいるが、後付けで整理するのは難しい。プロジェクトが始まる段階から標準化、統合化を念頭に入れた作業を進めておくことが重要である（図2-2-17）。

- ▶ バイオインフォマティクスとかバイオスタティスティクスについては、人材が非常に不足している（図2-2-17）。
- ▶ データ共有、特にヒトのデータをどううまく扱うのかというガイドラインを我が国で整備していくことも重要。NoSQLはいわゆるビッグデータを扱うときに、従来の関係データベースではなかなか扱えないデータがあるということで、このようなデータベースが今、幾つも出てきている。どのように使えばどんなことができるのか、何が難しいのかということ、今、技術的に詰めているところである。次世代シーケンサー対応で、大規模なスパコンが必要なように思われているが、必ずしもそうではないこともわかってきた。ビッグデータと呼ばれるものは何か特別な処理をしないといけないと思われているが、割とバイオ、ライフの分野は単純並列でいいケースが多く、いわゆるHadoopでの分散処理、NoSQLでのデータ管理で対応が可能ということが徐々にわかってきている（図2-2-17）。

（質疑応答ならびに討論）

- ▶ 決定論的な知見であれば、少数例でいいというケースもあるが、大量のデータを揃えておかないと何が本当かわからない。やはりまずデータを大量に扱って、そういう解析をやってみないといけない。多くの研究は、やはり確率論に入っていくのではないかな。
- ▶ それぞれのメタデータの質、診断基準を含めて、バックグラウンドを揃えたものを集めなければならない。そういった点に注意をして、最終的には統合的に活用したらいいだろう。

2012年10月31日
JST-CRDS
DB統合的活用WS

ライフサイエンスにおける オープンイノベーションのための データの共有と統合

東京大学 大学院新領域創成科学研究科
情報生命科学専攻

科学技術振興機構
バイオサイエンスデータベースセンター

情報・システム研究機構
国立遺伝学研究所DDBJセンター

高木利久

図2-2-1

発表の内容

- データ、知識、DBの爆発の現状
- データ共有、データ統合の必要性、重要性
- オープンイノベーションに向けた取り組み
 - 米国、欧州、中国、日本
 - センター、プロジェクト、共有ルール
- 医科学における活用、成果
- DB統合的活用に向けた課題

図2-2-2

データ、知識、DBの爆発

- データ爆発
 - 超高速ゲノムシーケンサー
 - コンピューターの性能の伸び(ムーアの法則)を凌駕
 - 1万を超えるゲノムプロジェクト進行中
 - 画像、動画データも急増中
- 知識爆発
 - 論文数 2,000万件
 - フルペーパーも利用可
- データベース爆発
 - 世界 1万
 - 日本 1千
 - 解析ツール 2千
 - 内容也多様化

ライフサイエンスは
peta(10^{15} =千兆)
のデータ量の時代に入

図2-2-3

ライフ分野におけるデータの共有・統合の重要性

- データベースを目指した大規模プロジェクト(ミレニアム)
 - 多額の研究費がデータに化けている
 - ゲノムコホート研究など情報爆発を加速化する動き
- 仮説駆動型からデータ駆動型(データ中心科学)へ
 - 法人化、ハイドロ法 → データの囲い込み
- データや知識の分断、断片化
- オープンイノベーションには共有・統合化必要
 - ブラウズ(検索)や一部のデータ公開では不十分
 - データ生産者以外の不特定多数のイノベーターの存在
 - 再利用、転用、マッシュアップが自由にできることが不可欠

図2-2-4

外国での取り組み

- DBセンター
 - 米国NCBI、欧州EBI、中国BGI
 - いずれも数百人規模の陣容、スパコンも
- プロジェクト
 - 米国 ビッグデータイニシアティブ、BIRN、caBIG
 - 欧州 ELIXIR
- 共有ルール、セキュリティガイドライン(例)
 - 米国 dbGaP Best Practices Requirements
 - 欧州 Wellcome Trust Sanger Institute DATA SHARING GUIDELINES

図2-2-5

我が国における取り組み

- 内閣府主導の統合データベースプロジェクト(H18)
 - 文科省、経産省、農水省、厚労省
 - 四省連携のポータルサイト
- 文科省の統合データベースプロジェクト(H18)
 - データを公共財化・コモンス化し、その価値を最大化
 - 中核センターの設立
 - H19～ライフサイエンス統合データベースセンター-DBCLS
 - H23～バイオサイエンスデータベースセンター-NBDC
 - CCライセンスの採用、統一フォーマット、辞書などの開発
 - カタログ、横断検索、アーカイブなど
 - 分野ごとのデータベース統合化進行中(現在11課題)
 - データ提供要請の公募要領への記載(ライフ課、CREST、科研費)
 - ヒト由来データの共有、セキュリティガイドラインの作成

図2-2-6



図2-2-7

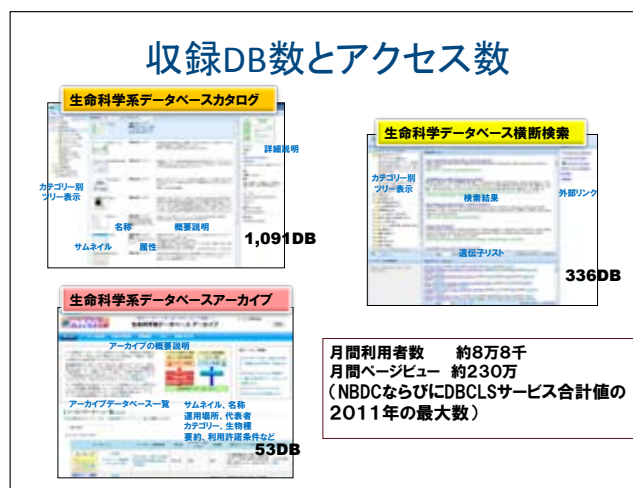


図2-2-8

大規模なコホート研究

- Framingham Heart Study (1948年～)
 - 心不全・心筋梗塞・脳卒中のコホート研究
 - 1万人規模のGWAS、血圧、血中脂質等の因子を発見
- British Birth Study (1946年～)
 - 1946、1958、1970、2000年生まれのコホート
 - 出生時体重、身長、学業成績、親の経済状況、本院の健康と経済状況を長期に調査
- 久山町 (1961年～)
 - 久山町の年齢、職業分布などが全国平均に近く平均的な日本人集団
 - 尿検査、血圧、貧血、心電図、脂質などのデータを収集
 - 剖検率も高く正確な死因
- 東北メディカルメガバンク (2012年～)
 - 7万人の住民コホート、8万人の3世代コホート
 - NBDCからデータの公開予定

図2-2-9

ケースコントロール研究

- 2型糖尿病
 - 日本人: 東大附属病院。理研他グループ (nature genetics, 2010)
 - フィンランド人: フィンランド国研、ミシガン大他グループ (science 2007)
 - Wellcome Trust Case-Control Consortium (WTCC) (nature genetics, 2007)
- リウマチ
 - Brigham Rheumatoid Arthritis Sequential Study (BRSS) CANADA (nature genetics, 2007)
 - Epidemiological Investigation of Rheumatoid Arthritis (EIRA) (カロリンスカ研)
 - WTCC (nature genetics, 2007)
- アルツハイマー
 - カーディフ大を中心としたグループ (nature genetics, 2009)
 - フランスを中心としたグループ (nature genetics, 2009)
 - WTCCC (nature genetics, 2007)

いずれもGWAS(genome wide association study)で疾患関連SNPを発見

図2-2-10

ケースコントロール研究のメタ解析

- 2型糖尿病のメタ解析
 - 16グループのデータのメタ解析 (Caucasian) 約2.8万人のcase (nature genetics 2008)
 - 18グループのデータのメタ解析 (East Asian) 約2.5万人のcase (nature genetics 2011)
- リウマチのメタ解析
 - 13グループのデータのメタ解析 1万人以上のcase (nature genetics, 2010)
 - 日本人データのメタ解析 1万人のcase (nature genetics, 2012)

いずれも、メタ解析によって新たな疾患関連SNPを発見

図2-2-11

データの品質管理の重要性

- 診断基準は多様であり、診断基準を合わせないと、疾患関連因子は見つからない。
- 大規模データはエラー多く、様々な角度から品質管理をする必要がある。
- メタ解析は、集団の構造化を補正する必要がある

QCなしではFalse Positive, False Negativeの増加

図2-2-12

海外での共有化の取り組み

海外での関連した主な取り組み

- 1) NCBI
 - ・dbSNP SNP情報を蓄積
 - ・dbVAR 構造多型のデータを蓄積
 - ・dbGAP GWAS, 次世代シーケンサー結果を含むgenotype-phenotypeに関するデータを蓄積
- 2) EBI
 - ・EGA GWAS, 次世代シーケンサー結果を含むgenotype-phenotypeに関するデータを蓄積
- 3) HGVS (human genome variation society)
 - ・LSDB (locus specific database) のリンク集
- 4) その他、PJ単位のDB
 - (1000genome PJ, 国際がんゲノムコンソーシアム等)



図2-2-13

H23年度採択

研究開発課題名	研究代表者	所属・役職
ヒト疾患関連データベース統合化研究	岩坪 威	東京大学 大学院医学系研究科 教授
メタボローム・データベースの構築	金谷 重彦	奈良先端科学技術大学院大学 情報科学研究科 教授
ゲノム情報に基づく疾患・医薬品・環境物質データの統合	金久 貴	京都大学 化学研究所/バイオインフォマティクスセンター長・教授
ゲノム・メタゲノム情報と基盤とした微生物DBの統合	黒川 順	東京工業大学 大学院生命理工学研究科 教授
ゲノム情報に基づく植物データベースの統合	田嶋 智之	かずさDNA・エヌ・エー研究所 副所長
ヒトゲノムバリエーションデータベースの構築	渡永 勝士	東京大学 大学院医学系研究科 教授
生命と環境のフェノーム統合データベース	豊田 智郎	理化学研究所 生命情報基盤研究部門 部門長
蛋白質構造データベースの国際的な構築と統合化	中村 春木	大阪大学 蛋白質研究所 教授
植物統合データベースと研究支援ツールの開発	成松 久	産業技術総合研究所 植物園工学研究センターセンター長
大規模ゲノム医学研究の統合情報基盤の構築	松田 文彦	京都大学 大学院医学研究科 附属ゲノム医学センター長・教授

H24年度採択

研究開発課題名	研究代表者	所属・役職
生命動機システム科学のデータベースの統合化	大瀧 修一	独立行政法人 理化学研究所・生命システム研究センター 生命動機研究チーム チームリーダー

図2-2-14

NBDC統合化推進プログラムによる GWAS, NGSデータの共有化



図2-2-15

ヒト多型DBの収録データ

- GWAS 10疾患 13 study
- 公開準備中 36疾患 50 study
- 50-100万SNPのチップを利用
- 検体数は1 study 200検体から1500検体程度
- GWASの健常者用のSNP は600検体
- human variation DBの変異は1万件以上

図2-2-16

DBの統合的活用に向けた課題

- ゲノム配列、タンパク質構造以外は統合化なかなか困難
 - 文脈依存データはメタデータの記載・標準化が不可欠
 - しかし多くは未整備
 - 例) パスウェイ、相互作用DBだけでも300以上のDB、表記法
- 活用可能なデータは(現時点では)必ずしも多くない
 - 標準化されていない、メタデータがない
 - 品質管理できてない
 - 解析ツールの部品化、オントロジーの整備必要
- プロジェクト開始時から標準化、統合化を指向すべき
 - プロジェクト終了時に標準化、統合化考えても遅い
 - 各プロジェクトにバイオインフォの専門家の配置すべし
- バイオインフォ、バイオ統計学者の人材不足の解消
- データ共有に向けたガイドラインの整備
- 画像データの解析、アーカイブ

図2-2-17

臨床系 DB の視点から

山本 隆一 先生（東京大学大学院情報学環准教授）

- ▶ ナショナルレセプトデータベース（NRDB、厚労省が管理している診療報酬明細データベース）、特定健診、特定保健指導の全数データベースの活用のお話する（図2-3-1~4）。
- ▶ 薬剤疫学ではカナダのデータベースが、例数は少ないけれども一番よく使われている。GPRDやTHINは基本的には特定の診療所と契約して、そのデータを全部集めている。カルテIDと結びつけられており、深く活用することができる。それに比べてアメリカは、保険者が加入者に対する何らかの知見を得るためのデータベースであるため、数は多いが内容はそれほど濃くはない。結構開放されているようである（図2-3-5）。
- ▶ 我が国では、医療機関が出すレセプトの99%、薬局のレセプトの99.9%、歯科の約半分が電子化されていて、その電子化されたレセプトを全部収納する。特定健診は、当初から電子的に収集することが決められており、対象が4,500万人のうち毎年2,700万件ぐらいが情報として得られ、格納されている。レセプト件数は年間18億レコード（図2-3-6）。
- ▶ NRDBは、もともと高齢者の医療確保に関する法律で作成が決められたデータベース。医療費適正化計画の作成が目的。厚労省もしくは都道府県しか使うことがなかったが、データベースをつくるときに、かなり有用なデータベースなのでそれ以外にも使えるようにしなければいけないという提言が出され、その提言に従って本来目的以外の利用が検討され、ガイドラインがつけられた。現在、有識者会議で審査をして、提供するかしないか助言をしている。最終的には厚労大臣の決定になる（図2-3-7）。
- ▶ 国は、個人が直接識別できる情報を持ちたくないということで、審査支払機関から出るときに一定の個人識別にかかわる情報をハッシュ化している。NRDBに入るときにもう1回ハッシュ化して、どちらも連結テーブルを持たない状態にしてレセプトに格納している。従って、例えば、ある薬の副作用が発見されたとしても、それを使っている患者さんには戻れないという状況となっている（図2-3-8）。
- ▶ ハッシュ化により容易に個人は特定できないが、では絶対できないと言われると、結構微妙な問題がある。この有識者会議では、「個人情報でないとは言えない」という立場でこのデータベースを扱っている（図2-3-9）。
- ▶ ハッシュ値が2つあるので、一応同じ人のデータは結びつけられるという前提に立つ。しかし、レセプトの中に書かれる番号も名前も医療機関によっていろいろで、例えばスペースが空いているところを2バイトにするか1バイトにするか、ダッシュをどう表現するかといったことがあり、実際はなかなか完全には一致しない。レセプト同士だと一致率が大体8割ぐらい。レセプトと特定健診になると、半分以下になる（図2-3-10）。
- ▶ 「匿名化データか？」ということが議論され、いわゆる疫学的研究倫理指針に言う匿名化された情報とは言えないという結論に達している。医療機関コードは入っているため、医療機関は特定できるデータベースになる（図2-3-11）。
- ▶ こういった診療報酬請求書のデータベースはアジアに3つある。韓国のHIRAデータ

ベース、台湾のNHIリサーチデータベース、それから日本。台湾、韓国は日本より少し早く、対象人口は図2-3-12のとおり。学術目的の利用は、それぞれ認めているが、韓国が一番厳しい。国のプロジェクトに限られていて、なおかつ厳しい審査がある。台湾は相当開かれていて、研修を受ければ利用できる。ただし、全数入っているデータベースに関してはオンサイトセンター以外では利用できない。我が国は、2014年度までを試行期間として、審査をして提供している状況（図2-3-12）。

- ▶ データ更新は、韓国は毎月で、日本の場合レセプトは毎月、特定健診は毎年。台湾は毎月ではなくて、年に1回更新。生年月日は、日本は日が入っていない。日本は、身長／体重等はDPC（包括支払いの請求書）か特定健診、保健指導には入っているが、いわゆる出来高払いのレセプトには入っていない。処方情報に関しては比較的きちっと入っているが、ワクチンに関しては保険適用でなければ入らない（図2-3-13）。
- ▶ 手術は独自コードを使っている、入院日はわかる、診断コードはICD-10準拠である、それが台湾の場合はICD-9であるといった違いがある。こういうデータベースが整備されている（図2-3-14）。
- ▶ こういうデータベースは本当に役立つのか、試行した結果の紹介。産業医大の松田晋也先生が行ったもので、二次医療圏ごとの疾患と患者の動態を調べている。全傷病では、その患者がいる二次医療圏の中が最も多いというデータが出ている（図2-3-15）。
- ▶ 脳梗塞も大体同じ傾向で、脳梗塞の場合は二次医療圏という制度自体が比較的有効に機能している（図2-3-16）。
- ▶ 心筋梗塞も、ちょっとばらつくが、二次医療圏で閉じている場合が多い（図2-3-17）。
- ▶ 乳がんでは全く二次医療圏で閉じない。こういう疾患に対して医療の情報提供や計画を策定するときには、二次医療圏は適切ではないということがわかる（図2-3-18）。
- ▶ 第1回の提供審査では応募が43あり、そのうち6つしか許可されなかった。この有識者会議には、レセプトを出す側の医師会、薬剤師会、歯科医師会が含まれる。医療機関コードが入っているため、ある医療機関がどういう診療をしているかが全部わかってしまうことになるので、かなり大量のデータを網羅的に行う、探索的研究というのは相当強い反対があり承認されることが難しかった。現在は、こういう提供に関して法的裏づけがなく、契約による安全性確保しかできないということから、条件をどうしても厳しくせざるを得ないということになっている（図2-3-19）。
- ▶ 多分それでは非常に不満が多いだろうということで、サンプリングデータセットをつくっている。まずは当座のニーズを満たすために、このような条件のサンプリングデータセットをつくって、申請のあった6つの研究グループに提供している。さらに目的別で、例えば糖尿病、がんなど、今年度、来年度で作成して提供する予定である（図2-3-20）。
- ▶ 平成26年度にデータベース自体を更新するが、いろいろな問題が明らかになってきた。制度を整備する必要がある（図2-3-21）。
 - 厚労省との契約によりデータを提供することになっているので、例えば東京大学だと全部局の案件を総長と契約しなくてはいけない。
 - 目的外に使ったとしても、どういう罰則をなし得るかということ、多分名前を公表するとかそのぐらいしかできない。
 - データベースの構造が、早く入力することを目指したものとなっていて、利用のほう

が効率的ではない。抽出にかなり時間がかかり、ぎりぎり回っている状態。

- 中途半端なデアイデンティフィケーションがこのままでいいのだろうか。
- レセプトは行った行為を全部カバーできる病名をつけないといけないという特性がある。したがって、請求に関係ない病名は落ちてしまう、あるいは若干拡大解釈した病名がつく、といった問題があり、診断に関する実態は余り正確でないといわれている。ただし、希少疾患に関して言うと、それなりの根拠を持ってつけられるわけで、比較的信頼性が高くなるので、そのような例では個人が識別できてしまうということでジレンマに陥っている。プライバシー・プリザービング・データ・マイニングを導入して何とかできないか。
- 研究者による情報の安全管理は信頼できない。NRDB情報を提供した17件で、情報管理状況の監査を厚労省が行った。すべてで何がしかの問題がある。軽微な問題だけの場合もあるが、中には、これでは盗まれても仕方ないという状況もあった。研究者は、情報管理のプロではない。安全な環境を提供する必要があると考え、オンサイトリサーチセンターの必要性も検討されている。

(質疑応答ならびに討論)

- ▶ 法的な整備をどうすればいいのか。高齢者の医療の確保に関する法律では「医療費適正化計画の作成のため」としか書かれていない。「公益利用」「学術利用」などと明記し、データを利用することの意義をもう少し正確に伝える努力、あるいはそれを立法化していくことが必要なのではないか。目的外利用のような違反があった場合の対応も法的にしっかり決めることが重要。
- ▶ 高齢者医療確保法、個人情報保護法も、基本的には情報を取得した者がリスクを負いながら提供していくという形の法律になっている。法的に問題となる部分を、どうやって小さくするかという方法を議論しなければならない。基本的には使えるという状況にした上で、ある一定レベルからは制限を加えるような形、つまりは現状の原則と例外を逆転するような方式をとっていくことも検討すべきである。
- ▶ 医療の個人情報保護法がつくられようとしているが、大規模データベースを議論するならば、これまでの個人情報保護法の構造を少し超えるような発想がいる。現行ガイドラインでは大量のデータを集められない。法律事項で考慮することになるだろう。

ナショナルレセプトデータベース活用の現状

2012年10月31日
東京大学大学院情報学環 山本 隆一

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-1

医療IT化の進展と行政の取り組み

- > 1970代 医事システム、レセプトコンピュータの普及
- > 1980代 オーダエントリシステム
- > 1990代 電子カルテの実験的開発
- > 1999 診療録等の電子媒体への保存
- > 2001 医療のIT化のグランドデザイン
- > 2003 外部保存容認(制限付き)
- > 2005 個人情報保護法、E-文書法、医療情報システムの安全管理GL、HPKI基本ポリシー
- > 2006 IT新改革戦略 (On-lineレセ電、EHR-DB、HPKI・・・)
- > 2009 i-Japan戦略2015、社会保障カード実証事業、Power Shift
- > 2010 民間事業者による外部保存容認通知

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-2



Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-3

PMDA MIHARIプロジェクト

- > Medical Information for Risk Assessment Initiative
- > 21年度～23年度試行調査、24年度～本調査
- > レセプトデータを用いた使用実態および安全対策措置の効果に関する試行調査
- > レセプトデータを用いた副作用発現状況に関する試行調査
- > レセプトデータを用いた機械的シグナル検出の検討
- > 病院情報システムデータの試行調査
- > DPCデータを用いた試行調査

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-4

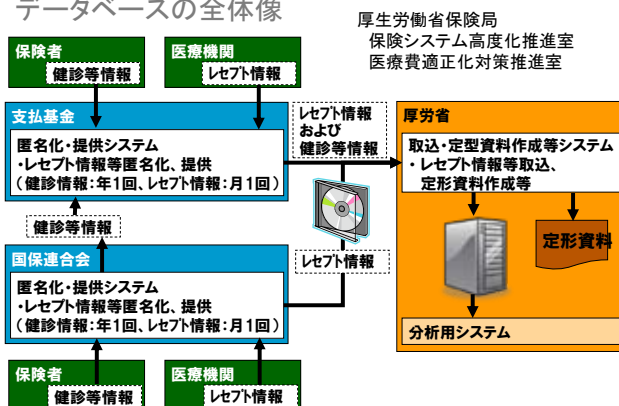
海外主要データベース概要

DB/DB運用組織名	国	規模	含まれるデータ	備考
OPRD	英	657万人	診療情報、処方、患者情報、検査結果等	MHRAが管理運営するDB 一般診療所400施設より情報収集
THIN	英	500万人	診療情報、処方、患者情報等	EPICがOPRDの代替として構築したDB GP300人より情報収集
PHARMO	オランダ	200万人以上	診療情報、処方、検査結果等	ユトレヒト大学、ロッテルダム大学が構築したDB
BMS Disease Analyzer	英・独・仏・豪	1570万人	診療情報、処方、患者情報、薬剤情報等	BMS Health社が構築したDB GP3000人より情報収集
id Aperio	米	3900万人以上	診療・処方レセプト、患者情報、検査結果等	保険会社ユナイテッドヘルス・グループの1部門である3のDB
Kaiser Permanente	米	860万人以上	診療・処方レセプト、患者情報、検査結果等	米国最大の非営利総合医療団体であるKaiser PermanenteのDB、7地域にリサーチセンターがあり、それぞれ独自のDBを所有する
HMO research network	米	4000万人以上	診療・処方レセプト、患者情報等	カイザーを含む14の保険会社のコンソーシアムが収集したレセプトデータのDB
id of Medicare, Medicaid	米	4230万人 4930万人	診療・処方レセプト、患者情報等	アメリカの公的医療保険制度の会員登録DB Center for Medicare and Medicaid Services (CMS)が両者を包括的に統括している
Health Services Databases in Saskatchewan	カナダ	100万人	診療・処方レセプト、患者情報等	Saskatchewan州地方保健局が保険情報より構築したDB

