

基盤の整備に関する具体的研究開発課題としては、代表的にはクラウドコンピューティング技術、大規模分散ファイルシステム技術、ストリーミング処理技術がある。また、次世代シーケンサの登場により、高速ゲノム解析が可能となり、個人のゲノム解析を実施して予防や治療に役立てる状況（パーソナルゲノム時代）が想定され、これに備え大量のデータを高速に解析し、可視化するためのインフォマティクス技術の開発も課題となる。

THIBの構築を具体的に推進するにあたっては、疾病を決め、それに関するデータを複数の病院でまず試験的に収集し、データ収集上の課題などをクリアにした上で、できるだけ多くの病院からのデータを収集する仕組みを構築する。あるいは、非常にまれな疾病に関しては、専門医のネットワークを構築し、個別データを収集する仕組みを構築するといったように、拠点をいくつか決めスモールプロジェクトでスタートし、インフラ構築を進めつつネットワークを拡げていくアプローチとする。さらに、標準化や共用ツール開発などのインフラ構築部分への投資や、恒久的運用に関する仕組みづくりの検討を実施する。

こうした基礎となるデータを収集し利用可能とするTHIBを構築することにより、基礎医学の進展、疫学の進展、および医療の高度化（個別化医療、医療経済における効率化、医療プロセスの高品質化）といった効果が期待できる。これらの効果を現実のものとするためには、医療制度改革など困難な課題も存在するが、まずは情報基盤を整えつつ、徐々にデータを収集することによって医療制度改革へのエビデンスとなり、より良い施策が実施され、健康で安心な社会、医学分野での技術立国という目標に大きく前進する。

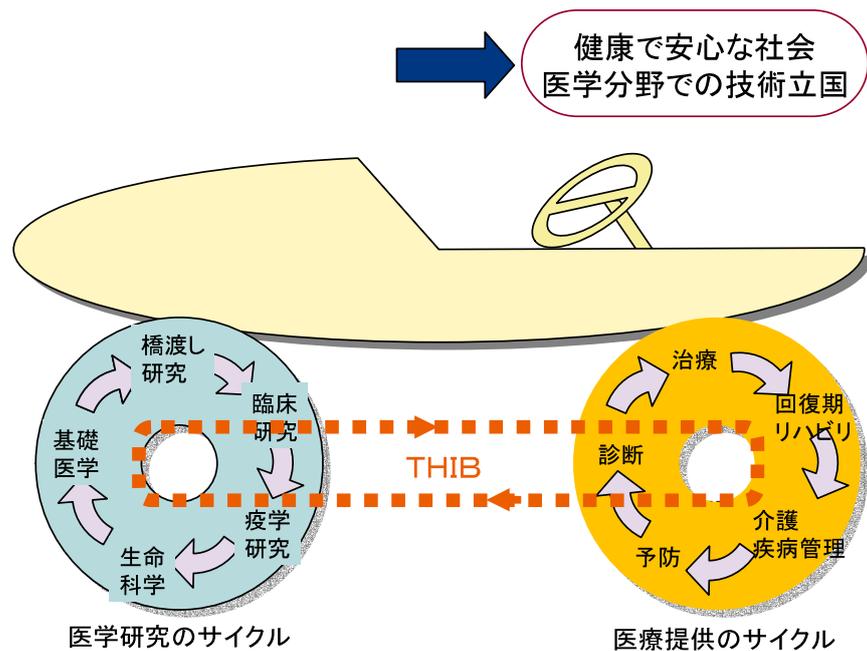


図 1. 概念図

## 2. THIBの構築と活用に投資する意義

THIBの構築と活用に投資する意義は、大きく以下の4つの点がある。

(1) 情報基盤の整備による医学研究・医療提供相互の有機的連携の強化  
 国の科学技術情報基盤の整備を進める足がかりとして、情報基盤であるTHIBを国として整備することで、医療・健康分野の研究に必要な主要なデータベースおよびそれらを連携したデータベースが利用可能となり、医学研究・医療提供相互の有機的連携の強化につながる。

(2) 医療・健康産業の国際競争力の強化

21世紀はグローバルに見ても、医療・健康産業が産業基盤として重要な地位を占めることとなる。しかし、わが国における医療・健康産業は産業基盤として強固であるとは言いがたい。例えば、個々の遺伝子の多様性をもとに、わが国の国民に対して最適な医療を提供するための医学研究や創薬、新規医療の研究を進展させることによって、医薬品や医療機器などの自給率を向上させ、わが国の医療・健康産業を国際競争力のある産業として伸ばしていくといったことが考えられる。また、こうした競争力の強化によって、さらには医療を通じた国際貢献にもつながることが期待される。

(3) 超高齢社会への対応（医療・健康問題解決）で世界をリード

世界に先駆けて超高齢社会に突入している我が国では、健康で安心な社会へのニーズが極めて強くなっていることから、医療の高度化や医療の質の向上に向けて、研究開発成果を社会に実装しその効果をいち早く試す必要があり、かつこれによって、こうした社会作りにおける課題解決型研究で世界をリードすることが可能となる。

(4) 医療の質の向上、医療経済学的効果

THIBにより、医療プロセスの各時点でのデータを大規模に収集し、そのアウトカムと比較することによって、医療プロセスの良否の検討が可能となる。こうした過程のなかで、標準的な医療プロセスの構築やアウトカム指標の策定が行なわれ、医療全体の底上げが可能となり、医療の質の向上に繋がるとともに医療経済の効率化といった効果も期待できる。



### 3. 具体的な研究開発課題

1章で示した概念図（図1）に点線のチェーンとして表現されているTHIBの実現イメージを図2に示す。