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-5

レセプト情報・特定健診情報等データベースの全体像



Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-6

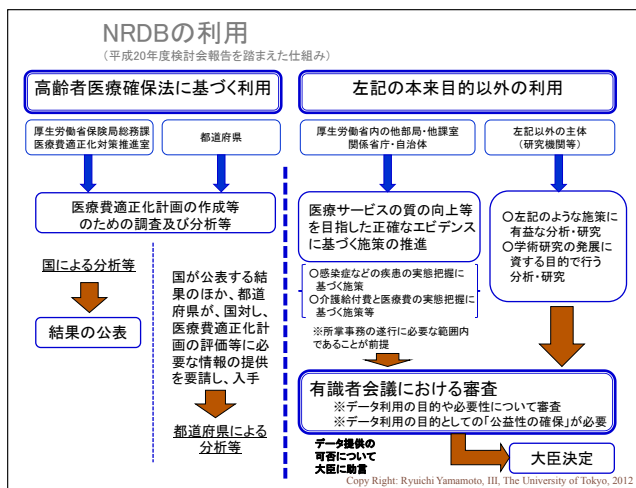


図2-3-7

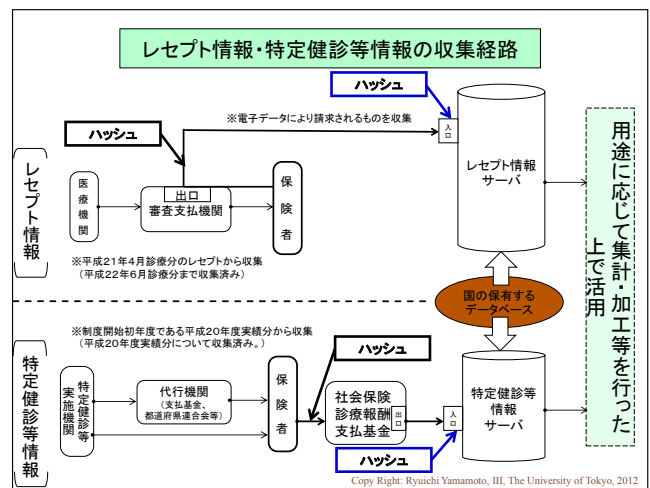


図2-3-8

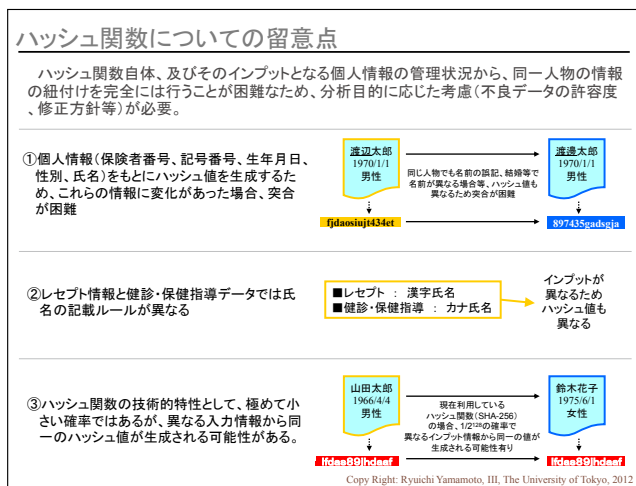


図2-3-9

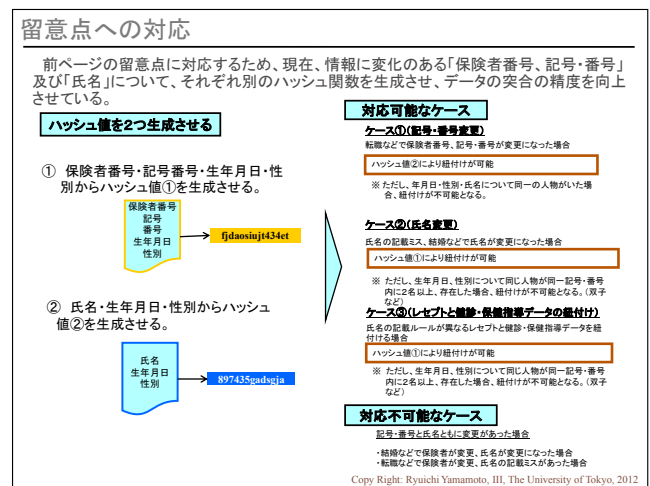


図2-3-10

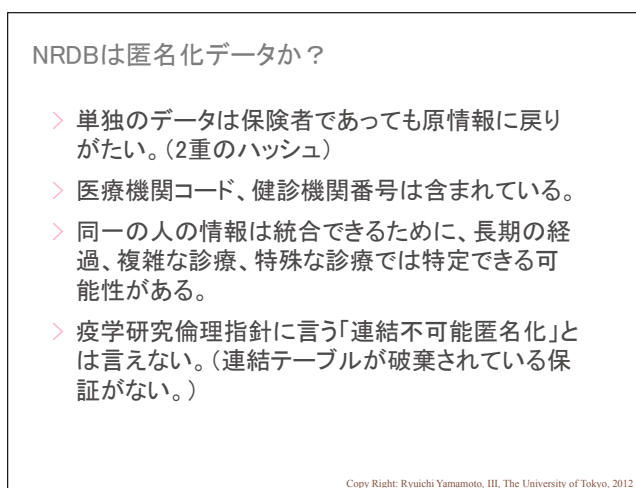


図2-3-11

	Korea	Taiwan	Japan
名前	HIRA DB	NHI Research DB	正式名称はない レセプトおよび特定健診・保健指導DB
収集範囲	全国レベルのレセプトデータを集積(ほぼ100%) (日本は特定健診・特定保健指導データを含む)		
他のデータソースとのリンク	不許可	患者の同意がある場合に限り部分的に許可	不許可
収集開始	2000年6月	1996年	2009年4月
対象人口	48,000,000	23,000,000	128,000,000
学術目的の利用	国のプロジェクトに従事している研究者に限られる。 さらにHIRA Claims Data Providing Review Committee)の審査を通過した場合。	目的と提供データ量が適切に判定されるが、基本的には学術利用は可	2014年度まで試行的に提供。 有識者会議の定めるガイドラインに適合した申請であることが審査される。

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-12

韓国 HIRA DB, 台湾 NHI Research DB, 日本 NRDB

	Korea	Taiwan	Japan
データ更新	毎月	毎年	毎月(レセプト) 毎年(特定健診・保健指導)
最新データ	前月	前年	2ヶ月前(レセプト)
誕生 年/月/日	Yes/yes/yes	Yes/yes/yes	Yes/yes/no
体重 / 身長	No/no	No/no	特定健診・保健指導および DPCデータには含まれる
調剤			
処方 月/日	Yes/yes	No/no	Yes/yes
調剤 月/日	Yes/yes	Yes/yes	Yes/yes
容量/規格 / 分量	Yes/yes/yes	Yes/yes/yes	Yes/yes/yes
ワクチン	No	Yes	保険適応ワクチンに限る

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-13

韓国 HIRA DB, 台湾 NHI Research DB, 日本 NRDB

	Korea	Taiwan	Japan
手術	独自コード	ICD-9 処置コード	独自コード
入院			
入院日	Yes	Yes	Yes
入院中投薬	Yes	Yes	Yes
退院時診断	Yes	No	Yes
診断コード	ICD-10	ICD-9	ICD-10
検査項目/結果	Yes/no	Yes/no	Yes/特定健診のみ

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-14

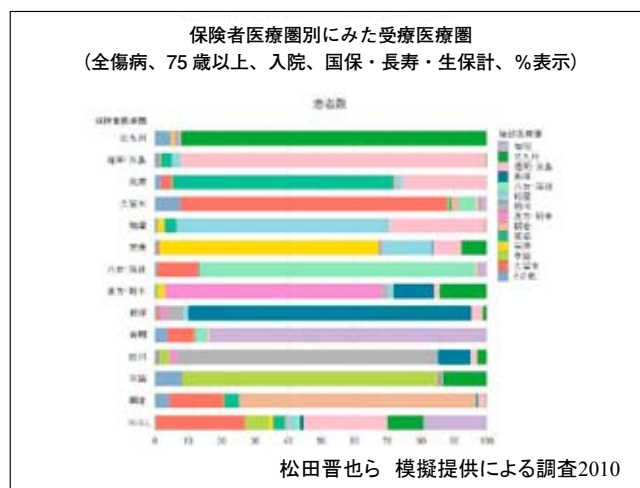


図2-3-15



図2-3-16



図2-3-17

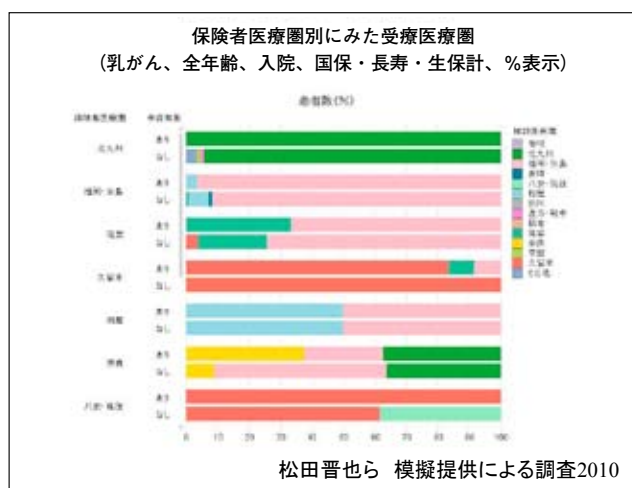


図2-3-18

使いやすいNRDBに向けての課題

第一回目の提供審査では43応募で6申請が許可

探索的研究は不許可にせざるを得ない。
セキュリティ上の不備は厳格に審査せざるを得ない。
そもそもレセプトの特性を理解していないものもあった。

- > 法的裏付けが明確ではない。(契約による安全性確保で罰則が難しく、条件を厳しくせざるを得ない。)
- > 探索的研究を可能とするためにはより安全なサブセットが必要ではないか。
- > データの持ち方がかなり特殊で、さらなる情報提供が必要ではないか。
- > データの持ち方はこれで良いか？

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-19

単純サンプリングデータセット

- > 以下入院、DPCレセ、医科入院外、調剤
- > 医科入院、DPCは10%、その他は1%抽出
- > 医科入院外は10月分11月分の調剤レセからハッシュで紐づけたものを同時に提供
- > 性別、5歳刻み年齢区分において母集団との構成比率を変えない
- > 傷病名コード、医科診療行為コード、診断群分類は出現頻度の低い順に総数が0.1%に達するまでダミー化
- > 高額レセ(入院70万点以上、その他5万点以上)は母集団から除外
- > DPCでは医療機関別係数、包括評価点、今月包括点数、合計点数は空欄に
- > 都道府県情報、医療機関情報、薬局情報、保険者情報、公費情報は削除
- > 病床数の記載がある場合はカテゴリー化
- > DPCでは入院時年齢、出生児体重、Burn Indexはカテゴリー化

目的別サンプリングデータセットは24・25年度厚労科研で。

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-20

今後に向けて

- > 制度整備
 - 契約による提供から制度に基づく提供へ
- > 26年度にはデータベースの更新予定
 - 今年度に取り方の検討を行う予定
 - 現状: 月々の入力と、本来目的の利用、調剤メディアス、社会医療診療実体調査などのための抽出で能力を使い切っている状態
 - 第三者提供には十分対応しきれていない。
 - Deidentificationは現状のままで良いか？
 - 稀少例を扱うためのPPDMの導入？
 - On Site Research Center(s)の必要性は？

Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-21

ありがとうございました。



Copy Right: Ryuichi Yamamoto, III, The University of Tokyo, 2012

図2-3-22

標準化・統合的活用の視点から

木村 通男 先生（浜松医科大学教授、附属病院医療情報部長）

- ▶ 病院情報システム（HIS）では、医用画像や患者基本情報、処方、検体検査、病名などの標準化は進んできている。特に医用画像のフォーマットの標準化は進んでいる（図2-4-2）。日本では病院のオーダーシステム（検査や処方の情報伝達システム）の普及が非常に高く（日本80%、米国25%）、それがHL7規格のデータであれば、ストレージのソフトは無料で提供している。2012年3月に117の施設がデータを蓄積している（図2-4-3）。
- ▶ 浜松医科大学で開発した「臨床検索エンジン」では薬剤の処方後の臨床検査値の変化を調べることができる。1人の患者さんのクレストールという薬物の投与前後のAST（代表的な肝機能の指標）検査値を例示した（図2-4-4）。
- ▶ HISのデータを解析する医薬品医療機器総合機構（PMDA）の医薬品の安全性に関する試行調査、MIHARIプロジェクトを2010年より5年間行っている（図2-4-5、6）。静岡県下の5つの協力病院のデータを用いて既知の副作用で調べたところ、血小板減少のような血液像の変化は見やすかったが、悪性症候群とか末梢神経障害は非常に出しにくいということが分かった。スタチンによる横紋筋融解（病名）、ファモチジンによる血球系減少（検体検査）の割合を例示する（図2-4-7）。参加病院が急性期病院であるため、血液像が悪いのはファモチジンの投薬だけが理由ではないと考えられる。
- ▶ さらにPMDAの指定する薬剤の検索プロトコールによって、国内10施設でオーダーエントリー情報から副作用を検索するセンチネルジャパンプロジェクトも2010年に始動した（図2-4-8）。しかし、解析結果の分析において、所見情報は含まれず、さらに、参加施設は急性期の病院が中心で、診療所等が含まれていないことに注意を要する（図2-4-9）。
- ▶ MIHARIプロジェクトやセンチネルジャパンプロジェクトで病院情報の解析における課題は、病名のいいかげんさ、処方テクニック（例えば1ヶ月分の処方を2週間に倍量の処方を出す）、検体検査の精度管理、前の病院情報の不在など。特に、前に通院していた病院の情報を調べるためには、共通IDが必要である。また、所見の構造化、例えば、血圧は数値データを扱う項目に入れる、といった構造化が必要と考えている（図2-4-10）。
- ▶ 臨床試験をEDC（臨床情報の、標準化されたHISからの電子的収集）で実施すると、所見情報は医師が書くので、紙ベースと比べてEDCでも基本部分の費用はそれほど安くないが、処方や検査の項目を1つ増やす追加額はEDCが安い。臨床試験の実施、ケースカード（調査票）の作成では、標準化によるこの費用対効果のことをぜひ考慮して欲しい（図2-4-11）。
- ▶ 臨床試験では患者さんの個人情報扱うため、大学病院の場合は医学・看護学教育及び症例に基づく研究ということで、これに関して匿名化をするように努めます、としている（図2-4-12）。
- ▶ 「あなたのカルテ情報が下記の用途に使われるのをどう思いますか？」というアンケートを一般の静岡県民で調査した。同じ質問をアメリカのRDDランダムダイアルでも調査した。厚労省、研究機関の独自研究、産学連携研究、保険会社、製薬企業が使う

場合、非匿名化が上段、匿名化が下段で、我が国では、匿名化であれば許容する人が大半で、アメリカでは匿名化されても嫌な人が多い、という結果であった（図2-4-13）。

- ▶ 「あなたのカルテを生涯カルテとして1つにまとめたいですか」ということで、メリットは重複が減ります、デメリットは一部分を隠すのがだめになります、として調査した。日本では、11.5%が嫌、7割ぐらいがグッドアイデアと考えるが、アメリカでは結構嫌な人が多かった。医療情報学会からは、医療情報の連携のための共通IDを診療の連携のためにも持つべきであると提言した。生涯カルテは患者さんのために進めるべき施策であり、二次的に研究のためになるということは理解される範囲ではあるが、オプトアウト、嫌がる人が不参加を選択できるように考えなければならない（図2-4-14）。
- ▶ 患者さんは目的を厳しく見ているので、研究者が思っているほど信用されていないということを、やはり我々は理解しなければならない（図2-4-15）。

（質疑応答ならびに討論）

- ▶ 静岡5病院データベースで、研究サイドから利用したいとリクエストがあるのは、データの入っている注射を含む処方情報と、検体検査結果と病名に対してDPC（急性期入院医療を対象とした診療報酬の包括評価制度）のコードとDPCの各種臨床付加情報を使って検索したい、というもので、検体検査結果で見えるものは、たいへんよく分かる。例としては、DPCの臨床コードを使って心不全の評価をする、という循環器の先生の検索希望があった。
- ▶ アンケート調査では、米国は不許容（not acceptable）とする人が多いが、オプトアウトは保険会社に関しては、保険加入の契約で許容されていない。その結果、不満はあっても、みんな医療情報データを使われている、ということが読める。日本はご存じのとおり、保険に関してどれに入っても扱いがそう変わるわけではないので、公平感があり、公益心がアメリカより強いと思われる。
- ▶ 医療情報の統合活用の目的からは少しそれるが、例えば倫理審査のあとの対応が日本ではよくない。本当にそれでやったかどうか、結果がどうだったか、ネガティブデータをちゃんと出すか。研究の側が患者の信頼にこたえるような透明性を確保することが必要と思う。病院情報システムに何を期待するか、との問いにNLM（米国国立医学図書館）のディレクターが「病院情報システムに期待するのは、要するにチーティングに対するエビデンスだ。」と言っていたが、病院情報システムは不誠実な医療への証拠物件となり得る機能がある。
- ▶ 患者さんが一番懸念するのは、自分の医療データが病院情報システムに登録されて自分にとって不利になるのではないかとということで、だからこそ、たとえ自分のデータが使われても、それが不利になるわけではなく、きちんと匿名化されて個人情報を守られていることを、わかりやすい形で示すことは重要だと思う。さらに、副作用情報とか、病気の発見につながるとか、そういったことをきちんと探し出して、研究成果として周知していくことが重要であろう。
- ▶ よく「被験者は研究のメリットを受けない」と言われる。でも、今の時代、慢性疾患などでは、そもそも薬が効いているのか効いていないのかわからないから研究するの

で、その結果は被験者に還元している。少し前の時代は、被験者に還元されないからそういう研究は抑制すべきだという意見が結構多かったように思うが、少し時代が変わってきた。それは慢性疾患が増えているということだと思う。

- ▶ 研究の透明性は、患者だけでなく、社会を対象にして持つべきである。
- ▶ 厚労省のデータベースのDPCによる分析で病院の特色が分かる。HISでも病院ごとの特徴が出る。データ統合時の病院による違いは考慮しなければならない。また、HISで連携することに対するインセンティブが要ると思っている。
- ▶ 個人情報保護に関して、違反は刑事罰の対象になるという部分は診療情報を扱う上で非常にネックになる、ということについて意見を述べる。刑事罰がいいとか悪いとかではなく、ここのところは少し考え方を変える必要がある。個人情報保護法の中には非常に不確定な概念が使われている。例えば「容易に」であるとか「特定」「匿名」というものも、どこまですれば匿名なのかという問題は常に出る。そこで、努力をしたものについては、違法性阻却をさせる手続があると思う。つまり「匿名の手続をとるためにこういう努力をした」ということがある場合には刑事罰から外すという発想をとってもいいのではないか。努力をしない者は刑事罰の対象になる、けれども、努力をしたところが不十分であったとしても、努力をしたところを評価する、といった法の組み込みは可能なのではないかと思う。違法性を阻却するような事情があれば法律違反にならない、つまり違法とは評価されないというのが刑法の解釈なので、法改正するのではなく、解釈で組み込む、という議論をしていくことができる。
- ▶ 法律化するとなれば、多分その努力の仕方を明記することになると思う。例えば、医療情報であれば厚生労働省令に基づいて「このようなことをすれば努力したものと評価する」といったものを定式化していく。多分、時代によってその努力の度合いが変わってくると思う。それを明記することになるのではないか。例えば、アメリカのHIPAAのように18項目を抜けば、それでもデータベースとしては匿名化したと絶対言えるかという、言えない。突合できる可能性がある。けれども、18項目を抜くといった努力をしたことを評価すると明記することで、違法性阻却を定式化できる。

・病院情報システムのデータの活用例 ・カルテ情報2次利用についての市民アンケート

日本医療情報学会理事長
浜松医科大学医療情報部
木村通男

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-1

病院情報システムが持つデータ (上ほど入手しやすい)

- ※画像 (DICOM規格)
- ※患者基本、処方、検体検査結果 (HL7規格)
- ※病名(但しどれが主病かわからない) (HL7)
- ※各種報告書、計画書(人に読んでもらうための書類) (HL7の文書規格CDA)
- ※所見、計画など(プログ्रेसノート) (規格なし)

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-2

標準機能としてHL7 v2.5でデータを出せるHIS稼働数、SS-MIXストレージ稼働数(2012/3現在)

富士通: FX(定期バージョンアップで対応), GX	:398	18
NEC: 電子カルテ: HR, オーダ: AD v.4 以降	:253	11
SBS: Doctor-X, PrimeKARTE	: 23	72
ソフトウェアサービス: e-Karte	:309	16
合計(その他含む)	985	117

※世界的に非常に高い、大病院でのオーダシステム普及率: 日本(80%)、アメリカ(25%)



Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-3

・検索モデル(画面例)



投与前1週間以内の検査では、ASTが30~180だったがクレタール投与後1週間以内の検査ではASTが180~500になった症例の検索。
※上のデータはデモ用のものであり、患者氏名・IDなどはずらしているものである。

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-4

PMDAのMIHARIプロジェクト SS-MIX 標準化ストレージデータを利用した 医薬品の安全性に関する試行調査

- ※電子診療情報等の安全対策への活用に関する検討会(2010より5年)
 - ☑PMDA 安全第1部 調査分析課
- ※22,23年度は静岡5病院対象
 - ☑D*D稼働中
 - ☑静岡県総合、静岡済生会、袋井市民、沼津市立、浜松医大

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-5

PMDA MIHARI year 1

【医薬品曝露群に関する調査】

- 候補テーマ①: オランザピンによる悪性症候群
- 候補テーマ②: スタチン系薬剤による横紋筋融解症
- 候補テーマ③: オランザピンによる高血糖
- 候補テーマ④: ファモチジンによる血小板減少または汎血球減少・無顆粒球症
- 候補テーマ⑤: フロセミドによる血小板減少または汎血球減少・無顆粒球症
- 候補テーマ⑥: アムロジピンベシル酸による血小板減少
- 候補テーマ⑦: スタチン系薬剤による末梢神経障害

【医薬品曝露群と非曝露群の比較】

- 候補テーマ⑧: フロセミドによる血小板減少または汎血球減少・無顆粒球症

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-6

結果1, 2

スタチン・横紋筋融解	5施設合計	浜松医大
実対象者(薬剤投与患者)	7552	1683
ケース	178	20
発生割合	2.36%	1.19%
ファモジチン・血球系減少	5施設合計	浜松医大
実対象者(薬剤投与患者)	17960	3604
ケース	684	162
発生割合	3.81%	4.50%

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-7



図2-4-8

厚生労働省医薬食品局: 医薬品等安全対策のための医療情報データベース基盤整備事業
(センチネルジャパンプロジェクト)

- ※平成23年: 基本仕様設計、1か所先行試験導入(東大病院)
- ※平成25年目途: 大規模拠点整備(10か所)
 - ☑東北大、千葉大、浜松医大、香川大、九州大、佐賀大、北里大グループ、NTT病院グループ、徳洲会病院グループ
 - ☑PMDAによる利用
- ※平成27年目途: 1000万人規模への拡大。

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-9

病院情報を研究素材にする際の課題点

- ※病名のいいかげんさ、処方テクニック、検査そのものの精度管理
- ※施設間にまたがる
 - ☑「初回処方」は当院での話であって、生活習慣病薬は、紹介元ですでに処方されている?
 - ☑マイ・ナンバーと異なる、医療IDの提案
- ※所見をいかにデータベース取り込み可能とするか。

図2-4-10

試算したケースカード単価と1項目増分の費用

	10処方・検査+10所見	10処方・検査+50所見	50処方・検査+10所見	1項目あたり増分
	5年2400枚	5年12000枚	5年2400枚	所見処方・検査
紙ベース	6,334	6,334	23,667	433
EDCによる	11,234	5,733	24,567	333
非標準HIS	11,125	4,892	24,458	333
HL7で出せるHIS	8,833	4,433	22,167	333
SS-MIXで出しているHIS	5,833	3,833	19,167	333

- ※大規模施設では特に、HIS系が紙にまざる
 - ☑SS-MIX標準ストレージは他用途にも使える
- ※EDCはスケールメリットがあり、その点は各HISと同じ
- ※1項目増分は、紙ベースだと医師の所見(433円)も処方など(200円)も同程度だが、HISを用いると大きく異なる
- ※非標準HISからHL7で出せるようにするには1~2千万、HL7で出せるHISがSS-MIXに出せるようにするには3~5百万(コードの標準化込み)

Michio Kimura M.D. Ph.D. Hamamatsu University School of Medicine

図2-4-11

院内掲示「患者さんの個人情報の扱いについて」
患者さんの個人情報は、各種法令に基づいた学内規定を守った上で、下記の目的に利用されます。

- ※(1)当病院での利用
 - ☑患者さんがお受けになる医療サービス
 - ☑医療保険事務
 - ☑患者さんに関係する管理運営業務(入退院等の病棟管理、会計・経理、医療事故の報告、医療サービスの向上)
 - ☑医療サービスや業務の維持・改善のための基礎資料
 - ※(2)当病院及び浜松医科大学の利用
 - ☑医学・看護学教育
 - ☑症例に基づく研究
 - ☑外部監査機関への情報提供
 - ☑職員の業務改善
- この利用に当たりましては、匿名化するよう努めます。

図2-4-12

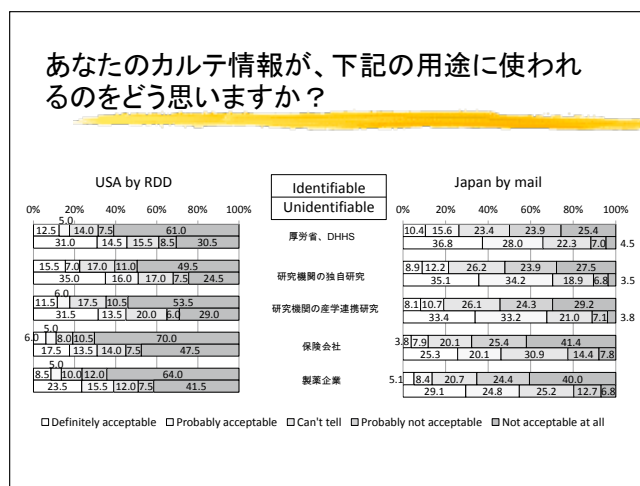


図2-4-13

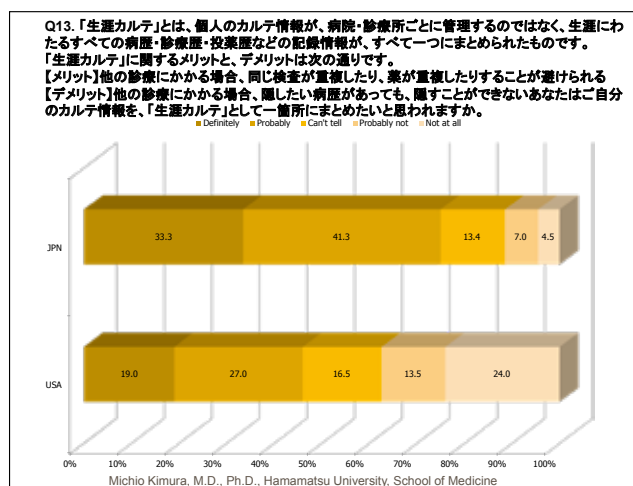


図2-4-14

Final remarks

- ※ 病院情報システムから、処方、注射、検査結果が出る基盤が揃ってきた
- ※ PMDAのMIHARIプロジェクト、医薬食品局のセンチネルプロジェクトはこの基盤を整備し、シグナル検出を目指す
- ※ 施設をまたがった患者のデータ突合、所見のデータベース化はこれからの課題
- ※ 患者は「目的」を厳しく見ている：研究者は自分が思っているほど信頼されていない

Queen's head rock, Taiwan

図2-4-15

セッション1 「データベースの活用目的」 討論

データベース活用目的に関する論点を図2Aにまとめた。病院のデータ、保険者のデータ、健診機関のデータ、研究機関のデータが提供され、研究者、研究機関、国、自治体、国民等々が活用する。「どのデータベースをどういう形で使って、誰にインパクトを与えるのか」の議論が希薄なまま、我が国のデータベース戦略が進んできたと考える。ここでは、その点を具体的に議論した。主な意見を以下に列挙する。

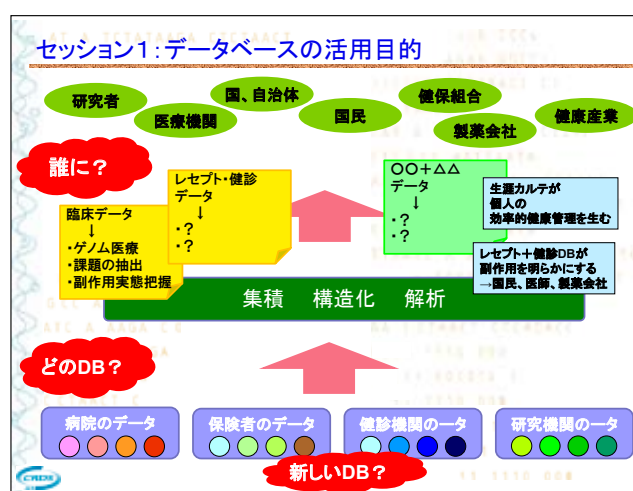


図2A セッション1討論資料

- ▶ 社会からは、研究者が研究のために臨床研究をしていると見られている。研究成果が国民にフィードバックされることを明示すべきだ。特に慢性疾患と長期間にわたり付き合う時代では、個人の生涯診療記録は個人にとって有意義であると同時に、治療の質を精査するためにも必要なものである。研究者はそのような情報活用の仲立ちをするのであって、自分の業績のためだけに研究をしているのではないことを明確にする必要がある。また、日本人の健康を守るためには日本のデータが必要である。欧米人で言われたことがすべて日本人に当てはまる訳ではない。
- ▶ 人種差はトピックスであり、コーケジアン、それから東アジアと南アジア、あとはアフリカ系の黒人の大きくは4つに分けられて議論されている。ただ、東アジアでも、細かく考えれば、日本人と韓国人、中国人は異なる。国々で、食習慣、気候、そして遺伝形質にばらつきがある。特に有名なのはヒト白血球抗原（Human Leukocyte Antigen、HLA）領域の多様性である。環境や遺伝の風土的要因が、疾患予測や診断に与える影響は無視できない話であり、日本人のテーラーメイド医療を作るためには、日本人のデータを備えて、それをリファレンスして医療にあたるのが正攻法である。
- ▶ この分野の動向が早い中で、我が国の対応は遅れている。GWAS解析の結果と個人ゲノム解析を組み合わせ、医療に向かう動きがある。シーケンシング技術を診断に使用することについて、社会インフラとしての整備、個人ゲノム解析の倫理対応など、先行しているアメリカ等の状況にはとても追いつかない。中国では倫理が問題化しにくく、集中的に資金も投入できるので、個人ゲノム領域では、“エイヤ”で東アジアのトップランナーになる可能性が高い。我が国が、どこまでやるのかあるいは、どこ

- までやれるのかは、国として明確な方針を出す必要がある。
- ▶ 少なくとも米国NIHが次に目指しているのは、医療情報の1コンポーネントとして個人の全ゲノム情報を使うことである。電子カルテ情報にゲノム情報を組み込む方向に進んでいるが、それがどこまで有効かについては議論が必要である。しかしながら、ゲノム情報は少なくとも診断ツールとして活用できるとの認識が専門家の中では広がりつつある。
 - ▶ ゲノム情報と突合できるあるいは、突合可能な我が国の医療情報データベースとしては、大規模なものではレセプトデータベースがある。ただし、レセプトは確かに網羅的に収集されているが、情報の中身はあくまでも医療機関が診療報酬を請求するための情報であり、医学的にも学術的にもそれほど正確とは言えない。ただ、どういう薬が使われたかについては確実に把握できるので、「こういう治療が行われた人」というレベルではあるが網羅的データベースとして価値がある。病院のDPCデータベースも学術研究用途に提供することになっており、提供方法を有識者会議で議論中である。DPCにはカルテ情報が入ってくるので、提供側からかなり強い抵抗があり、少し時間がかかると思うが、最終的にはデータが提供されるようになるだろう。他に利用可能なデータベースとしては、特定健診、特定保健指導のデータベースがある。これは網羅的ではなく、2,700万人の受診者のみであるが、血液検査の結果なども含まれている。検査結果もある程度コントロールされているので、一般の病院の検査データよりは精度が高いデータベースになるかと考える。
 - ▶ 染色体上の変異を比較してみると、日本人と北ヨーロッパ系の人との間で明らかな差異がある。この事実から考えて、やはり日本人のデータベースをつくっていかなければならない。ゲノムデータをきちんと整備した後、突き合わせるデータは、クオリティコントロールされているものでないと意味を成さないだろう。
 - ▶ 企業の立場から見ると、先生方の出された成果をどう利用するかという視点になる。その意味では、エビデンスや知見の移転に関する法整備が必要であろう。健康情報をビジネスとして活用する場面では、「医療費の削減」が前面に出やすい。しかしながら、削減にも限界がある。情報提供ビジネスを継続性のある産業とするためには、健康増進のための情報提供など、継続性のあるビジネスが必要だ。そこで、社会インフラを回しながらデータをためていって、それを研究に使っていただくという仕組みが理想的である。

セッション2：「統合的 DB 活用に向けた技術課題と制度課題」

企業の視点からの課題

佐藤 恵一 先生（日立ソリューションズHB 事業推進部グループマネージャ）

- ▶ 企業の立場でできることを紹介する。疾病よりも未病の視点である。
- ▶ ライフイノベーション分野では、がん等の社会的に重要な疾患の予防や改善及び治癒率の向上などが挙げられている中で、予防医療等の重要性が科学技術アクションプランで指摘されている。また、長期の追跡で継続的に得られる情報を統合していかないと健康長寿モデルの構築が難しいということも、日本学術会議等で行われている（図2-5-1）。
- ▶ 健康長寿社会の中での社会インフラというキーワードは、医療機関だけでなく健診機関、健康事業者（健康を意識した消費行動に対するサービスを提供する）といった三つ巴の取り組みが必要である。医療機関は、医療技術で我々の健康に大きく寄与しており、健診機関も、未病の中で疾病期に移るのをモニターするという観点で重要な役割を持っている。一方、経済的に重要な健康事業者にも役割を持たせなくては行けないが、今の日本では状況がまだ醸成されていない（図2-5-2）。
- ▶ 健康事業者醸成の1例として、利用者の健康管理情報を匿名化してITインフラ基盤のDBに預けることが考えられる。その匿名化した情報を研究機関で使えるようにすれば、いろいろな研究成果が上がる。健康事業者が研究成果に基づいたサービスを提供して、利用者にインセンティブを与える。こういうサイクリックな仕組みができると、継続可能なモデルができる（図2-5-3）。
- ▶ 事業を検討し進めている中で見出している1つ目の課題は、法規制に関する検討である。事業者が安心して事業を推進できるようにしないといけない。2つ目はビジネスフレームワークの構築で、健康事業者の役割も明確にしないとフィージビリティを担保した世界はできない。3つ目は、匿名化技術の検討である。漏洩時のリスク低減と個人の特定をどのように避けて利用者に安心感を与えるかを考えなければ行けない（図2-5-4）。
- ▶ 法規制面での検討は、医師法や薬事法などでグレーゾーンが多く、個人情報保護法の中で企業が自信を持てないような形になっている。企業が安心して取り組める法整備、ガイドラインの整備が必要である。
- ▶ ビジネスフレームワークの構築は、健康管理情報がITインフラ基盤の上に置かれると、まずは健康管理情報だけを利用したサービスが基点になる。その上で各企業のエビデンスを蓄積する、あるいはサービスの付加価値を上げていく。血液検査や遺伝子検査、バイオマーカーなどの検査が入ってくるだろう。情報がたまとサービスコンテンツもどんどん発展し、より付加価値の高いサービスが出てくる。このように段階的に進展していくのが一番予測しやすい形だと考える（図2-5-5）。
- ▶ 最初に種になる情報が入ってこないとサービスも回らない。まずは、種になる健康管理情報を基に様々なプレーヤーがサービスを提供できる基盤を用意し、その中で法制的な基盤というものを担保した上でサービスを継続していく必要がある。

- ▶ この「サービス継続サイクル」が回ってくると、ITインフラ基盤の中にデータがたまり、そのデータを利用することによってコンテンツを開発できるようになる。また、研究機関の役割が大きくなる（図2-5-5）。サービス継続サイクルと新産業創設サイクルをうまく回していきながら、社会全体に知見というものをフィードバックしていく仕組みをつくる必要がある。補助金頼りではない形を目指すべきである。
- ▶ 事業者としては、匿名化技術において機微情報を持ちたくない。そのため、一次匿名、二次匿名と段階を踏んで匿名化を強固にするのではなく、最初の段階で個人情報から機微情報を分離する必要がある（図2-5-6）。
- ▶ 機微情報を切り離した個人情報は、インフォームドコンセントと一緒に企業がしっかり管理をして、それに基づいた情報の権限をITインフラ基盤の中で統制する。機微情報はITインフラ基盤の中に閉じ込める。さらに、暗号化技術を組み合わせれば、よりセキュリティレベルを向上することができる。
- ▶ 事業の継続性という観点でクラウド基盤の活用は有効的である。いきなり大きい基盤を作るのではなく、ある程度規模の伸縮可能な、スケールアップ、スケールアウトが可能なクラウド基盤を利用していく必要がある。
- ▶ このような仕組みをとることで、継続したデータ集積が可能になり、さまざまな研究や調査に対する拡張性とデータの連続性が生まれる。
- ▶ 実際にプラットフォーム化を考えると、さまざまなアプリケーションへの対応やさまざまな企業が参加できる状態をつくらなくてはいけない。その意味ではデータベース上で標準化し、共通API上にアプリを自由につくれる、審査をした上で許可を与えていくといった形が必要になる。
- ▶ 行政や医療機関とのシームレスな連携ができることも必要。
- ▶ 健診機関、健康事業者、医療機関がシームレスな情報の連携をすることで、健康長寿社会が実現できると考えている。副次的には、重複計算や医療費の削減とか、予防医療という産業が新たにできることで経済も心身の健康も発展していく。
- ▶ こういった技術を匿名バンクという形であらゆるところに利用している。実際にさまざまな検査結果を使ったサービスにも取り組んでおり、サービスを提供する上での課題を日々検討している（図2-5-7、8）。
- ▶ 例えば人事情報と健康管理情報が結びついてしまうと社内で人事査定等の不公平を生む可能性もあるため、社内にはデータを残さないという形で、人事情報と健康管理情報を分けて、保健師と産業医だけが結びつけられるシステムをサービスとして提供できる（図2-5-9）。

（質疑応答ならびに討論）

- ▶ 図2-5-6の匿名化手法の中で、個人情報を企業に渡すのは産業を起こすために必要か。
- ▶ 匿名化された個人情報が必要である。企業は匿名化されていないと何かがあったときのリスクを負い切れない。損害レベルが高くなり、賠償金にするとかなり損害度の大きいものになる。
- ▶ 図2-5-6でインフォームドコンセントの取得は当然医療機関が行うと思うがそのつなぎ合わせはどうするのか。
- ▶ 現状のサービスでは、産業レベルでインフォームドコンセント取得を行っている。企

業の中で閉じている。

- ▶ その場合、何が機微情報で何が個人情報なのか微妙な問題があるのではないかな。
- ▶ 経済産業省から、ガイドラインが出ているため、それに従っている。独自につくことはできないので、あくまでそういったガイドラインに従って、識別できる、できないということを考えている。ただ、後追いして結びつけられるかどうかに関しては、担保できていない部分もある。ITインフラ基盤に関しては、さらに対策が必要である。
- ▶ そういうサービスがよかったというためには、やはり重篤な病気、合併症や脳卒中などを防ぐことが最終的なエンドポイントになる。そのためには病院の情報や、病名などそういう長期にわたって確実に障害が起こったという情報が必要になってくるが、そこはどのように仕組まれるのか。つまり、検査値がよくなるだけだと、本当にそれがいいことかはわからない。そこが今、臨床で問題になっていて、例えば、糖尿病で教育入院すれば血糖もよくなる。けれども、長期的には何もよくなっていない。それは単なる見せかけの改善であって、やはり最終的には入院したかとか生きているかとか、そういうハード・エンド・ポイントをとらえないといけない。
- ▶ 疾病期に移行したときに人の命がどう延びるかとか、完治するかということに関してこのインフラが適用・活用されるのは、恐らく病院との連携ができてから。ただ、健康情報、未病の情報というのは民間の取組みでないと難しい。民間から医療へのデータのやりとりは壁が厚いが、未病から疾病期に移ったときの情報がうまくマイニングできると、疾病期に移る期間を遅らせるという観点で、機能してくるのではないかな。
- ▶ 医療IDがあれば病院情報と健診情報がつながる。その取り組みが非常に重要だ。
- ▶ 医療IDで一元的に管理してしまうと、それがなくなったときのリスク等を考えていかななくてはならないし、各企業、各病院がそれぞれのID体系で運用している面もある。そこをうまく連結できる仕組みを取り入れて管理していくべきである。

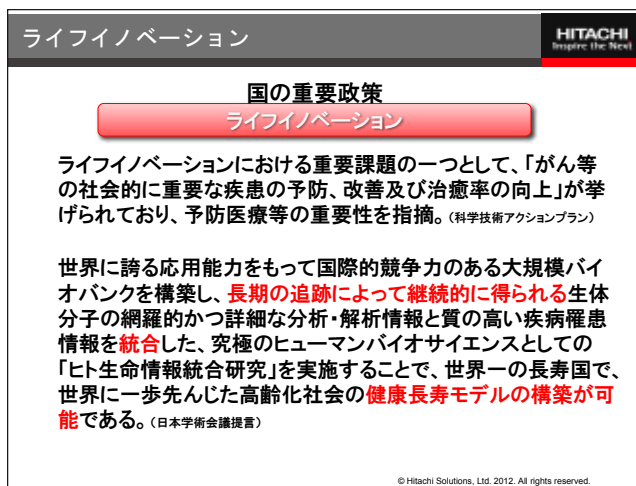


図2-5-1



図2-5-2

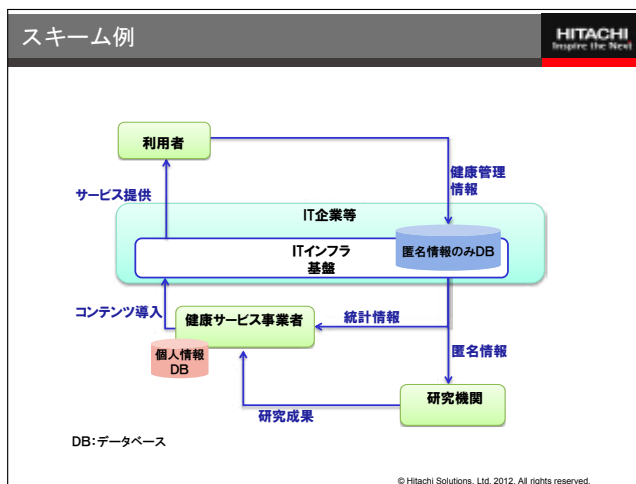


図2-5-3

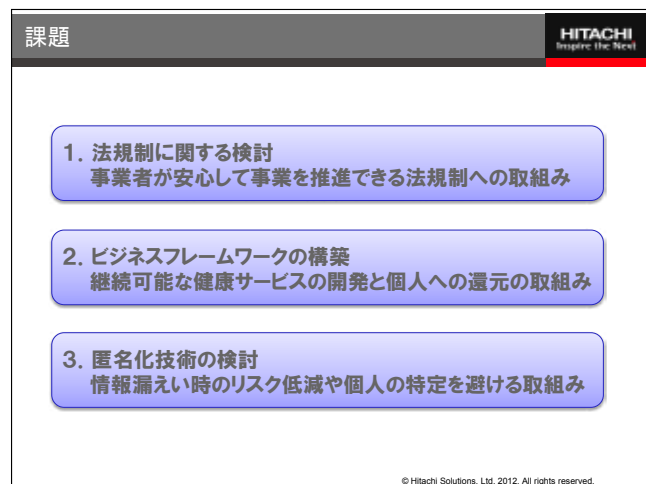


図2-5-4

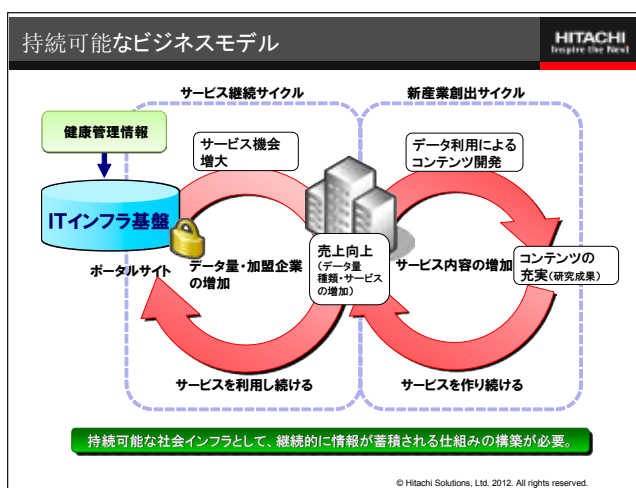


図2-5-5

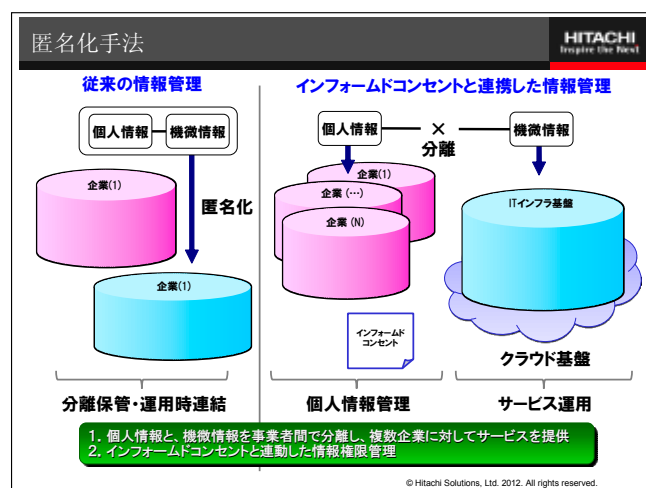


図2-5-6

```

graph LR
    User[利用者] -- 申し込み --> Provider[サービス事業者]
    Provider -- 検体送付 --> CheckCo[検査会社]
    CheckCo -- 検査結果①, ②, ③ --> Provider
    Provider -- 情報 --> Bank[(匿名バンク)]
    Bank -- 検査結果 --> Provider
    User -.->|お勤めの〇〇はこれです!!| Provider
  
```

The diagram illustrates the architecture of Hitachi's Employee Health Management Cloud Service. It shows the flow of data from an employee's smartphone to a cloud server, then to a company's management center, and finally to a management officer. The process involves secure data transfer via a VPN and secure storage in the cloud.

1. 業務現場にシステムが普及すれば、情報データのやり取りが迅速になり、業務の効率化が図れます。

2. クラウドサービスを利用することで、ネットワークコストを削減できます。

3. 個人が特定できない情報のみをクラウドサービスにて管理することで、情報管理リスクを軽減できます。

企業・経営管理センター

作業室変入！

従業員、医師、作業室変入、作業室変出、医師・作業室変入、医師・作業室変出

クラウドデジタルソリューション

セキュアなデータ管理

VPN

人事情報はネットワーク上に渡りません。

クラウドサービス

人事・経営部門

人事DB（人事情報）

自動送信

伝票

© Hitachi Solutions, Ltd. 2012. All rights reserved.

独立行政法人科学技術振興機構 研究開発戦略センター

個人情報保護法を中心とした制度課題

村上 康二郎 先生（東京工科大学教養学環准教授）

- ▶ 法的あるいは個人情報保護の話を中心に発表する。
- ▶ 医療や健康分野を検討する前提として、プライバシーとか個人情報保護制度をきちんと理解する必要がある。また、現在厚生労働省を中心に医療等情報個別法の検討が進んでおり、注目する必要がある。医療分野では匿名化が重要（図2-6-1）。
- ▶ プライバシー権は、「私生活を公開されない権利」という消極的な権利だと考えられてきたが、近年の情報化の進展等を踏まえ、自己に関する情報をコントロールする権利として認めるべきという学説が有力となっている（図2-6-2）。
- ▶ その対象となる情報も広がっており、権利としても強力になってきている（図2-6-3）。
- ▶ 日本では2003年に個人情報保護法制の整備がなされ、幾つか法律ができた。民間部門にはいわゆる個人情報保護法、国の行政機関には行政機関個人情報保護法、独立行政法人には独立行政法人個人情報保護法が適用され、それぞれ法律の内容が微妙に違っている。その結果、私立病院には個人情報保護法、国立病院には行政機関個人情報保護法、国立大学法人の病院には独立行政法人個人情報保護法が適用される。しかも、県立病院等には地方公共団体の各条例が適用される。この条例が1,800程度あると言われており、しかもその中身が微妙に違っている。そうすると、複数の病院の情報を連携して使いたい場合に、適用される法律が微妙に異なることで、障害が発生する（図2-6-4）。
- ▶ 個人情報の定義は、特定の個人を識別することができる情報。また、その情報だけでは特定個人が識別できなくても、他の情報と容易に照合することで識別できれば、それも個人情報となる。特に医療系では機微情報が問題となる。海外ではセンシティブデータと呼ばれ、人種・民族や病歴など、健康や身体に関する情報は氏名や住所よりも重要な情報であると一般的に考えられている。日本の個人情報保護法には機微情報について何も規定がないため、医療分野の特別法が必要ではないか（図2-6-5）。
- ▶ 日本の個人情報保護法は、個人情報取扱事業者に様々な義務を課している。本日のテーマと特に関連が深いのが第三者提供。原則として、本人の同意を得ないと事業者は第三者に提供してはいけないため、データベースを作成したり連携したりする時の障害になる可能性がある。また、例えば治療目的で収集した情報を後で研究目的で使うなど、目的外利用も問題（図2-6-6）。
- ▶ 現在、マイナンバー法案が審議されようとしている（図2-6-7、8）。
- ▶ いわゆる通常のマイナンバーである社会保障・税に関する共通番号とは別に、医療等に関するIDを作る動きがある。厚生労働省の検討会で今年9月に報告書が出ているが、まだアバウトな内容で、直ちに法案になるとは考えにくい（図2-6-9）。
- ▶ 論点は多岐に亘るが、本日のテーマとの関連が深いのは、本人同意の問題。第三者提供を行う場合は本人同意が必要だが、医学研究や公衆衛生の向上のためには、少し緩和してよいのではないかという議論がある。医療の分野では、本人が意識不明だったり死亡していたりすることもあり、同意を得るのが難しい場合もあるため、その辺の対応が課題になる（図2-6-10）。

- ▶ もう1つ、匿名化も重要な論点である。匿名化した後、本人同意を行わずに情報を適切に活用できる仕組みが必要。厚生労働省の検討会では安全に匿名化すべきと言われるが、それが具体的にどういう状態かは不明確で、今後詰める必要がある（図2-6-11）。
- ▶ あくまで現時点の暫定的な私案だが、全てにおいて本人同意を要求したら成り立たないところもあるため、学術研究や公衆衛生の向上など、国民全体の利益につながるものについては、要件を少し緩和しても良いのではと考えている（図2-6-12）。
- ▶ ここからは匿名化の具体的な話題提供を行う。日本及び海外において、一定の匿名化がなされた場合には規制対象から外れるという議論があった。例えばEUもそうだし、アメリカのHIPAA法では、（1）有識者等がお墨付きを与えている場合、或いは（2）個人に関する18の識別子を削除して匿名化した場合は、法の適用対象外となる。ただ、本当に18の識別子で良いのか、という批判はある（図2-6-13～16）。
日本では、対応表の有無で連結可能、連結不可能と言われており、国内の医学系の倫理指針やガイドラインでもそのように表現される。しかし、海外では「匿名化」は連結不可能匿名化だけを指すため、連結可能匿名化は「仮名化」と呼ぶ方がよいのではない（図2-6-17）。
- ▶ 匿名化の手法は様々だが、中でもk-匿名化が国内外で注目されている。単純に氏名などを振り落とすのではなく、そのレコードの持っている情報が最低k個以上存在するように集合的に匿名化し、特定の個人に絞り込めないようにする手法である（図2-6-18）。
- ▶ 現行の個人情報保護法を前提にした場合、匿名化によってその情報を非個人情報にできるのかを、経済産業省のプロジェクト等でかなり検討してきた（図2-6-19）。
- ▶ まず問題になるのが、「容易照合性」への対応である（図2-6-20）。
- ▶ 次に、本人同意なしに第三者提供をできるかについてもグレーな状況（図2-6-21）。
- ▶ アメリカのポール・オーム氏の最新の学説が非常に注目されている。匿名化をしても他方で再特定化技術、リアイデンティフィケーション技術が非常に進歩しているので、リスクは有るとの主張である。さらに、HIPAA法もかなり批判しており、結局これはリスク評価の問題ではないかと指摘している（図2-6-22）。
- ▶ オームを全面的に支持するわけではないが、考慮すべき課題だろう。従来、一定の匿名化をすれば個人情報ではないとの方向性はあったが、単純にそうだと言えなくなってきた。つまり、リスクがどの程度高いか、低いか、の問題だと考えている。従って、何らかの第三者機関がリスク評価を行い、その匿名情報の利用に関するお墨付きを与えるような仕組みが必要だろう（図2-6-23）。
- ▶ 個人的な見解だが、リスクの評価基準について。従来は、匿名化技術ばかり注目されてきた。しかしそれだけでは不十分で、誰にどういう範囲で公開するのか、その情報の種類は何か、目的は学術研究なのか営利なのか、などを考慮した総合的な判断が必要だろう（図2-6-24）。

（質疑応答ならびに討論）

- ▶ 解析研究の情報セキュリティを審査する場合、例えば、各事業者や医療機関内の倫理委員会で判断する方法もあるだろう。ただ、公正さや厳密性の面での不安はある。匿

名化しても、種々の情報を突き合わせると再特定化されるリスクが残る点が難しい。従って、ケース・バイ・ケースの判断をせざるを得ない部分があるため、やはりある程度信頼できる中立的な機関が必要なのかなと思っている。いずれにせよ、この辺については国家的な議論が必要。

- ▶ マイナンバー制とは別に医療のIDを作ろうという点について、そのことで根本的な壁が出来てしまう恐れがあるのではないか。
- ▶ 個人的には、いわゆるマイナンバーとは別に医療分野だけの医療等IDを作った方が良く考えている。マイナンバーとは利用目的、リスク等がそもそも異なっている。根本的な壁という観点からは、これまでの個人情報保護法制は、行政や民間などで法律が異なっていたが、医療分野の法律を一本化していった方が良い。
- ▶ マイナンバーと医療等のIDは別に作られる。これは医療側、サービス側から見のではなく、国民側から見たときに、税を払うための番号と検診など医療に関わる番号を、恐らく多くの方は一緒にしたくないであろうということで、それは別に整備すべきであろうということになっている。それから、学術利用に関しては憲法で保障された学問の自由なので、そのこと自体に影響を及ぼすような法制度にはできない。ただし、公益利用、学術利用に関しては善意の同意をもう少し尊重したいということがあって、曖昧な目的による同意、つまり公益あるいは医学研究に用いるということで同意をいただけたら、もう同意を得たと見なすという考え方が一方にはある。ただし、それで本当に良いのか、この研究だけには使ってほしくないという人が出るかもしれない。そこで、オプトアウトの権利を保障しようと思うと、ご本人が自分の情報がどこにどうあるのかを知らないといけない。その場合には医療等IDを使う以外に知る方法がない。医療等IDとつながった医療・健康情報というのはマイナンバー法で言う特定個人情報に当たるので、番号が付与された個人情報で、これは現行の規制よりも厳しい規制がかかる可能性があるという点が現在の議論で、まだ結論が出ていない。

I はじめに

- 医療分野における個人情報保護を検討する前提として、**プライバシー権**や**個人情報保護制度**の概要を理解する必要がある
- 医療分野においては、現在、厚生労働省を中心に**医療等情報個別法**の検討が進められている
厚生労働省「社会保障サブワーキンググループ及び医療機関等における個人情報保護のあり方に関する検討会の合同開催」など
- 医療分野においては、**匿名化**が重要であるが、最近の議論動向が注目される

図2-6-1

II プライバシー権について

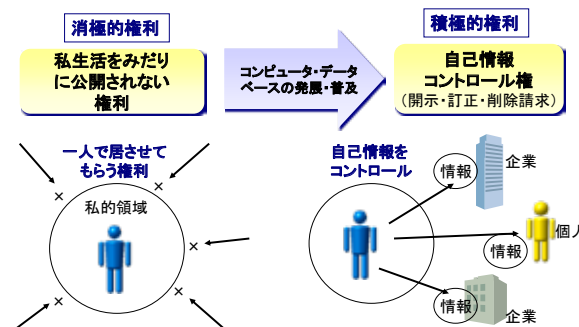


図2-6-2

伝統的プライバシー権と現代的プライバシー権

	伝統的プライバシー権	現代的プライバシー権
定義	私生活をみだりに公開されない権利	自己に関する情報をコントロールする権利
対象情報	私生活上の事実・一般人が公開を欲しない・非公知	①原則として 固有情報 ② 個人情報一般 が対象になる
侵害行為	公開する行為	取得・収集、保有・保管・利用、開示・提供
法的性質	自由権的権利	自由権的側面 請求権的側面（開示・訂正・削除請求）

図2-6-3

III 個人情報保護制度の概要

個人情報保護法制の整備（2003年5月）

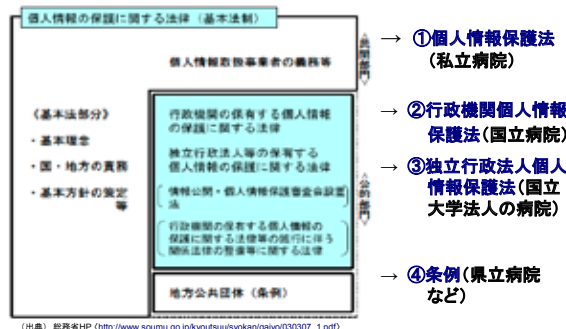


図2-6-4

個人情報保護法における個人情報概念

個人情報とは何か？

・個人情報の定義（個人情報保護法2条1号）

「生存する個人に関する情報であつて、当該情報に含まれる氏名、生年月日その他の記述等により**特定の個人を識別することができるもの**（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む）」

個人情報に該当するための要件

- ①**生存性**・・・生存する個人に関する情報
- ②**個人識別性**・・・特定の個人を識別することができるもの
- ③**容易照合性**・・・他の情報と容易に照合することによって特定の個人を識別することができる

→ **機微情報（センシティブ情報）**についての定めはない

図2-6-5

個人情報取扱事業とその義務

個人情報取扱事業者

個人情報データベース等を事業の用に供している者（2条3号）

	個人情報取扱事業者の義務
15条	利用目的を特定しなければならない
16条	特定された目的に必要な範囲を超えて取り扱ってはならない
17条	偽りその他不正な手段によって個人情報を取得してはならない
18条	利用目的を本人に通知し、または公表しなければならない
19条	個人データを正確かつ最新の内容に保たなければならない
20条	個人データの安全管理のために必要かつ適切な措置を講じなければならない
21条～	従業員、委託先に対する必要かつ適切な監督を行わなければならない
23条	本人の同意を得ないで、第三者に提供してはならない
24条	保有個人データに関する事項を本人の知り得る状態に置かなければならない
25条～	本人から開示、訂正、削除などの求めがあった場合、応じなければならない

図2-6-6

IV マイナンバー法案について

社会保障・税番号制度の概要

- マイナンバー法により、国民の個人識別番号である「社会保障・税番号制度」を導入することにより、国民の権利と義務の公平性、透明性を確保するとともに、国民の利便性の向上を図ることが可能となるほか、行政の効率化、透明化に資する効果も期待される。
- 個人番号（マイナンバー）**
 - 国民の個人識別番号として、国民一人ひとりに一意に付与されるマイナンバーを指し、国民により本人に通知、通知、通知等の通知を受けることにより、国民の権利と義務の公平性、透明性を確保するとともに、国民の利便性の向上を図ることが可能となるほか、行政の効率化、透明化に資する効果も期待される。
 - マイナンバーの通知は、国民一人ひとりに一意に付与されるマイナンバーを指し、国民により本人に通知、通知、通知等の通知を受けることにより、国民の権利と義務の公平性、透明性を確保するとともに、国民の利便性の向上を図ることが可能となるほか、行政の効率化、透明化に資する効果も期待される。
 - マイナンバー法に規定する通知を除き、本人にマイナンバーの提供を受けることは、本人からマイナンバーの提供を受ける場合、個人番号カードの交付を受ける等の本人確認を行う必要がある。
- 個人情報保護**
 - マイナンバー法の規定によるものを除き、特定個人情報（マイナンバー付きの個人情報）の収集、保管、加工、利用、提供、開示、滅失等の取扱いについては、行政機関等は情報提供ネットワークシステムでの情報提供などマイナンバー法に規定するものに準拠する。ただし、行政機関等は情報提供ネットワークシステムでの情報提供などマイナンバー法に規定するものに準拠する。ただし、行政機関等は情報提供ネットワークシステムでの情報提供などマイナンバー法に規定するものに準拠する。
 - 情報提供ネットワークシステムでの情報提供を行う際の適格性としてマイナンバーを用いないなど、個人情報が第三者に提供されないよう努める。
 - 国民が自宅のパソコンから情報提供等の取扱いを受ける仕組み（マイナンバーポータル）の提供、特定個人情報保護評価の実施、個人情報保護委員会の設置、罰則の強化など、十分な個人情報保護策を講じる。
- 法人番号**
 - 国民の個人番号は、法人等に法人番号を通知、法人番号は国民の個人番号と紐付けされる。法人番号は、国民の個人番号と紐付けされる。法人番号は、国民の個人番号と紐付けされる。
 - 法人番号は、国民の個人番号と紐付けされる。法人番号は、国民の個人番号と紐付けされる。法人番号は、国民の個人番号と紐付けされる。

http://www.cas.go.jp/jp/houan/120214number/gaiyou.pdf

図2-6-7

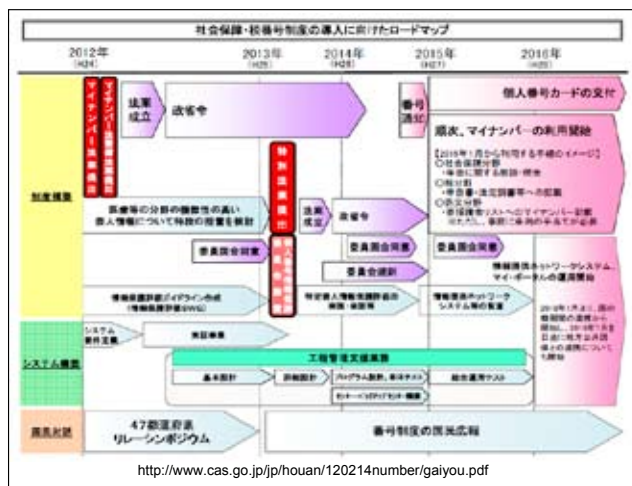


図2-6-8

V 医療分野における法整備の動向

- 厚生労働省「社会保障サブワーキンググループ及び医療機関等における個人情報保護のあり方に関する検討会の合同開催」

「医療等分野における情報の利活用と保護のための環境整備のあり方に関する報告書」（平成24年9月18日）
<http://www.mhlw.go.jp/stf/shingi/2r9852000002k0gy.html>

図2-6-9

医療等分野における情報の利活用と保護のための環境整備のあり方に関する報告書 概要3-1

II. 医療等情報の利活用と保護に関する法制に盛り込むべき事項について

情報の取扱いについて国民が安心でき、医療等情報の取扱者が情報の利活用に要しないための法制

- 情報の取得・活用について、本人への医療等サービスの提供等の場合、明示等その旨及び情報管理責任者等を明示することで同意を得たことによる取扱いを明確にする。また、専ら公衆衛生や学術研究目的の場合、一定のルールに従って取り扱うようにすることが適当であり、安全に匿名化された情報の取扱いとあわせて、具体的に検討する。
- 情報の保管、委託等における安全管理措置等
 現行ガイドラインの内容を基本に、医療等情報の利活用が促進されることに伴い必要となる安全管理措置等の具体的な内容について、引き続き検討する。
- 国民の信頼・安心を確保し、情報の取扱者が要しないための罰則のあり方
 個人情報保護法での罰則の取扱いをどうするか、医療等ID（仮称）取扱者の適正な取扱い、不正な手段によるデータ取得の防止の措置、漏失の場合に際する罰則の適用の検討など、悪しき原因を持つ者による個人の権利・利益の不当な侵害を未然に防止・抑制しながら、情報取扱者が情報の利活用に要することがないための仕組みを検討する。
- 主務大臣・第三者機関の関与の仕組み
 医療等情報を保護するための監視・監督（行政機関も含む）を専門的・公平に行うとともに、個人情報の取扱いをめぐる個人からの苦情等に対応する機関として、第三者機関が必要と考えられる。本人の権利利益の侵害のおそれがある場合に通告を受け、主務大臣に対して必要な権限行使を求めるなどの仕組みについても検討する必要があると考えられる。

http://www.mhlw.go.jp/stf/shingi/2r9852000002k0gy-at/2r9852000002k0la.pdf

図2-6-10

医療等分野における情報の利活用と保護のための環境整備のあり方に関する報告書 概要3-2

II. 医療等情報の利活用と保護に関する法制に盛り込むべき事項について

法制の適用のあり方

- 医療等に関する個人情報の範囲
 社会保障・税番号大綱等での記載も考慮し、基本的に医療・介護分野における「生命・身体・健康」に関する個人情報に限定し、死亡等の情報についても、一定の措置の対象とする方向で検討する。
- 死者の情報の取扱い
 医療や介護分野では死者の個人情報についても何らかの安全管理や目的外利用等の制限に配慮する必要がある。現行ガイドラインも参考に、死亡等の個人情報についても、一定の措置の対象とする方向で検討する。
- 安全に匿名化された情報の取扱い
 医学等の研究や一層進展し、医療者が管理することなく情報を取り扱うことが可能となるよう、「安全に匿名化された状態がどのようなものか」について明らかにし、引き続き検討し、そうした匿名化措置により法で規制される個人情報でないことを明確にするべきである。
- 小規模事業者に対する適用
 現在個人情報保護法に規定されていない小規模事業者も一定の措置の対象とする方向で検討する。
- 医療等の個人情報を取り扱う主体に共通するルール
 今回の法制では民間部門のみならず、行政機関、独立行政法人等や地方自治体を含めたルールを定めるべきである。
- 適用除外に関する考え方
 個人情報保護法が適用除外である学術研究について、医療等分野での研究の重要性から、患者等がデータを提供することへの安心感を確保し、個人情報の取扱いに関する義務の確保と、学術の自由の確保のバランスをとる観点から検討を要するべきである。例えば、安全管理措置等の履行確保のための仕組みの一部に属して法制の対象とするとも考えられる。

http://www.mhlw.go.jp/stf/shingi/2r9852000002k0gy-at/2r9852000002k0la.pdf

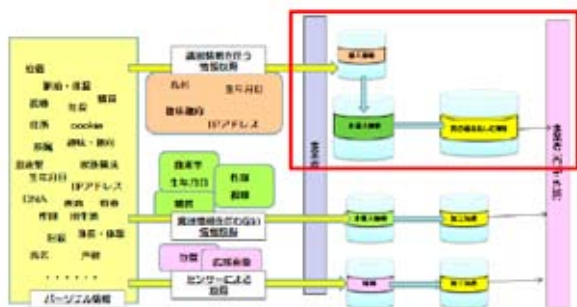
図2-6-11

本人同意に関する暫定私案

- 学術研究および公衆衛生の向上を目的とする場合は、一定の条件のもとに、**本人同意の要件を緩和**すべきではないか
- 仮名化の場合（匿名化については後述）
 - 学術研究または公衆衛生の向上を目的とすること
 - 第三者に提供する目的を本人に対して通知し、または公表すること
 - 対応表は、個人情報取扱事業者が安全に管理し、第三者に提供しないこと
 - 仮名化されたデータの提供を受けた事業者は、当該データの安全管理のために必要かつ適切な措置を講じること

図2-6-12

VI 個人情報の匿名化に関する課題



出典：パーソナル情報の利用ガイドライン(案) 利用のあり方に関する提言

図2-6-13

EU個人データ保護改正規則提案(2012)

規制の対象となる個人データの範囲を拡大する傾向があるが、匿名化については、どのような態度をとっているか

(23) The principles of protection should apply to any information concerning an identified or identifiable person. To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual. **The principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.**

図2-6-14

アメリカの議論状況 — HIPAA法

HIPAA法(Health Insurance Portability and Accountability Act of 1996)

健康情報の個人識別性が否定される場合

(1) 統計および科学に関する適切な知識、経験を有する者が、当該情報が個人を識別するために利用されるリスクが非常に小さいと判断する場合で、かつ当該判断の正当性を文書で証明する場合

図2-6-15

(2) 個人などに関する18の識別子が削除される場合

- ① 氏名
- ② 州よりも小さな全ての地理的区分(ただし、郵便番号の最初の3ケタは除く)
- ③ 直接的に個人に関係する全ての日付(年を除く)
- ④ 電話番号
- ⑤ FAX番号
- ⑥ 電子メールアドレス
- ⑦ 社会保障番号
- ⑧ カルテ番号
- ⑨ ヘルスプラン受益者番号
- ⑩ 口座番号
- ⑪ 認証・ライセンス番号
- ⑫ 車体識別番号・シリアル番号・ナンバープレート
- ⑬ デバイス識別番号
- ⑭ URL
- ⑮ IPアドレス
- ⑯ 指紋および声紋を含む生体認証情報
- ⑰ 完全な顔写真
- ⑱ その他の個人を識別する番号、特徴など

図2-6-16

匿名化の意義・分類

(1) 連結可能匿名化

必要な場合に個人を識別できるように、その人と新たに付された符号又は番号の対応表を残す方法による匿名化をいう

(2) 連結不可能匿名化

個人を識別できないように、その人と新たに付された符号又は番号の対応表を残さない方法による匿名化をいう

(厚生労働省「臨床研究に関する倫理指針」、その他、文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」、経済産業省「経済産業分野における個人情報保護ガイドライン」などにも類似の定義をしている)

「連結可能匿名化」→ 仮名化
「連結不可能匿名化」→ 匿名化

図2-6-17

匿名化の方法・技術

■ 単純匿名化

個人単位のレコードから、直接個人を識別できる氏名やIDなどの情報を切り落とし仮名化、あいまい化などによって取り除く方法

■ 集合匿名化

住所や性別、年齢などの組み合わせによって個人識別につながる可能性のある属性情報を加工し、どう組み合わせても個人を特定できないよう、すなわち一定数未満の人数に対象を絞り込めないように匿名化する方法

■ k-匿名化

間接的に個人を識別できる可能性がある属性項目を一般化し、同じ属性情報を持つレコードが最低でもk個以上存在するように、パーソナル情報を集合匿名化すること

出典：日立コンサルティング「基盤利用マニュアル」平成22年3月

図2-6-18

匿名化に関する個人情報保護法の解釈論

個人情報とは何か？

・個人情報の定義(個人情報保護法2条1号)

「生存する個人に関する情報であって、当該情報に含まれる氏名、生年月日その他の記述等により**特定の個人を識別することができるもの**(他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む)」

個人情報に該当するための要件

- ①**生存性** … 生存する個人に関する情報
- ②**個人識別性** … 特定の個人を識別することができるもの
- ③**容易照合性** … 他の情報と容易に照合することによって特定の個人を識別することができる

→ **匿名化によって、これらの要件を否定できるか**

図2-6-19

匿名化に関する課題① — 容易照合性



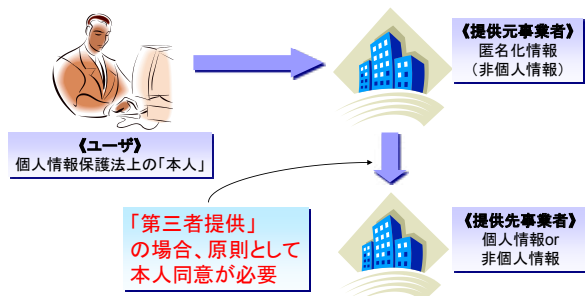
加工の元となった個人情報と匿名化した情報の間に容易照合性が残っていると、個人情報該当性を否定できない

《経済産業省ガイドラインQ&A14》

「経営者、データベースのシステム担当者などを含め**社内の誰もが規程上・運用上、双方のデータベースへのアクセスを厳格に禁止**されている状態であれば、「容易に照合することができ」とはいえない」

図2-6-20

匿名化に関する課題② — 第三者提供



提供元にとって非個人情報であれば、**提供先にとって個人情報であっても、本人同意が不要となるのか**

図2-6-21

アメリカの議論状況 — 最新の学説



ポール・オーム
http://www.wired.com/images_blogs/photos/uncategorized/2008/09/04/paulohm.jpg

「破られたプライバシーに関する約束—匿名化における驚くべき失敗について」(Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 UCLA Law Review 1701 (2010))

- ・匿名化技術全般について懐疑的な立場
- ・HIPAA法の匿名化の基準についても批判的
- ・プライバシーリスクを評価する5つの要素(①データ加工技術、②私的開示と一般公開、③量、④動機、⑤信頼)

図2-6-22

匿名化に関する考察①

- 一定の匿名化がなされていれば、個人情報に該当しなくなるという手法には課題もある
- 安全な匿名化がなされていても、**再特定化**されるリスクが残るため
- 匿名化の問題は、プライバシーに対する「**リスク評価**」として考えた方がよいのではないか
- 何らかの第三者機関が、どの程度のリスクがあるのかを判断するようなスキームが必要ではないか

図2-6-23

匿名化に関する考察②

■ 匿名化に関するリスク評価のための要素

- (1) **匿名化の程度**
単純匿名化か集合匿名化か、kの値はどの程度か
- (2) **公開する範囲**
一般公開か特定の者に対する開示か
- (3) **情報の質と量**
対象となる情報は機微情報か、多いか少ないか
- (4) **動機**
学術研究の目的か、経済的な目的か
- (5) **信頼**
データの受領者は信頼できるか(学術研究者か、政府か、広告機関か)

図2-6-24

バイオデータ構造化の課題

金 進東 先生（ライフサイエンス統合データベースセンター特任准教授）

- ▶ データベースの統合活用という本計画の特徴は、規模が大きいことと、対象にするデータの性質が結構違うデータベース群が対象になっていること、そして、この大規模で多様性のあるデータを長期間にわたって開発する必要があることである。よって、開発の期間中、状況がどう変わるかわからないところが大きな課題である（図2-7-3）。このようなデータベースの計画に当たって重要になるのがデータ生産の分散。分散して生産し、蓄積しながら開発する。開発中に状況が変わっても、その変わった状況に柔軟に対応していくアプローチが必要になる（図2-7-4）。
- ▶ 最近広がっているのが、リンクド・オープン・データ（linked open data）とかシマンティック・ウェブとか、ウェブ・オブ・データとかRDFとか、いろいろな呼び方があるが、一番理解しやすいのがウェブ・オブ・データという用語。我々が日常生活で使っているWWWというのは、ウェブ・オブ・文書、ウェブ・オブ・ドキュメント。これに対して、データが今のWWWになるモデルをウェブ・オブ・データと言っている。WWWが分散されている文書がつながっていることによって有用性を持つというモデルに対して、データも分散して開発し、それをつなげることによって効率高く開発ができるといったモデル。データベースの統合化には、リンクド・オープン・データ・システムがふさわしいと考えている。
- ▶ リンクド・オープン・データにはRDF（Resource Description Framework）という表現形式を使う。リレーショナル・データベース（RDB）とよく比較されるので、RDBに対しては理解している前提で話す。例えば、ある病院でこんなRDBを持つと想定する（図2-7-6）。男性の患者さんで、35歳で、身長、体重がわかっていて、担当医は山口先生で、その担当医の所属は内科で専門は腎臓である。よく見ると、このデータでは山口先生に対する情報が重複していて、余り効率が高くない。
- ▶ 普通はテーブルを分ける。分けて管理することによって効率がよくなる。分けることによって先生のデータは別のテーブルになって重複がなくなる（図2-7-7）。テーブルは分かれてもフォーリン・キーというものがあることによってテーブルがつながる。例えば「太郎さんの担当医の専門は何ですか」という質問にもフォーリン・キーを使ってテーブルをあわせて答える。
- ▶ このモデルを続けていくと、もっと分けることが可能である（図2-7-8）。何がいいかというと、例えばこのデータ、例えば「太郎さんは男性で、年齢は35歳ですよ」という情報は、実は役所で管理したほうがいい情報で、例えば身長とか体重は保健所等で管理したほうがいい。このように情報コンテンツにより管理場所を分けて、分散して管理させたほうが効率が上がる。
- ▶ もっと極端に分散させて、分けることが可能であればもう全部分けておいて、必要に応じてテーブルをジョインして質問に応じて答えるというモデルが考えられる（図2-7-9）。これを普通データベースの理論からは垂直的な分散と言い、データを分散させて、開発と管理も分散させて、コストも分散させてリスクも分散させるというモデルが最近、広がっている。
- ▶ 垂直的な分散だけではなくて、実は水平的な分散もできる（図2-7-10）。例えば、こ

のような役所が管理しそうな情報も水平的に分けてしまって、太郎さんが新宿区に住んでいたなら新宿区の役所が管理して、花子さんが文京区に住んでいたなら文京区の役所が管理してと、分散して開発もしくは管理ができる。

- ▶ このように極端にデータを分散させる技術がRDFである。ここまでデータを分散させると、実はすべての情報が3つ組で表現できる。人1の名前は太郎で人2の名前は花子みたいに、すべての情報が3つ組で表現する。RDFトリプル (triple) と呼ぶ (図2-7-11)。
- ▶ RDFは、分散させて開発のコストとリスクを減らす方法。データを分散させる技術すべてをインターネット基盤で、分散されているデータをつなげるところにインターネット技術を使う (図2-7-12)。RDBとRDFを比較してみると、テーブルモデルのRDBは目的に最適化してテーブルを設計するモデル。だから1つの目的に対しては最適化されていて早い。一方、RDFの技術はいろいろな応用に柔軟に対応するモデル。1つ1つの運用には遅いがいろいろな目的に使えるという長所を持っている (図2-7-13)。それを簡単に言うと、組み立て自由のための分解とも言う。
- ▶ 以上のRDFの技術を使って本計画にどう対応できるか考えてみる。例えばデータ収集のときに集まった情報をRDFで表現して、データ収集の段階とその後の構造化 (分類体系) の段階を分けて、データを生産するところはデータの生産に集中して、分類体系をつくるところは分類の体系づくりに集中する。それによってデータを最適化して実際に使える公共データもしくはサービスを開発していくというモデルがいいのではないかと提案する。医療関係のデータはセキュリティ管理が重要な問題なので、セキュリティ管理が必要なところは分ける必要がある。
- ▶ このモデルのいいところは、データが分散されて役割も分散されること。国の役割はインフラ提供とか方向性を提供するとか、分担を決める等の役割が重要で、セキュリティの管理と保障も重要だと思う。開発の側面で一番重要なのはデータの生産である。WWWの成功の1つの理由は、データの生産が簡単だったという側面がある。
- ▶ 技術的な面ではデータ生産の生産性が上がるような技術開発が必要になる。データ生産性を上げるためには、データ生産自体が普通の業務に密着している必要がある。

(質疑応答ならびに討論)

- ▶ 先ほどのデータの増え方は爆発的。これからもさらに爆発的に増えていくと思うが、これからはどのようにデータベースを集めるあるいは、整えていけばいいか。
- ▶ 最近話題になっているビッグデータの問題は、爆発的に増えたデータをどう処理するかということだが、それはデータ収集の技術が発達したためである。逆に言うとビッグデータ問題を持っている分野はデータの収集、もしくは、生産が進んでいる分野との意味である。医療分野は生産されるデータの量は大きいがその電子化がもっと進む余地がある。また、セキュリティの問題で共有できてないデータの共有化ももっと進めてビッグデータ問題に持って行く必要もあると思う。
- ▶ 分散型にしたときに、その分散そのもののマネジメントはどう考えるのか。
- ▶ リンクド・オープン・データの形でデータを分散させるとデータは分散されるけどアクセスは標準のプロトコルを使ってインターネット経由でできるのでマネジメントが容易になる。しかしながら、セキュリティの問題は残るので、関連の技術開発が必要。

- ▶ RDFは遅いとあるが、例えばRDFトリプルという形で、3つの項目である程度集約化して並列分散をかけると、速度の問題は解決できるという解釈でよろしいか。
- ▶ RDFのモデルはデータの生産に向いている技術。とにかく多様な性質を持っているデータを分散して生産する、最適化はその後の段階で実施するモデル。だから、後で実際のサービスをつくるときRDBを使ってもいい。レイジー・オブティマイゼーションという技術である。RDBも1970年代に提案されたときは使い物にならないぐらい遅いと言われていた。それが技術革新でどんどん早くなって、今、日常的に使える。だからRDFも標準化すれば、多分それに向けたアーキテクチャが出てくるので、技術が追いついてくるだろう。

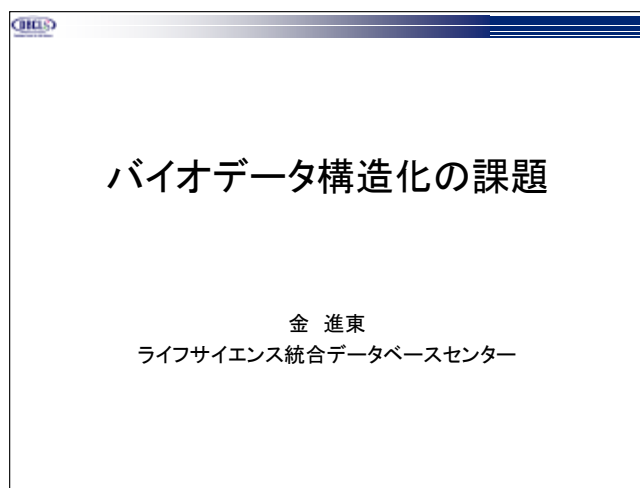


図2-7-1

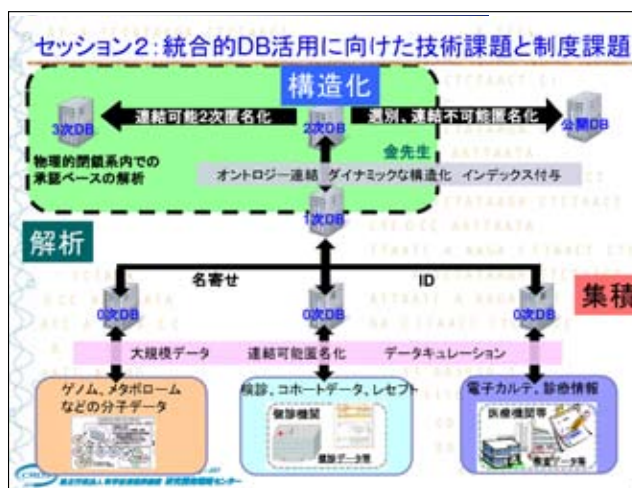


図2-7-2

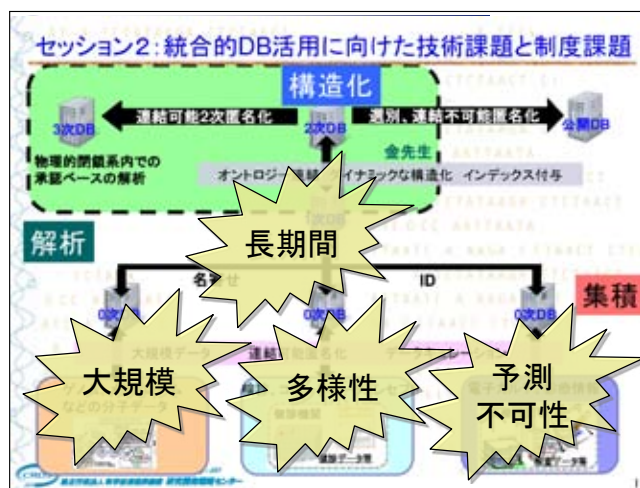


図2-7-3

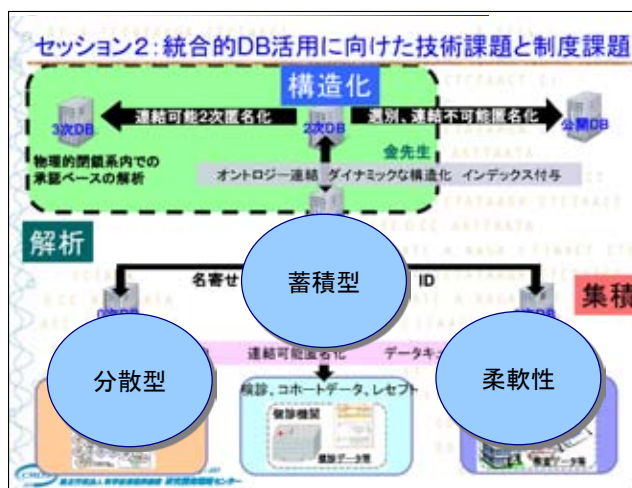


図2-7-4

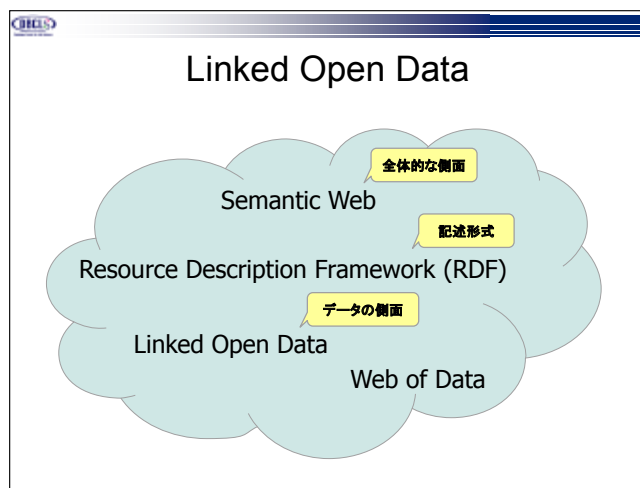


図2-7-5

RDB vs RDF (1/6)

ID	名前	性別	年齢	身長	体重	担当	所属	専門
1	太郎	男	35	170	68	山口	内科	腎臓
2	花子	女	29	161	53	山口	内科	腎臓
...	...	...	...	...	...	...	...	...

図2-7-6

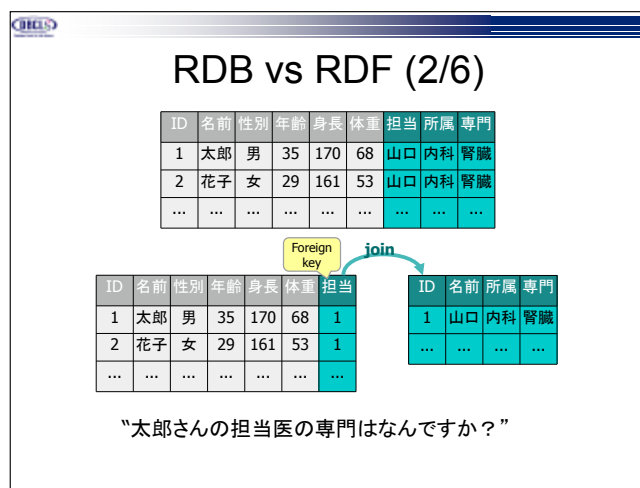


図2-7-7

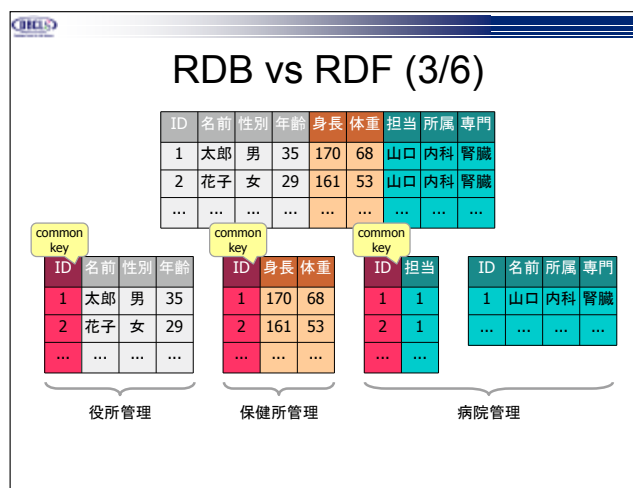


図2-7-8

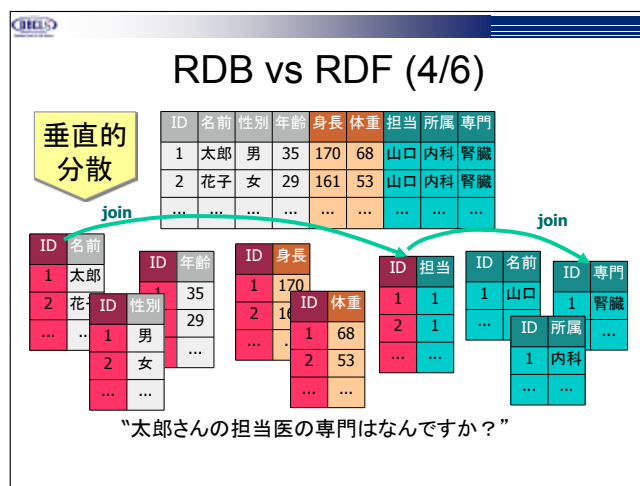


図2-7-9

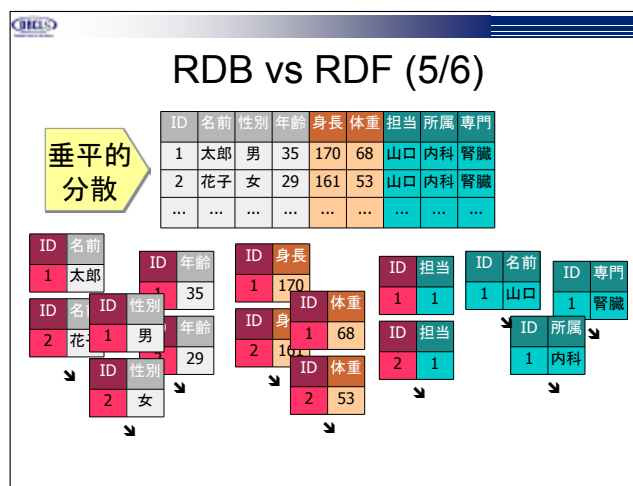


図2-7-10

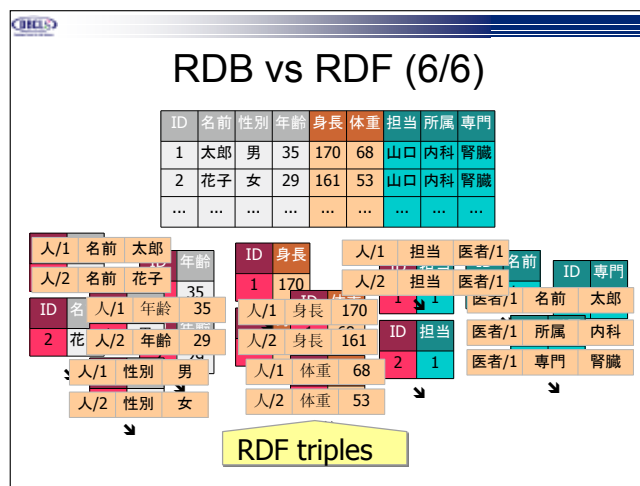


図2-7-11

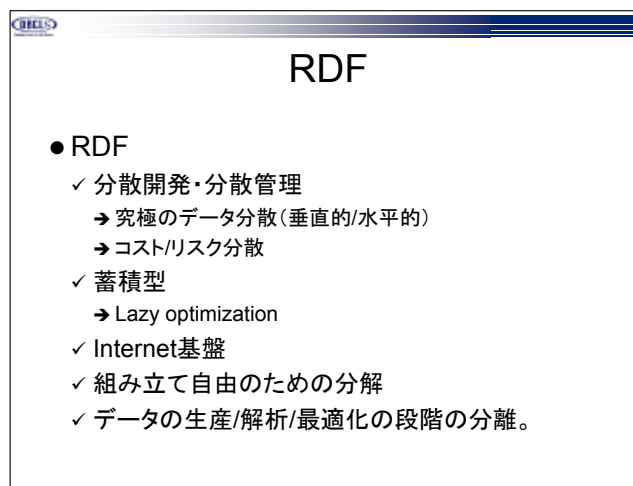


図2-7-12

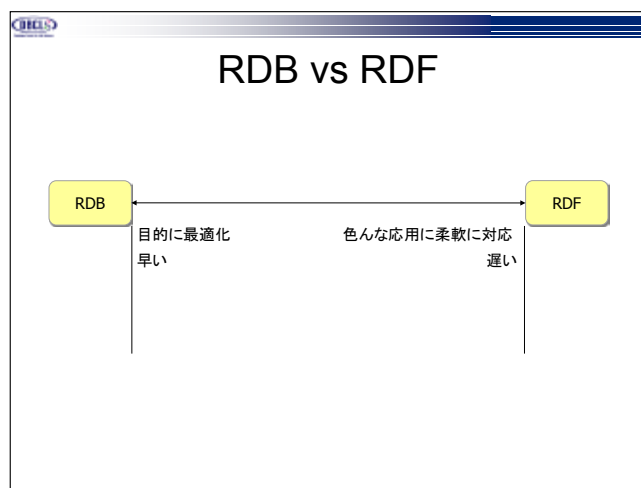


図2-7-13

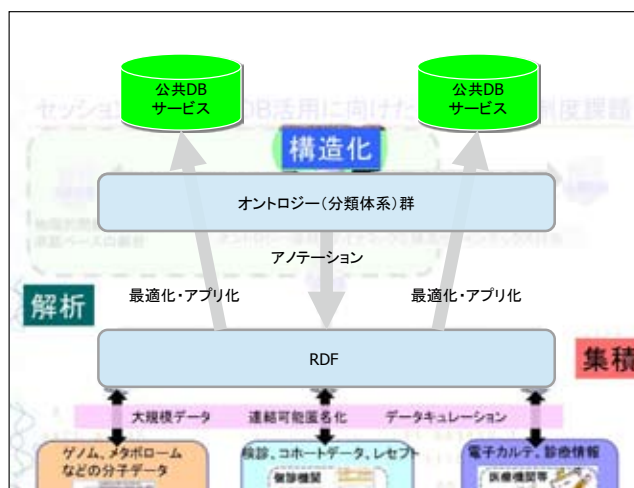


図2-7-14

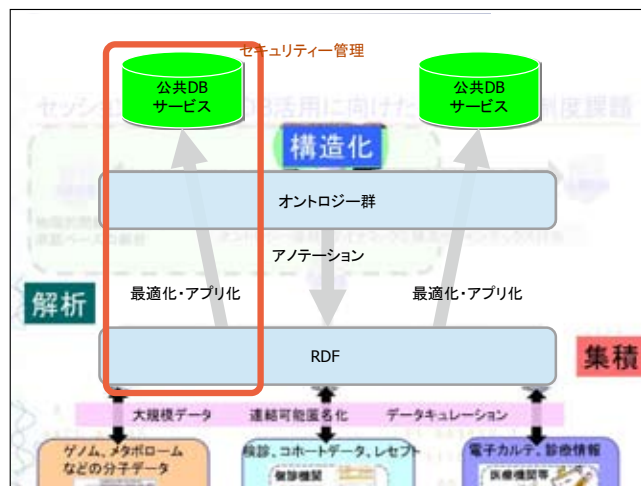


図2-7-15

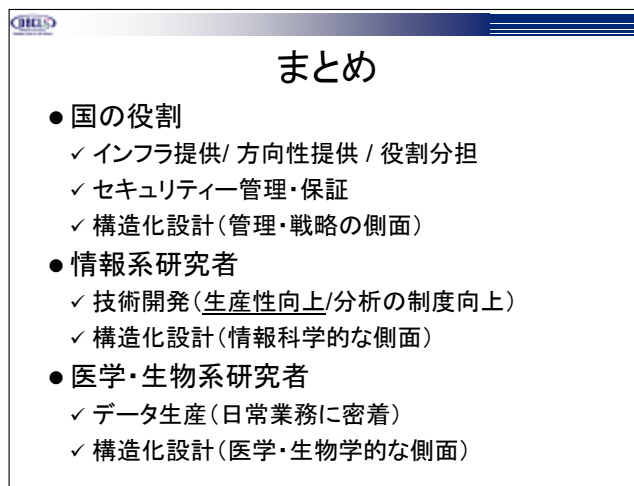


図2-7-16

ゲノム疫学における課題

加藤 規弘 先生（国立国際医療研究センター研究所遺伝子診断治療開発研究部長）

- ▶ ゲノムの文字情報を読むという構造的な理解（ヒトゲノム計画）から、書かれている意味の生物的理解（エンコード計画：2012年発表）へと移ってきている。GWAS手法の本格導入から、治療薬の効果的な使い分けや副作用の把握などへの応用も試みられており、今後の10年間で最も期待されるのは個別化医療の進展だろう（図2-8-1）。
- ▶ 構造的な理解を進める過程において、ヒトの1000人ゲノム計画で最も注目されたのは大きく4つの人種（白人、東アジア人、南アジア人、黒人）、或いはヒスパニックを入れると5つの人種である。欧米からすると、疾病ゲノムのアジア方面では、人口の多い中国（東アジア）、インド（南アジア）が大きな柱となりつつある（図2-8-2）。
- ▶ 2003年にヒトゲノム計画が終わって、その後は個人差、バリエーションを探すHapMap計画が始まり、そこで得られた成果をGWASで解析を行った。GWASによってバリエーションと健康情報などの関連づけができつつある中で、技術を磨いていってエキソームシーケンス、そしてダイレクトシーケンスに移っていった（図2-8-3）。
- ▶ 中国がゲノム医学に相当力を入れている。イギリスのUKバイオバンクになぞらえた形で、カドーリバイオバンクスタディを中国は行っている。既に51万2,000人をサンプルとして集めている。また、中国は、自分たちだけで十分それを集めたり加工することはできないので、オックスフォード大学からマネジメント面も含めたバックアップを受けている（図2-8-4）。
- ▶ 「NEJM」「Lancet」「Science」「nature」「Nat Genet」の5誌に発表されているものを人種別（東アジアについては中国と日本）に分けたところ、欧米が圧倒的に多くて119報、中国が18報、日本が11報、アフリカ系アメリカ黒人を含めたものが2報であった。2007年には日本が優位だったが、昨年あたりはもう圧倒的に中国からのパブリケーションが増えている（図2-8-5）。
- ▶ 個別化医療開発には4つの研究段階がある。1つ目は、SNPやバリエーションの探索をランダムに行う段階。2つ目は、それがバイオロジカルに病気との意味づけがあるかを検証する段階。3つ目は、それが個人の病気に対してどれぐらいのインパクトがあるかを評価する段階。4つ目は、それが費用対効果として本当に保健医療あるいは療養指導等に使えるかを検証する段階（図2-8-6）。
- ▶ また、個別化医療開発には、コホートも密接に関係してくる。例えばアメリカでは、RCTと連動させたゲノム情報の収集と解析、電子カルテへのビルトイン等も始めている。こういった取り組み等と関連して構築される知識集積型データベースはアメリカが主導しているが、東アジア人を対象としたものはまだ無い。従って、日本、中国、あるいは韓国も含めて、今後どう取り組んでいくかが今後の課題となるだろう（図2-8-6）。
- ▶ 日本が海外と対等に進めて行くためには、日本の実データをきちんと持つ必要がある。日本で、DNAを用いたゲノム解析で論文化されたものは既に多数ある。これらを寄せ集めると25万人に達する。その中には、責任者が変わってしまっただけでサンプルが行方不明になったものもあるだろうが、例えばGWASの1つのツールとしてこ

れらサンプルを提供し、結果を共有データベースに蓄積して日本のデータを増やしていくやり方もありうるだろう（図2-8-7）。

- ▶ **GWAS**を方法論として普及させ、実データを次々に蓄積して行ったのが、**GAIN**という米国における**GWAS**のネットワークである。これは、公募して**GWAS**の費用を出す、成果を全て登録することが義務付けられている。**GAIN**の継続事業として**GENEVA**が始まったが、こちらでは更に対象疾患を増やしている。更にその次のステップとしては、白人だけでなく他の人種のことを考慮しないといけないという発想が出てきて、他人種を対象とした試みが出てくる。ただ、これらは全部、**GWAS**の普及と成果データの登録といった探索フェーズだけである。その後の検証フェーズに入ってきているのが現段階である（図2-8-8、9）。
- ▶ 臨床応用への展開として**eMERGE**がある。これは電子カルテとゲノム情報を連動させようという試みで、現在第2期に入っている（図2-8-10）。
- ▶ **GARNET**という米国内のネットワークもある（図2-8-11）。
- ▶ 最後に今後の課題を列挙する。まず、米国中心に欧米人のデータベースが整備されつつあるが、それに対して、日本はどう差別化し、相補的なデータベースを構築していくのか。そして、東アジアの中で中国や韓国に対してどう優位性を確保していくのか。新規収集分だけでなく、既存のサンプル／コホートを活用して実データをある程度上乗せし、厚みをつけるなどの戦略。探索フェーズから次第に臨床応用フェーズへ移行している中で、どのようなデータベースを構築すべきか。最後に、ラーニングデータベースという考え方（図2-8-12）。

（質疑応答ならびに討論）

- ▶ 既存サンプルの活用について。非常に重要だと思うが、それぞれ基準が違うから合わせられないといった議論が必ず出る。**GWAS**の場合は問題ないのか。
- ▶ 細かい診断基準を言い始めると切りがない。ただ、広いカテゴリーの中ではある程度分類できるので、まずはそこから始めると良いのではないか。
- ▶ もう1つ必ず出てくるのが、昔の試料に対して**IC**をどう取るのかという点。
- ▶ 病気の種類によって状況は異なるが、連結不可能匿名化にし、当面は其中で得られるような成果とする。世界的でかつ、急速な流れからいくと、5年、10年と時間をかけてやっても仕方がない話なので、2～3年の短期間で日本独自の厚みのあるデータベースを構築すると良い。
- ▶ 日本では、データを出す側とデータベースを作る側が別々のプロジェクトで、別々に戦略を立てられている。総合的に戦略を立てないと、海外に太刀打ちできないと思う。
- ▶ 人を育てるためどうすればよいか。若い人たちが入ってくるような仕組みというのは、ポジションなのか、給与なのか、研究費なのか、或いはキャリアパスなのか。この問題について、危機感を持っている。
- ▶ 日本人は工夫が得意なので、各人が与えられた仕事を淡々とやるのではなく、考える余地、インセンティブをうまく盛り込むべき。指摘の中で一番重要なのはキャリアパスであろう。
- ▶ ゲノム研究拠点の集約化について、中国以外の動向は？
- ▶ 欧州では、英国のサンガー研究所と**EMBL**が中心機関として機能している。米国は、

4カ所程度の拠点（Baylor College of Medicine、University of Washington、Broad Institute、Johns Hopkins Universityなど）が、合議体を形成して進めている。韓国は中国を見習って、後発ながら拠点を作る動きがある。アジアでは、シンガポールも拠点を作った。アジアの拠点は、研究要素をできるだけ少なくして、解析能力を卓越させ、解析を海外から受託している。米英への留学者が、ノウハウを持ち帰って、各拠点の運営に活かしている。

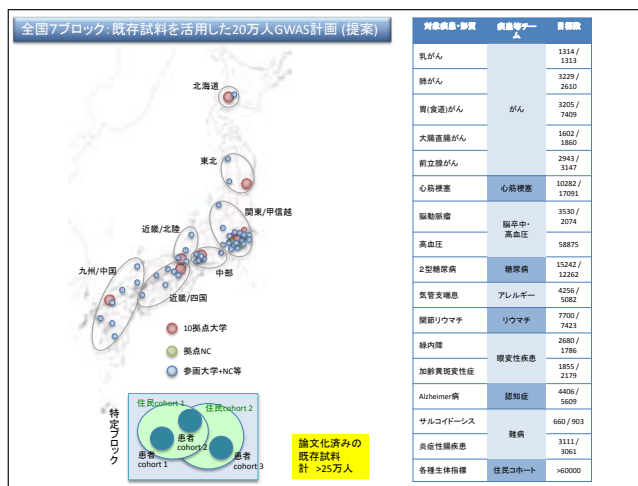


図2-8-7

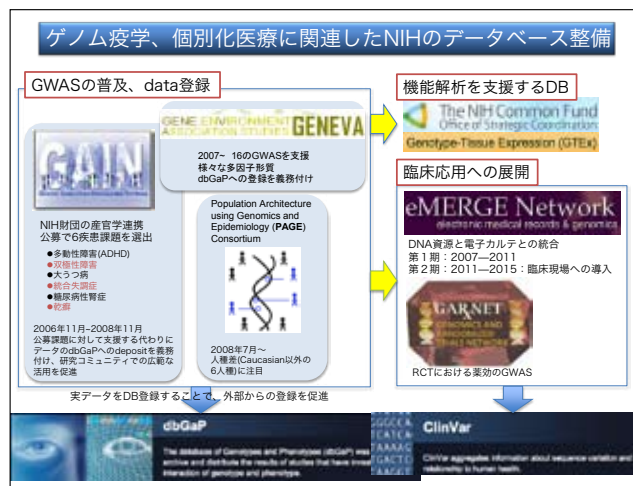


図2-8-8



図2-8-9

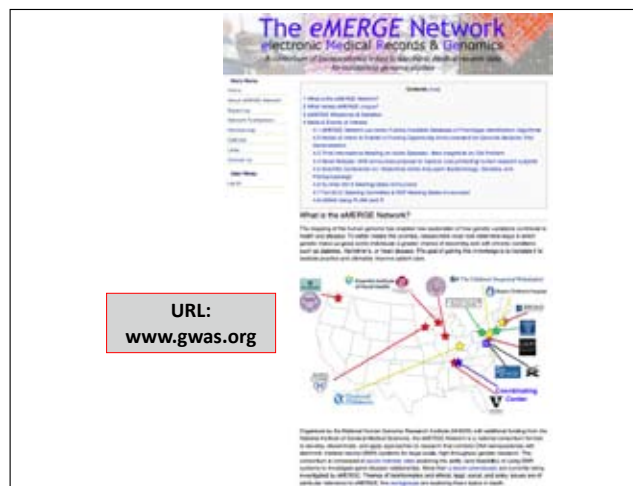


図2-8-10

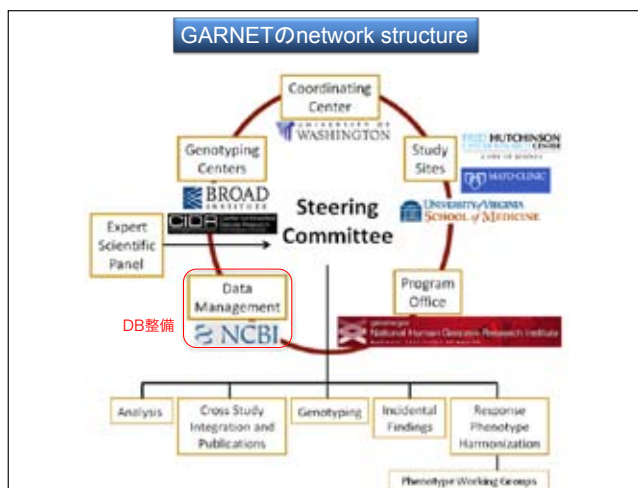


図2-8-11

- 課題
1. 米国NIHを中心に整備されつつある、欧米人のDBとの差別化・相補性
 2. 東アジアのなかでの、中国、韓国に対する優位性の確保
 3. 新規収集分だけでなく、既存のサンプル/コホートの有効活用：国としてのライフサイエンス戦略
 4. 「臨床応用への展開」というフェーズでの時宜に合ったDB構築
 5. 基盤的知見のDB化とともに、臨床研究等との連携・組織化が重要：“learning DB”の考え方

図2-8-12

セッション2 「統合的データベース活用に向けた技術課題と制度課題」 討論

セッション1、2の議論で指摘された技術面、制度面の課題を図2Bにまとめた。幾つかのカテゴリーに整理して掲載している。本討論では、優先順位が高い課題を中心に議論した。取り上げた主な課題は、法制度（医療等情報個別法、共通ID、医療等IDなど）、セキュリティ技術（匿名化技術、暗号化技術など）、オンサイト解析センター、データベースの維持（技術、費用）、人材育成などである。

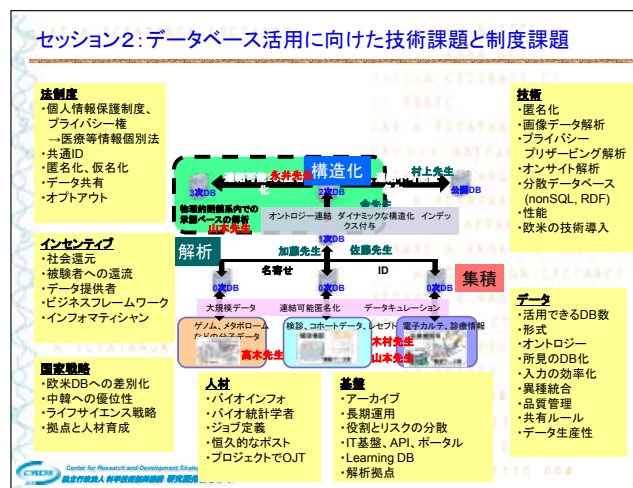


図2B セッション2討論資料

法制度（医療等情報個別法、共通 ID、医療等 ID など）

- ▶ 現時点の我が国政府の動向としては、医療に特化した新しい法律をつくる方向で進んでいる。厚生労働省が中心となり、内閣官房と連携しながら進んでいる。学術研究や医学研究のために臨床系データのデータベース化が必要で、それを連携して使いたいという要望を厚労省に明確に提示するべきで、その点を法律に盛り込まないと、法律ができてからの対応では遅いだろう。当初スケジュールでは、2013年中には法案が出来上がるロードマップになっている。実際は遅れていると思われるが、早く意見出しをした方がよい。医師会からの意見が強く反映されると、患者情報の保護が前面に出てくる。研究等への利活用とその公益性を強く発信した方がよい。ただし、JSTあるいはこのWSからの発信が、厚労省対文科省という対立構造を生まないように配慮するべき。
- ▶ 個別法を作る必要性を訴えたのは、情報活用を進めたい側のステークホルダーであり、医療情報を学術、公益に結び付けるべきとの理念を持っている。しかし個別法の議論を進めると、医療側からは保護に重点を置いた意見が返されてくる。また文科省は、刑事罰を盛り込むことに反対の立場である。しかしながら、活用を考える場合に医療等IDは必須であり、それを法律で規定しなければならず、そこに付随する情報には一定の規制を入れなければならない。確かに医師会、歯科医師会は保護に傾いた立場であるが、全体の論調としては、いわゆる公益利用、学術利用を否定するのではなく、安全に、すなわち患者さんに被害が及ばない形で情報活用を進めるべきとの立場である。情報を提供する個人、患者へ、本人たちが望まない影響を与えずに活用を

進める方策を示して、理解を浸透させることが重要である。その方策の主要論点は、匿名化のリスク評価である。匿名性に関するリスク評価には2つのレイヤーがあると考えている。1つはコントロールされた状況下での情報利用時でのリスク。情報解析を進める際には、匿名性や機微情報の秘匿をコントロールした上で、研究者に提供し、さらに研究者の行動規範に依存して全体のリスク程度が決まる。ここは合理的に基準を作れるだろう。2つ目は、アンコントロールな状況下でのリスクである。これは、解析結果を公表する段階に相当する。公開すれば情報というのは、もう全く制御が効かない。この段階で、その情報の主権者である患者さんに望まない被害が出ると大問題になる。ここは、かなり厳しいリスク評価をしておかなければならない。2つ目のレイヤーに関しては、1つ目のコントロールが故意に破られた場合や、まったく予期せぬ形で個人情報特定された場合などに相当するので、本WSの趣旨にはそぐわないだろう。ここでは、コントロールドユースの場合のリスク評価を議論することが重要だろう。1つ目のレイヤーでのリスクを低減するためには、オンサイト解析センターというセキュリティ管理された空間を使用することが、リスクを下げるという意味では相当有効な手段になるだろう。

セキュリティ技術（匿名化技術、暗号化技術など）

- ▶ セキュリティ技術というのは、100%完全なものはないという前提で話さなくては行けない。次に、その技術のリスク評価を定量的に行う必要がある。例えばドアの錠前だと、専門家が来て開けるのに何分かかるといって定量化ができる。その上で、必要に応じてそのセキュリティレベルを補う制度を作ることになる。例えば、この鍵は30分持つので、必要なセキュリティレベルを確保するためには、30分に1回ぐらい警備をするようにする、という対応を考える。技術に対する定量的なリスク評価と制度面での補完で、安全性を担保することになる。
- ▶ 暗号化技術についても、基本的には破られるという前提のもとでエンハンスをしていくべき。これは、多分皆が考えている。匿名性については、 k -匿名化という評価手法で、「 k 」が増えれば増えるほど推測されにくいと定量的に評価できる。そこに暗号化を組み合わせると、セキュリティ強度は、累乗的に強くなり、かなりの強度が担保できると考えている。暗号化のセキュリティについて定量化する際には、暗号を破るためのPC稼働時間を考えるが、コンピュータの計算速度にもよるため、水準を決めるのが難しい。それゆえ、技術だけではなく、運用面や制度でカバーする必要がある。

オンサイト解析センター

- ▶ 解析センターは、メタ解析を行うわけで、他機関との情報交換が発生する。情報交換の際に、漏出があった場合にどうするかについても議論が必要。米国では「遺伝情報差別禁止法（Genetic Information Nondiscrimination Act: GINA）」が2008年に成立していて、もし情報が漏れたとしても、遺伝子差別はされないという安全弁がすでに備えられている。
- ▶ 厚生労働省関連のナショナルセンターを結んだバイオバンクネットワーク事業の中で、国立国際医療研究センターに中央データベース管理部門を設立した。そこでも、

解析研究を行うべく議論を進めている。主要論点は、セキュリティの担保である。

- ▶ セキュリティの議論は、技術面の話と、心理面の話があると考ええる。オンサイト解析センターの話は、技術的な対応だけでなく、心理的な安心を与える効果も大きいのではない。技術面のことで、金融関係などで使われている暗号化技術をライフサイエンスで使うことが、十分か不十分かという議論が具体的、現実的ではないか。
- ▶ 医療とか医学データに特別な技術が必要だとは、私は全く考えていない。既存の技術で十分対応できる。問題は、個人の識別情報をいつの段階で外すかということである。ゲノムや臨床情報の解析には、個人の識別が重要であり、HIPAAの個人識別性が否定される条件の2番目にある18項目を解析の途中過程で削除することは現実的でない。ある程度個人が識別できるリスクを含んだ情報を使って分析しないと、研究が進まない。「結果は安全な状態として出すけれども、経過途中はやはり多少リスクはありますよ、ただ、そのリスクは他の方法で対処していますから安心してください」といった形をつくるのが具体的な方策であり、重要なことだ。

データベースの維持（技術、費用）

- ▶ ライフサイエンスのデータはシーケンスに代表されるように大きなものだが、特別な技術を開発する必要はなく、技術面では、一般のビッグデータの研究開発成果を活用すればよい。
- ▶ 長期的にそのデータを使うことを考えると、競争的資金で稼働している我が国アカデミアの多くのデータベースの維持が心配である。
- ▶ アーカイブするためのセンターは必要であり、NBDCやDDBJがその役目を果たすはず。DDBJは最近スーパーコンピュータに入れ替えて、ディスクも2年後に12.5ペタに増強する。それで5年は何とか運営できるだろうが、その先はまだ想定していない。
- ▶ 臨床情報のデータベースをつくっているが、研究費が続かないという問題にいつも直面する。データベースの維持は事業費的にするべきである。国に依存するのではなく、保健制度なり何らか医療制度の中に診療情報データベースを組み込む必要があると考えている。基幹的なデータベースは、制度の中で継続する必要がある。今は研究費レベルでデータベースが構築されているから、中止せざるをえないデータベースが多い。
- ▶ 中国のBGIのように、企業経営として遺伝子関連のデータベースを運営することはできないか。
- ▶ データベースについている研究費の規模が、民間の事業経費で賄えないほど大きい。データ生産とデータベース維持の原資を事業で回収するスキームが難しい。
- ▶ データの二次利用のところに企業が踏み込めるようになれば、データ維持を企業が負担する可能性があるが、今はその段階にまだない。二次利用で新たな収益源として見込める部分があれば、それに対し企業がリンクすることはある。例えば副作用のサーベイなどに製薬企業が出資することはあるのでは。

人材育成

- ▶ 圧倒的に安定なポストがない。大学が用意していない。関連する特任ポストは増えていて、そういうところに引っ張りだこだが、研究費が終わると雇用も終わる、未だに

そういう状況が続いている。まず、安定なポストをつくる必要がある。

- ▶ 情報をバックグラウンドにしている研究者にとっては、バイオインフォという学問は自身の研究成果をアプライする対象としてみており、情報研究の中の専門性によりかわり方も変わる。一方で、ライフサイエンスの中でのバイオインフォ人材というのは、情報関係は何でもできる人という見方になっている。ライフサイエンス研究者からの要求は、情報研究者の目線から見ると何をすればいいのかわからない焦点が絞られていない話が多い。ジョブディスクリプションあるいは研究内容を明確にして、議論する必要がある。
- ▶ 今後、個人ゲノムなどを含めて大規模データがどんどん出てくるのだが、それに対する漠然とした危機感から人材不足の話がでてくるのだと思われる。恐らく我が国が直近で目指さなければいけないのは、既に開発済みのソフトウェアを使いこなしてデータ処理ができる実務的な人材を確保するということだ。すぐにでも人を充足すべきである。
- ▶ いわゆるプロでなくても、中間的な人をいかに養成するかでよいと考える。アメリカで聞いた話しでは、実験系の人たちにドライな研究をやらせている。それで通用している。
- ▶ 私たちが、バイオインフォマティクスの人材を募集するときには、バイオインフォマティクスの専門家に応募してもらえることはなく、情報のバックグラウンドの方あるいは、実験系の方を採用して、それぞれがやっていないものを後から教育するという形でバイオインフォマティクスの仕事を担当してもらっている。恐らくバイオインフォマティシャンという分類に属する方々がポストを得る機会は少なく、それゆえ人材も少ないはず。現状では、やはりバイオの実験系の学生にITを植え込むのが、効率的な良い人材育成法だろう。

3. 総合討論（セッション3）

オリエンテーション

永井 良三 先生（自治医科大学学長、CRDS 特任フェロー）

- ▶ ゲノムデータを活用した研究として、東大の小川誠司先生のがんゲノムデータ研究を紹介する。小川先生は、悪性貧血や骨髄性白血病が非常に高頻度で出てくることで知られている骨髄異形成症候群という、前がん状態の疾患の多くに、様々なスプライシング因子の変異があることを明らかにした。学術的には非常に高い研究であっても、その変異の種類や位置がどのように臨床経過に影響を与えるのか、治療に結びつけるには、多くの臨床例で検討しなければならず、臨床データベースが必要となる（図3-1-1）。
- ▶ 臨床医は、自ら時間をかけてゲノムデータを作成するのではなく、臨床データも含めて、もっといろいろなデータを統合して利用することが重要である。データは、改めて入力するのではなく、日々の仕事の中でデータベース化することが重要である（図3-1-2）。
- ▶ 年間2,000件ほど行う心臓冠動脈のカテーテル検査では、さまざまな画像情報、デバイスの情報、薬の情報、病名の情報、等々を日々「心カテーテルレポートシステム」に落としながらデータベース化している。このレポートシステムソフトを無料でオープンソフトとして、6つの病院で使って、ゲノム情報と合わせたりして、予後調査を進めている。いずれ匿名化したデータを統合して大きな臨床研究をしよう、と考えている（図3-1-3）。
- ▶ データベース化活動の中で、データの標準化が問題になってくる。病名、検査、処方、こういう基本的なことでも実はコードが病院ごとに全く異なっている。そこで、東大の大江先生、山本先生たちと、SS-MIXという標準化を進めている。SS-MIX形式を用いれば、複数施設の臨床研究が容易になり、ゲノム情報と合わせた研究も推進することができる（図3-1-4）。
- ▶ いろいろなデバイスによって、あるいは診療の場所によって診療成績が大きく異なる。多施設の統合データベースがあると、どこに問題があってどの数字を目標にして改良していけばよいかといったことがわかる。こういうところは、実はニーズがたくさんある。
- ▶ ちょっとしたパラメータの数値が高い人と低い人がどう違うのかといったこともデータベースがあると、今は半日とか1日あればだいたい調べることができる。昔なら集めて解析するのに1年2年かかるところで、データベースによる迅速化も通常の臨床活動に非常に役に立っている。
- ▶ データベースの活用例として、糖尿病に網膜症の合併症を既に起こした人を調べてみると、ほとんど症状はなくても重症の冠動脈硬化を持っていた、ということがわかった。ここをターゲットとした治療薬の開発も可能になる（図3-1-5）。
- ▶ また、患者血清の蛋白をプロテオミクス技術で調べることができるようになった。血中のBNPというペプチドの断片（バイオマーカーの1つ）が病態とどう関係してくる

かといったことも非常に迅速にわかるようになった（図3-1-6）。

- ▶ 疫学は疾患の予防にも重要である。左が従業員2万5,000人の自動車販売店、右が8,900人のメーカーで、肥満の度合いと、検診の、リスクなし、低リスク、高リスクの3群から重症疾患（脳卒中、心筋梗塞）を起こした人がどこから出ているか調べた。もちろんリスクは高い方が重症疾患は増えるが、低リスクからの発症もある。痩せている人と太っている人を比べると、同じリスクなら自動車販売店では太っている人の方が重症疾患を起こしやすいが、メーカーでは痩せている人のほうが起こしやすい。現実というのは非常に複雑で、自動車販売店の営業の不規則でストレスを溜める生活では肥満の方が重症疾患は増えるが、メーカーでは体型に気を遣って痩せていても血压管理がしっかりできていないために重症疾患が多かった。いろいろな因子を組み合わせると、健康指導は難しい（図3-1-7）。
- ▶ ゲノムは非常に重要で、データベースの統合活用にこれを欠かすことはできない。希少疾患（rare diseases）と一般的な病気（common diseases）はそれぞれをしっかりと研究すれば良い。ただ、希少疾患の治療薬から一般的な病気の治療薬が出てくることがしばしばあり、希少疾患をしっかりと研究することで新たな薬物や治療法への取りかかりができる（図3-1-8）。希少疾患をしっかりと診ていくためには、ゲノム解析を増やす必要がある。
- ▶ 大規模な医学・医療データベースでの利益は、患者さんに還元しないといけない。患者が自分の医療や健康に関して、自分で判断できるようになることが重要である。ゲノム研究や疫学研究は、国民が自律的に判断できるように支援するという性格がある。この点を、もっと強く打ち出すべきと思う。
- ▶ 医療者にとっても、医療の実践の根拠をデータベースに求めることができる。
- ▶ 医療は、集団では効果があるけれども、個々の患者に対する効果は全くわからない。集団をサブグループ化して、そのサブグループに適切な医療（個別化医療）を明らかにすることが、医療にとって重要である。
- ▶ 動物で明らかにされた知見をヒトでも検証してみる、あるいは基礎研究のテーマを探る上でも新たな課題を抽出する上でも、データベースあるいは疫学が必要だと言える。アメリカでは、このごろは基礎研究でも、テーマを選ぶときに疫学を組み込まないと研究費の申請が通らない、という話を聞いた。アメリカでも、ヒトを意識した医学研究という意味では疫学あるいはデータベースが必要になっている。
- ▶ 企業は、まずアンメットニーズをよく見て開発して欲しい。何か開発してから「どう使いますか」ではなくて、どこに課題があるかを知り、現実を見るという意味で、データベースを活用して考えていただきたい。
- ▶ 新規医療技術の検証は当然だが、実は既存の医療技術にも、怪しいものがある。本当に意味がある医療であるかを検証するためにも、迅速な薬剤開発とか副作用の問題にも、あやふやな情報や怪しい情報への備えにも、研究資源の有効活用のためのエビデンスにも、医療・医学データベースが必要ではないか（図3-1-9）。
- ▶ 我々が理解しているメカニズムはまだほんの一部であって、わからないことがたくさんある。そこに切り込んでいくには、現実を見ながら、そして仮説を立てて研究することだと思う。仮説先導ではなくてデータが先だという時代でもあり、我々がチャレンジするためには相当大規模なデータを集積しないといけない。

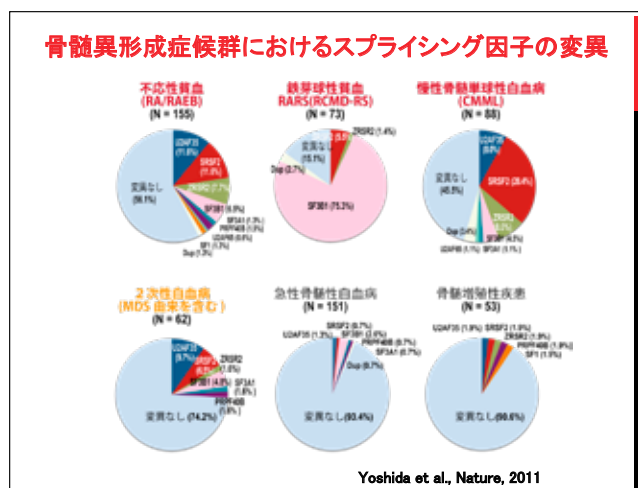


図3-1-1

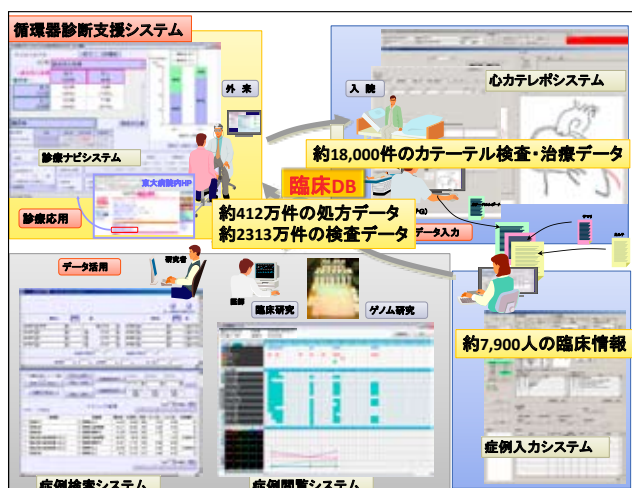


図3-1-2



図3-1-3

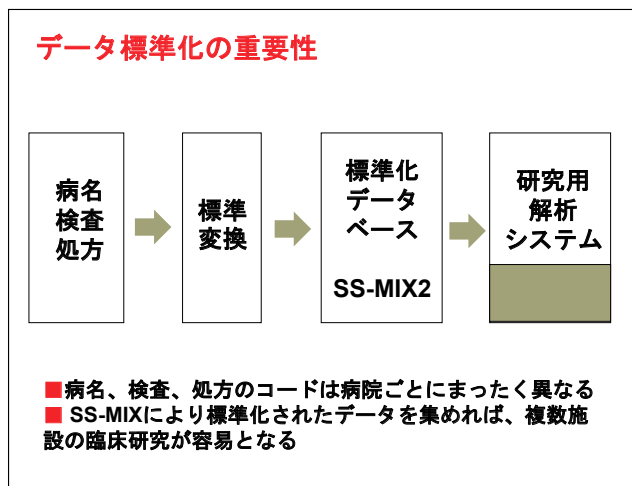


図3-1-4

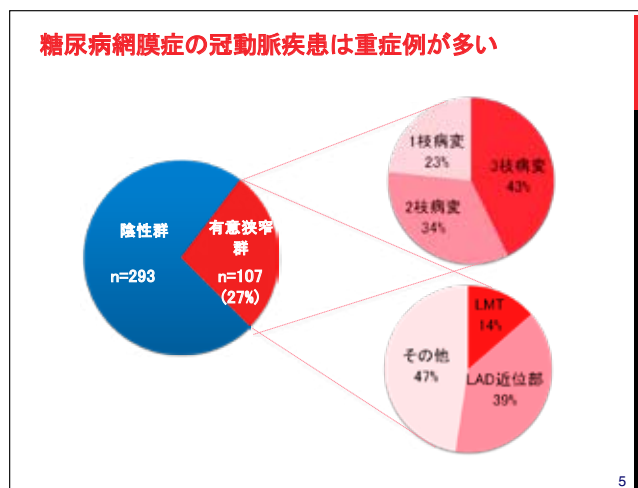


図3-1-5

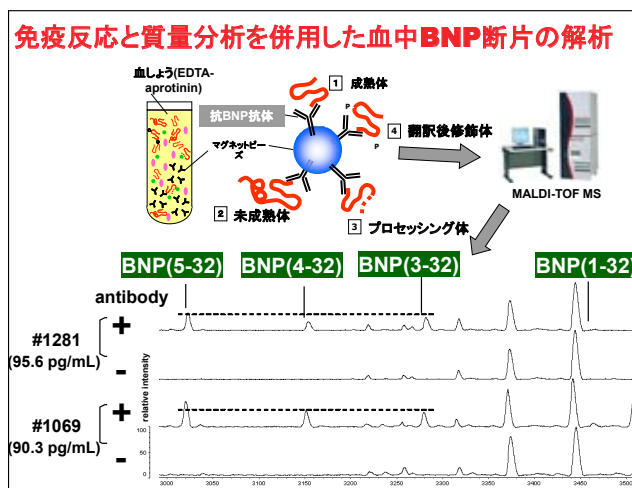


図3-1-6

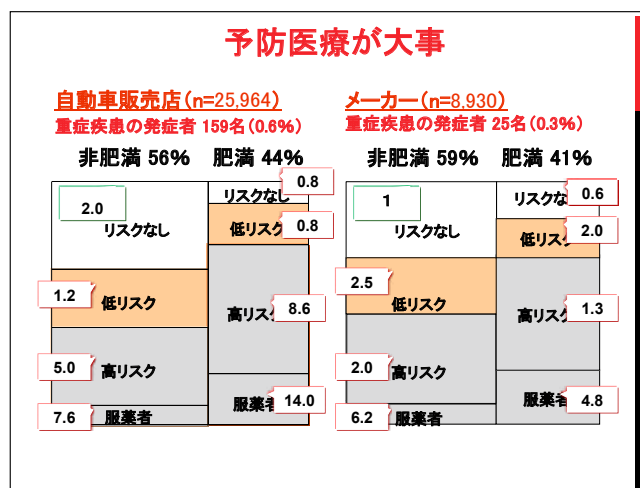


図3-1-7

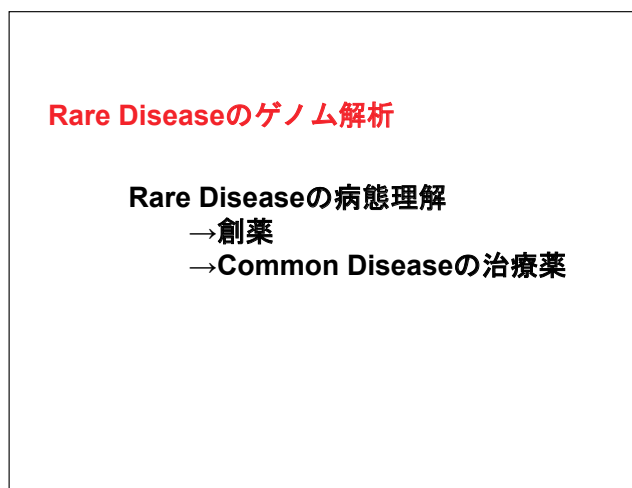


図3-1-8

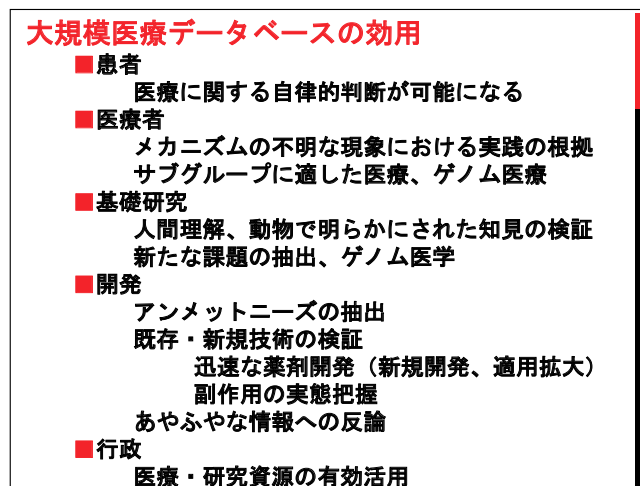


図3-1-9

総合討論

セッション1と2の議論の結果ならびに、オリエンテーションを受けて、総合討論における主な視点などをJST-CRDSが以下のように整理した。主に以下の視点を中心に、参加有識者から意見を頂いた。

- 1) 統合すべきデータベースと、そのインパクトは誰に
 - GWAS 研究、遺伝子診断から医療へインパクトを与える
 - データベースを用いた臨床試験の加速
 - 疫学研究のためのデータベース
- 2) 制度課題
 - 医療等IDについて
 - 医療等の個人情報保護個別法について
 - 包括同意について
- 3) 技術課題
 - データ交換・共有
 - 機械学習
 - 医師と情報処理研究者とデータマネージャーの協働
- 4) どのような成功例を最初にだすのか
 - アンメットメディカルニーズへの対応
 - がんへの対応
 - 希少疾患への対応

1) 統合すべきデータベースと、そのインパクトは誰に

- ▶ GWAS研究、遺伝子診断から医療へインパクトを与える
 - ヒトを対象にした研究の中では、ゲノム研究は非常に重要である。心臓の研究では、マウスやラットの心臓を取り出せるが、ヒトではそれができない。ヒトの疾病研究では、まず血液や尿、そしてゲノムは直接研究ができることが重要である。ただ、多因子性の心臓病をゲノムから見ても、オッズ比では1.2倍から1.4倍程度のリスクである。それでも、ゲノムからヒトを見る研究領域には、多くの研究者の科学的好奇心を引き付ける可能性が秘められている。
 - 例えばハンチントン病は、特定の遺伝子変異の影響を強く受ける常染色体優性遺伝型の神経変性疾患である。一方、糖尿病は生活習慣によるところが大きく、遺伝型の影響は比較的少ない。すべての疾患が遺伝子のみで診断できるわけではない。
 - 医療という視点から遺伝子診断の必要性を考えると、抗がん剤などの重篤な副作用が出やすいかどうかを診断するファーマコゲノミクス（Pharmacogenomics、PGx）やメンデル型遺伝性疾患など、特定遺伝子の変異だけである程度その説明ができる疾患

の診断目的では使えるだろう。糖尿病、高血圧、心筋梗塞、緑内障など複合的要因に起因するような疾患の診断には、ゲノム情報だけでは足りない。しかしながら、そのような疾患に関しても、ゲノムから解決策の探りを入れるアプローチには意味がある。特定のSNP（Single Nucleotide Polymorphism、一塩基多型）と環境要因が重なって疾患が発症するケースも数多く存在すると予想されており、それらを見出そうとする試みも続けられている。医療という面では、遺伝子診断の限界もわかりつつあるが、まだ過渡期であり、今後遺伝子診断の活用が広がる可能性が十分にあると考える。

- ゲノム情報あるいはエピゲノム情報は、臨床情報と結びつけてこそ、その潜在価値を引き出せるという理解で良い。
- ゲノム情報だけで疾患すべてを理解することが出来ないことは明らかになった、しかしながら、ゲノム情報に臨床情報や家系情報を重ねることで、疾患原因に迫れる可能性があり、現在それが追究されている。

▶ データベースを用いた臨床試験の加速

- 患者データベースを活用して、治験を迅速化できる。必要な患者数を迅速に集めることができれば、治験のスタートが早まる。ところが、日本には大規模な患者データベースがない。
- どこにどういう患者さんがいて、どういう頻度で何が起こっているのかがわかるだけで、実は治験は加速される。治験が1年早く終了すると、メーカーにとって利益は大きい。薬剤に対するパテントの有効期間が1年延びることになる。治験の効率化は、画期的なイノベーションではなく、インクリメンタルなイノベーションの部類である。また、治験のスピードアップは、いつか画期的な薬剤の発見につながるだろうと考える。ファーマコゲノミクスの話もあり、臨床データベースとゲノムをセットにして持っていれば治験は大きく加速されるだろう。
- 米国のGARNETというプロジェクトは、ランダム化比較試験（Randomized Controlled Trial、RCT）とゲノム情報をつないだデータベースのネットワーク化という事業として2～3年前に始まった。新薬の申請の際に、ファーマコゲノミクス情報が必要とされるという話も出ており、我が国は対応が遅れるのではないかという気がする。医療でのゲノム利用について、我が国は米国に対し、5年遅れている感覚だ。中国に対しても、技術的なアドバンテージはもうないだろう。それから、中国は我が国よりも人口スケールが1桁大きいので、何を集めるにも中国の方がやりやすい。

▶ 疫学調査のためのデータベース

- オリエンテーションの中で紹介された武田薬品の事例で、糖尿病治療薬が膀胱がんの発症を高める可能性が示されたが、遺伝子の多面作用を示唆しているようで興味深い。糖尿病リスクを高めるけれども、膀胱がんのリスクを下げる遺伝子があり、それに触る薬剤と考えると説明ができる。
- そのようなことがあるかもしれない。ただ、あのデータからは何も言えない。武田薬品は自社調査の結果を早急に出すべきだが、なかなか出てこない。フランスが出してきた大規模なデータを、国内でとることは難しい。そこで海外を使うということになり、対応が遅れる。我が国のレセプトデータは問題が多いが、DPCデータも活用す

れば、大規模追跡調査が可能になるはず。

- ・フランスは、1980年代から病院活動情報化計画（PMSI）を実施しており、既に全国を網羅する医療関連データベースを完成、運用している。それを管理しているのは行政である。
- ・疫学論争で、我が国が意図的にターゲットにされるケースがあったとしても、自国のエビデンスで対抗できるようにする必要があるのではないか。
- ・アイスランドも医療データの電子化が進んでおり、疫学研究や情報の一元化に向けた先進的な取り組みが行われている。

2) 制度課題

▶ 医療等IDについて

- ・医療等IDの導入には、やはりサクセスストーリーが必要だろう。がんゲノム、心臓カテーテルレポートは、データベースとIDが必要とされる例にはなるかと思う。やはり、個別化医療というかサブグループ化に現在の医療が向かっている以上、患者さんを特定できるIDは必要と考える。

▶ 医療等の個人情報保護個別法について

- ・医療等の個人情報保護個別法における罰則規定について。医療やアカデミアの中に刑事罰が入ることへのアレルギーは強い。我が国の医療は余裕がないところで個別に頑張っている、刑事責任に耐えるだけのインフラがない。GCP（Good Clinical Practice）が導入されたときに、治験が一斉に止まった経験がある。これを回復するのに10年ぐらいかかった。だから個別法も、法律をつくった人が予想もつかない反応は起こり得る。
- ・刑事罰というものが後ろに控えていると、どうしてもすくんでしまう。しかしながら、現在の議論を大局的に見ると、刑事罰条項を盛りこまなければ、今の膠着状態から前に進まないと感じている。
- ・いざと言うときに、サポートしてくれる人がいれば随分違うだろうが、現場に人的余裕はない。
- ・各施設、病院ごとでなくとも、刑事罰条項に対する一定のコンサルテーションができる人を地域レベルでも配置してもらえばよい。
- ・少し誤解があるかもしれない。個別法で議論されていることは、情報を提供する側は一定の義務を果たせば免責しようということだ。何か事故があったら謝るのは病院長だという考え方があるが、一定の配慮がされていればそれは免責しようという議論をしている。「重大な過失」に対して刑事罰を科するという項目は、盛り込まないという議論も続いている。
- ・厚労省の合同会議が終了し、その後、個別法に関する審議は、社会保障審議会に部会がつくられて法案を検討することになっているが、まだスタートしていない。医療等IDには三師会が反対しているので、その説得を先にするという手順である。

▶ 包括同意について

- ・新しい分析機器が医療に次々と持ち込まれ、疾病解析の切り込み方がより先鋭化する

現状では、医療情報が後から追加されることを予め想定し、できるだけ広く同意を取ったサンプルを集めるという手法が主流になりつつあると思う。包括同意も重要な論点である。

- 包括同意には、倫理面からの議論が必要だ。
 - 同意の問題は、法的な議論が続いている。基本的には、利用目的等を十分説明した上でかつ、本人の自由意思による同意である。自由意思かどうかは重要である。例えば上司とか企業からの要請でDNAを出すことは、半強制的な同意と考えられ、問題が残る。包括的な事前同意をとったとしても、その内容が曖昧だと、どこまで使われるのか本人に予測がつきにくいので、できるだけ利用目的を具体的に明示する必要がある。その具体性の水準が定まっていない。提供した情報やサンプルが、漏洩しないで安全に管理されていればいいが、漏洩は大きな問題になる。管理の安全性についても担保された上での同意であり、管理についても水準を定める必要がある。
- ▶ 遺伝子例外主義について
- 法律の世界では、遺伝子情報の取り扱いにいろいろ議論がある。アメリカでは、遺伝子情報と他の医療情報との比較で、「遺伝子例外主義」といって、遺伝子情報を非常に厳格に保護すべきだと話が出る。しかしながら、一方で遺伝子は特別ではないとの意見もあり、未だに対立している。遺伝子については、議論になりやすい対象であるというところは、踏まえておいたほうがいい。
- ▶ プライバシー権について
- 個別法化での医療情報の所有者はだれになるのか。
 - 所有権というのは民法上の権利で、形のあるものに対する権利だ。プライバシー情報とか個人情報というのは情報なので、基本的に、所有権というのは概念的に使えない。ただ、プライバシー権とか個人情報保護法制の中でどういうふうに取り扱っていくかについては、基本的には患者の情報であって、患者に一定のコントロール権とか関与権が認められるというのがオーソドックスな考え方。
- ▶ 情報ネットワークの必要性と国家戦略
- 我々が、個々の病院の中で患者さんとの関係や体制をつくるよりも、全国的な医療機関、医育機関等がネットワーク化され、標準化されたプロトコルで動かせる体制が必要だと思っている。個人ゲノムシーケンスのステージになって、初めてつくるというのも1つの考え方だが、なかなか生易しいものではないという危機感がある。
 - 国家戦略としてのゲノム研究はどのような形が理想的か。分散型にするのか、拠点型が望ましいのか。
 - できれば数カ所ゲノム研究拠点を育てるべき。
 - テーマは国やファンディングエージェンシーが介入して、「あなた方はこれをやりなさい」というトップダウン型でよいのか。
 - 今の段階では、ある程度トップダウンでないと、ゲノム分野での国際競争に生き残れない。一気呵成で個人ゲノムデータを揃え、基盤を固めるというのが1つの方策だと思う。

3) 技術課題

▶ データ交換・共有

- 電子カルテの自由記載部分を標準的な形式で出すというのは、今はできない。SS-MIXをやや拡張したところまで標準化がすすんでいる。次は画像と波形を加える段階である。そこまでくると、情報量で言えば、臨床で生じる情報の9割以上はカバーできる。センチネルジャパンプロジェクトでもこのレベルの情報であれば十分出せるだろう。
- これからいろいろな病院で、SS-MIXのデータを蓄積し、情報交換できるようにしてほしい。
- 厚労省の補助金も、SS-MIXによる情報交換が条件になっていて、既に200病院以上が対応しており、その数も増大している。問題は、そこからデータベース化する段階である。要するに、病院として情報を管理する分には全く問題ないが、その情報を病院から出して、集めて大規模データベース化するというのは、患者さんに対する責任感とか現行個人情報保護法の問題で、なかなか抵抗が多い。やはり制度的な手当等、あるいは政策的な誘導が必要ではないかと思う。

▶ 機械学習

- 昨年Googleが発表した人工知能関係の機械学習成果では、画像をランダムにコンピューターに読み込ませて、機械が自身でネコの認識に成功した。医療情報のようなテキストの自動構造化や、画像情報の解析に使える技術だろう。
- たくさんのデータを得るときに、病気の形が分けられるなら分けたほうがいい。それは医師の非常に専門的な知識が要る。しかし、専門化、細分化して病名を増やすと、結局抜けが多くなる。そこをITで置き換えることはできないか。
- 幾らITが進歩しても、そこはまだ医師の領分だと考える。ただ、カルテの書き方からサマリーの書き方まで教育が進んでくると、ITで置き換えうる部分が出てくると考える。
- 診断の難易度と論文掲載は相関している。例えば皮膚疾患で視覚的に病気の切り分けが容易なものは、当該フェノタイプに対しての論文が最初に出てくる。それに続き、医療技術として要求の高い診断が必要な疾患の論文が続くことが一般的である。難しい診断に関する情報は、機械学習できるほど数が集まらない。
- ゲノムの解析を考えれば、カルテに遺伝性を記載した方がよい。例えば、「両親はどうでしたか」とか「お祖父さんお祖母さんは同じような病気を持っていましたか」ということが書いてあるか、書いていないかだけで解析時の検出率が全然違うはずだ。そういうところまでのトレーニングなりバックアップ体制が必要だと考える。

▶ 医師と情報処理研究者とデータマネージャーの協働

- 病気を対象とした情報処理に関しては、医療機関でかつデータマネージャーと情報処理研究者がいるような施設でないと難しい。医師でないと、病気やそのフェノタイプ解析が困難である。現場の臨床医療者とデータマネージャーと情報処理研究者の協働が必要。
- 情報工学の専門家だけでは限界があると思う。どこに臨床的問題があるのか、長年の

経験に基づく交通整理が必要なので、解析者と臨床的なスキルを持った人が二人三脚で情報交換しながら、医学、医療への展開を目指した解析を進めるべき。

- 情報工学の専門家と医師が融合できる場所が1つの拠点となる。もう1つの拠点はシーケンシングセンター。シーケンシングセンターは、コストパフォーマンスも考慮し、アウトソーシングも検討するべき。
- データマネージャーが重要である。集めているデータ、フェノタイプが本当に正しいのか裏をとるとか、抜けているところがないか、そこを担当する専任者が必要である。現場の生情報は洗い直しをしないと使えない。
- データマネージャーは他の医療機関との情報連携も担当する。情報連携にはデータの標準化と、担当者の教育が必要となる。
- 治験では実際そういう体制になっている。ゲノム解析や一般臨床研究でもそこまでのレベルにする。今、日本から出る臨床研究は非常に質がいいと言われているので、ちょっとした支援があれば十分に対応できると思われる。

4) どのような成功例を最初にだすのか

- ▶ アンメットメディカルニーズへの対応
 - ゲノム研究がイノベーションにつながる場合は、インパクトの大きい遺伝子に限られるであろう。実際には、非常にインパクトが小さい、オッズ比が低い遺伝子が多数を占めている。
 - アンメットメディカルニーズをきちっと探って、それに基づいたゲノム研究をすべきだ。我が国は高齢化が進んでいるし、糖尿病が大きな問題となっている。そうすると、糖尿病の合併症を起こしやすいゲノムというのはどういうものかという研究を、我が国で展開することが必要だ。現場からの課題に応じたゲノム解析が必要ということだ。
 - もちろん国内のアンメットメディカルニーズについてきちんと対応すると同時に、世界的な動きの中でも我が国の存在感を高める必要があると考える。
- ▶ がんへの対応
 - がん情報はデータベース化するいい対象だろう。がんは結構ばらばらな診断、治療がなされている。まず、データベース化して基本的なところで治療の均一化を図るべき。その上で、個別化や最適なクリニカルパスが計画されるようになればと考える。
 - がんの場合は、がん登録が法律で決められているのでデータを集めやすいが、他の疾患についてはそれだけの強制力はない。そこを今後どうするか。特に厚労省の挙げている4疾病5事業についてはデータベース化を進めるべきだ。それから、100以上の特定疾患が指定されている。そこででの成果をエビデンスベースで検証するため、データベースが重要だろう。「健康日本21」のような国家事業的なものにも、データベース化と、それを利用した検証作業が必要だ。
- ▶ 希少疾患への対応
 - 成功例として一番わかりやすいのは、希少疾患だと考える。IDや医療情報を用いた何らかの成功例が出ればよい。例えば、小児疾患であったり、いろいろな代謝異常症は極めてわかりやすい。その領域では、遺伝子変異との相関が強く、オッズ比が10

倍とか20倍になるような疾患群が見出しやすいと考える。IDを導入した後に、フェノタイプとジェノタイプの統合を進めることが重要。フェノタイプをデータベース化するためには、かなりの数のデータマネージャー、情報研究者が必要とされる。

- 個別の患者さんのデータを個票としてデータベースに出すことに対して、少し抵抗がある。特に難病の場合、数が少ないので個人の特定につながるから、セキュリティがきちんとしていないところには出したいくないとの主張がある。
 - セキュリティが心配だから情報を出さないという話は多い。だから、1カ所でもいいからセキュリティ管理が十分な解析サイトを作る必要がある。そこで成功例をつくった上で広げていくということが、多分、サクセスストーリーになるだろう。
 - JSTとしては、どういうところを推進したいのか。とにかく成功例を出さないといけない。漠然と進めるよりも、「これは出るだろう」みたいなものをまず提示することが重要だ。
 - もちろん法整備も加速すべき課題であり、研究開発投資の前提条件となる。CRDSとしては研究投資を主眼に置いており、法整備を前提として研究投資先を探すのであれば、がん、希少疾患などを中心にデータを集め、そこをターゲットにした研究投資のプロトコルを組むということが方針の1つにはなるだろう。
- ▶ コモンディジーズを含むターゲットポートフォリオが組めないか
- もっとコモンディジーズへの研究投資も考えるべき。
 - 臨床のデータベースをつくることで、我が国の国民の健康が増進し、産業力が出て来るというポートフォリオを上手く組むべき。
 - 突破口、いわゆる先陣を切って目に見えるようなプロジェクトと、それほど派手さはないけれども、底上げして、あるいは第2陣として進めていくようなプロジェクトとをうまく組み合わせた構えを作る必要がある。確かに、がんゲノムとか希少疾患は、取り組む対象としてはよい。しかしながら、それは頻度の高い病気ではない、あくまでもレアな部類である。だからコモンディジーズとしては、4疾病5事業がまずは対象となる。それらのデータベースをしっかりとつくり、そこにゲノムが入ることでさらに研究が加速される体制づくりが必要である。4疾病関連は、患者数が桁違いに多い。世界で何十億人もいるため、マーケットは巨大である。そこに欧米の製薬会社が目をつけており、現在利益を上げている。次の段階としては、そこをさらに細分化、サブグループ化、できれば個別化というところに持ち込んで治療する、そのためにはデータベースが必要になる。それはゲノムの研究推進にも役に立つはずだ、そんなロジックを想定する。
- ▶ 遺伝子診断ビジネスは成功例になりうるか
- 遺伝子診断ビジネスが既に始まっている。大手の健康食品会社は、肥満関係の遺伝子診断をして、あなたにはどういうダイエットキットが適していますよとアドバイスしながら商品売っている。中国の遺伝子診断メーカーが性格とか能力とか、そういうことも含めてどういう可能性があるのか、易学に近いようなことを遺伝子検査結果から情報提供している。競争馬の距離適性を見るような遺伝子も調べられていて、その情報をブリーディングに役立てるといったことも既に行われている。ビジネスとして

の遺伝子診断は、データベース利用の1つの最終形だとも言える。

- 米国の23andMe社では、顧客が自分の頬の細胞を郵便で送って、その解析を請け負っている。つまり、顧客が自ら「調べてくれ」と依頼しているわけだから、インフォームドコンセント実施義務はその時点でクリアされている。DNAチップを用いたSNP解析結果を顧客に返すが、23andMe側でそのデータを蓄積しており、ラーニングデータベースという手法で、新しいものが加わってきたときにデータベースの知識レベルを向上させ続けている。だから顧客に対する最初の情報提供の後に、「今度こういうことを調べたいから新しく情報をくれ」という顧客の新しい要求に答えることもできる。2011年6月でユーザー数は10万人を超え、ゲノム情報がどんどん蓄積している。米国でこのような遺伝子診断のDTC（Direct to Consumer）形式が容認されている理由は、公的な研究目的でボランティアを集めようと思っても個人情報保護法の問題がありなかなか進まないため、DTCを利用してゲノム情報収集を民間に肩代わりさせているのではとも考えられている。つまり、米国では完全に資本主義化している遺伝子診断であるが、そこで蓄積したデータを医療へ応用したいというのが米国政府の本音であると思われる。中国も、ゲノム情報収集を民間企業BGIに任せている。BGIは、コンプリートジェノミクス社という米国の有名な受託ゲノム解析会社を買収の話を持ちかけているが、米国政府はそれに「ノー」と言う可能性もある。つまり、米国としては次世代医療につながる自国のゲノム研究の展開に、海外資本を持ち込みたくないわけである。
- 国内の肥満遺伝子検査は、数十万人が既に受けているだろう。我が国でも、遺伝子と疾患に対するリテラシーは確かに上がっているだろう。実施する側も、エビデンスベースで慎重に取り組んでいる。一方で、才能遺伝子とか子供の才能にかかわるところに関しては、倫理を強く意識する必要がある。例えば、韓国では科学的立証が不確実な外見や性格に関する遺伝子診断を禁止したうえで、診断を禁止・制限する遺伝子をブラックリスト方式で具体的にリストアップしている。
- いろいろなところで遺伝子診断、遺伝子を使った情報ビジネスが広がっていった状況を見ると、我が国でも、より一般市民にとって理解しやすいような形でデータが出てこないかと期待する。
- 個人遺伝情報取扱協議会（CPIGI）で、遺伝子診断に関する企業活動のガイドラインを議論している。
- 我が国の現行制度であれば、生死にかかわる遺伝子の診断は、企業活動としては請け負えないだろう。米国で23andMe社が持て囃されている理由の1つは、自分の家系に近い人、全然知らなかったけれども3代4代遡ったときの近親者がゲノム解析で見つけられて、希望があればその近親者のFacebookアカウントを紹介しますというサービスにある。23andMeにはGoogleも出資しており、SNSビジネスともうまく組み合わせている。
- 血縁関係がでてくると、養子縁組が大きな問題となる可能性がある。「絶対に誰から養子を貰ったか秘密にしてほしい」と思っている、その子が遺伝子診断すると血縁関係が明らかになる。これは社会的な問題だ。
- ゲノムデータの活用からは、さまざまなアウトプットがでるのではないかと。
- 多数の遺伝子により決定されているフェノタイプの解析、診断は困難だ。酒が飲める、

飲めないは1つの遺伝子に起因するが、その程度のことしか言えないのでは、遺伝子診断の成功例とはいえない。

5) その他

- ▶ 国際競争、どうすれば米国や中国に比肩できるのか
 - 遺伝子を治療に結びつける研究は、米国が先導している。その中でスタンダードを、自分たちのいいように決めている。その中で、東アジア人のスタンダードは中国人である。中国をスタンダードだと言われても、それが我が国に当てはまるかは別問題。我が国の遺伝子研究のポジションを確保する必要があると考える。
 - 中国の戦略は非常に単純で、受託シーケンシングビジネス一点集中主義から広げようとしている。「中国で一番最初にやった」と、パンダゲノム決定のようなアドバルーンをどんどん上げて、最終的にBGIのシーケンスビジネスが潤うようにし、そこから広げて、いつの間にかその周辺を含めた大きな分野の大御所というか、オーソリティ的になって、他国を飲み込もうとしている。結構影響力の高い戦略である。欧米も中国に視点が向いており、コーケリアンゲノムに対する東アジア人ゲノムのリファレンスは、多分中国人ゲノムになると言われている。
 - 臨床情報のきめ細やかさは日本のほうがはるかに進んでいる。
 - 臨床試験のスピードアップが重要だと考える。あるメーカーに、私が思いついた適応拡大のアイデア伝えてから、2年半で治験が終わった。これは早い事例であり、やればできる。我が国でも、治験なり臨床研究の迅速化は可能である。質のよいシーズを出し、治験の迅速化を進めれば、画期的なものが出てくる。
 - 米国で、バイオインフォマティクス研究から、卵巣がんの薬が肉腫に適用できるのではないかという指摘があり、すぐ治験が始まった。米国は、患者データベース、ゲノムデータベース、それと薬のドッキング解析など、情報を駆使した、早い動きを見せる。
 - 米国は、医療が市場原理で動いている。必要があれば、マーケットからお金が流れてくる。我が国は、国民皆保険を維持するために市場原理は排除している。したがって米国と同じような市場原理によるスピードアップはできない。我が国独自のアプローチで、スピードを上げる必要がある。
 - 米国と我が国では、がんの発症パターンが随分違う。我が国独自のアプローチの中に、少なくともナショナルデータベースが含まれるはずである。
 - 米国は、ゲノムをビジネスととらえている。ヒトゲノム計画の波及効果で、米国は50兆円以上儲けたと言われている。シーケンサーを販売する会社の売り上げ、そこで雇用されている人の人件費とかを合算すると、のべ50兆円以上になる。創薬におけるファーマコゲノミクスの話も、医療レベルを上げるために必要と言うが、それと同時に、ゲノムビジネスの計算差分がそこにある。
 - 米国の学会等で、ホールゲノムの話は日常的に出てくる。これは当然メーカーが付いて、シーケンスビジネスが展開するということだろう。
 - 米国の小児病院などで、特に希少疾患を早期に診断する手法として、全ゲノムシーケンシングを試験的に始めているところが幾つかある。実際にそれで確定診断がついて早期から外科的治療に入れた例も出ている。将来的にナノポアシーケンサーが出てきて、1回のヒトゲノムシーケンシングが1万円以下になれば、診断のためのホールゲ

ノムシーケンシングも費用対効果で評価できるようになる。

- そういう時代に備えて、病院の中で患者さんのインフォームドコンセントのとり方など、今から考えておいたほうがよい。

4. まとめ

セッション1において、「どのデータベースをどういう形で使って、誰にインパクトを与えるのか」を主題として、話題提供、議論を進めた。本セッションでは、まず、ライフサイエンス・臨床医学領域でのデータベースの必要性の高さが認識共有された。本領域の研究では、理論だけでは予測しえない事象（例えば薬の副作用）があり、実践データを解析し、さらに研究を進化させるデータサイエンスが重要である。データサイエンスを進めるにあたっては、精度を高めるためにデータベースの規模拡大を図らなければならない。また、複数のデータベースを統合的に活用する必要性も提言された。主要な論点、提言を以下に列挙する。

- 臨床系のデータベースは大規模化が進展しており、統合活用にむけた大きな課題は個人識別番号（医療等ID）の導入
- 基礎系の国内データベースも統合化、大規模化を図っているが、データ規格標準化が課題である
- レセプトデータベース（NRDB）の公益目的での二次活用研究が進められているが、その中で推進上の課題も明らかになっている、安全な解析環境（オンサイト解析センター）も必要
- 病院内あるいは地域連携レベルで臨床関連情報が統合的に解析されている例としては、MIHARIプロジェクト成果があり、薬剤の副作用追跡など公益性の高い研究成果が得られている
- 静岡県民へのアンケート調査結果からは、生涯カルテに賛同する意見が7割であり、個人は概して生涯カルテに対してその効果を期待している
- データベースを使った研究成果は国民にフィードバックされるべきである
- 個人の生涯治療記録が個人にとって有意義であると同時に、その二次活用が公益（治療の質の向上など）に結びつく、個人生涯治療記録データベースは重要である
- ゲノム研究の進展により、個人ゲノム情報データベース構築が現実味を帯びている
- 個人ゲノム情報、レセプト、病院DPCや特定健診のデータベースを統合解析することにより、ゲノム情報を診断ツールとして活用できる可能性が高い

セッション2において、「技術課題、制度課題」の洗い出しを行った。まず、話題提供として、企業、個人情報保護法、情報工学、ゲノム情報と臨床情報の統合活用、それぞれの視点からの課題が提示された。討論では、法制度、インセンティブ、国家戦略、人材、社会基盤、データ、技術という項目を重要議題として、議論を進めた。セッション2の発表と討論を通して議論の中心となり、特に重要と考えられた課題は以下の通りである。

- 医療分野において個人情報保護法個別法制定が必要
- 上記個別法の中で個人IDとして医療等IDを制定する
- データ提供インセンティブ（個人・社会への還元、ビジネスフレームワーク）
- オンサイト解析センターの設置と専任の情報安全管理者によるセキュリティ確保
- 長期的にデータを保管するアーカイブセンター
- バイオインフォマティクス研究者の恒久的ポストを大学に拡充

- バイオインフォマティクス研究者のジョブ定義を精査して育成
- 国家戦略として欧米中韓との差別化
- 匿名化技術
- 分散データベース
- 国内開発にこだわらず海外技術の導入も選択肢に組込む
- データ規格の標準化
- データ品質の管理

セッション3では、法制度整備を含めた広い視野から、データベースの統合的活用に向けた研究開発戦略の骨子となるべき部分を議論した。オリエンテーションにおいては、いくつかの臨床系データベースの例示があり、それらの重要性と、治療へと結びつけるための臨床研究の紹介があった。大規模医療データベースの効用として、患者が自律判断のための情報を得られるようになること、医療者へは医療実践の根拠を与えることなど指摘された。現在は、仮説先導ではなくてデータが先だという時代でもあり、相当大規模なデータを集積して、新しい切り口を見つけていく時代であるとの提言があった。総合討論で議論を進め、共通認識が得られた課題と対策を以下にまとめた。

- データベースを統合的に活用することで、国民、医療、行政などと幅広い対象にインパクトを与えることができる
- 個人のプライバシーを保護しつつ、研究開発成果を国民生活に還元するという理念が必要
- 統合すべきデータベースは、臨床系、基礎科学系、ゲノム情報にまたがる
- 医療関連の個人情報利用には、個人情報保護法の個別法が必要
- セキュアな情報ネットワーク（規格化、標準化、共有化を含む）が必要
- 機械学習などの最新情報処理技術の導入が必要
- 医師と情報処理研究者とデータマネージャーの協働が必須
- 我が国のデータベース戦略を国家戦略として立案し、国際競争に備えることが急務

JST-CRDSでは、本ワークショップでの議論を受け、更に調査活動をつづけ、データベース活用を促進するための研究開発提言作成を行う予定である。本ワークショップで主要議題となった制度課題については、先の理念も踏まえた上で提言の一部として組み入れ、総合的な研究戦略提案の企画、立案を目指す（平成24年度末に発刊し、JST-CRDSホームページにも掲載予定）。

5. 参考資料

① ワークショッププログラム

開催日：2012年10月31日（水）10：30－17：00

場 所：JST東京別館A②会議室

趣旨説明 10：30－11：00 森 英郎 フェロー CRDS

セッション1 11：00－12：45 「データベースの活用目的」

話題提供 11：00－12：00 各発表10分、質疑5分

永井 良三 先生（自治医科大学 学長、CRDS 特任フェロー） 医療・医学の視点から

高木 利久 先生（東京大学 教授、NBDC 副センター長） 基礎生物学の視点から

山本 隆一 先生（東京大学 准教授） 解析の視点から

木村 通男 先生（浜松医科大学 教授） 標準化・統合的活用の視点から

討 論 12：00－12：45

休 憩 12：45－13：00

CRDS挨拶 13：00－13：05 浅島 誠 CRDS 上席フェロー

セッション2 13：05－14：50

「統合的データベース活用に向けた技術課題と制度課題」

話題提供 13：05－14：00 各発表10分、質疑5分

佐藤 恵一 先生（日立ソリューションズ グループマネージャー） 企業の視点からの課題

村上康二郎 先生（東京工科大学 准教授） 個人情報保護法を中心とした制度課題

金 進東 先生（DBCLS 准教授） バイオデータ構造化の課題

加藤 規弘 先生（国立国際医療研究センター研究所 部長） ゲノム疫学における課題

討 論 14：00－14：50

休 憩 14：50－15：10

セッション3 総合討論 15：10－16：50

オリエンテーション 15：10－15：20 永井 良三 先生

討 論 15：20－16：50 ファシリテーター 永井 良三 先生

閉会挨拶 16：50－17：00 永井 良三 先生

② ワークショップ参加者リスト

本ワークショップの参加者を以下に列挙した（50音順、敬称略、役職は当時）

【有識者】

稲葉 一人	中京大学法科大学院 教授
小原 雄治	国立遺伝学研究所 所長
加藤 規弘	国立国際医療研究センター研究所 遺伝子診断治療開発研究部長
金 進東	ライフサイエンス統合データベースセンター 特任准教授
木村 通男	浜松医科大学 教授、付属病院 医療情報部長
興梠 貴英	東京大学医学部付属病院 特任助教
小西 宏	日本対がん協会 広報担当マネージャー
佐藤 恵一	日立ソリューションズ HB事業推進部グループマネージャ
高木 利久	東京大学大学院新領域創成科学研究科 教授、NBDC 副センター長
永井 良三	自治医科大学 学長、CRDS 特任フェロー
村上康二郎	東京工科大学教養学環 准教授
山本 隆一	東京大学大学院情報学環 准教授

【府省関係者傍聴】

文部科学省 ライフサイエンス課

【JST-CRDS 統合データベースチーム】

浅島 誠	ライフサイエンス・臨床医学ユニット	上席フェロー
川口 哲	ライフサイエンス・臨床医学ユニット	フェロー
鈴木 響子	ライフサイエンス・臨床医学ユニット	フェロー
辻 真博	ライフサイエンス・臨床医学ユニット	フェロー
茂木 強	電子情報通信ユニット	フェロー
森 英郎	ライフサイエンス・臨床医学ユニット	フェロー
波羅 仁	戦略研究推進部	主査

その他、JSTの関連部署より数名程度の傍聴参加があった。

■ワークショップ報告書作成メンバー■

浅島 誠	上席フェロー	(ライフサイエンス・臨床医学ユニット)
森 英郎	フェロー	(ライフサイエンス・臨床医学ユニット)
川口 哲	フェロー	(ライフサイエンス・臨床医学ユニット)
鈴木 響子	フェロー	(ライフサイエンス・臨床医学ユニット)
辻 真博	フェロー	(ライフサイエンス・臨床医学ユニット)
茂木 強	フェロー	(電子情報通信ユニット)
波羅 仁	主査	(戦略研究推進部)

※お問い合わせ等は下記ユニットまでお願いいたします。

CRDS-FY2012-WR-16

科学技術未来戦略ワークショップ報告書

ライフサイエンス・臨床医学分野における データベースの統合的活用戦略

平成 25 年 3 月 March 2013

独立行政法人科学技術振興機構 研究開発戦略センター
ライフサイエンス・臨床医学ユニット
Life Science / Clinical Research Unit, Center for Research and
Development Strategy
Japan Science and Technology Agency

〒 102-0076 東京都千代田区五番町 7 K's 五番町
電 話 03-5214-7481 (代表)
ファックス 03-5214-7385
<http://crds.jst.go.jp>
©2012 JST/CRDS

許可なく複写・複製することを禁じます。
引用を行う際は、必ず出典を記述願います。
No part of this publication may be reproduced, copied, transmitted or translated
without written permission.
Application should be sent to crds@jst.go.jp. Any quotations must be appropriately
acknowledged.
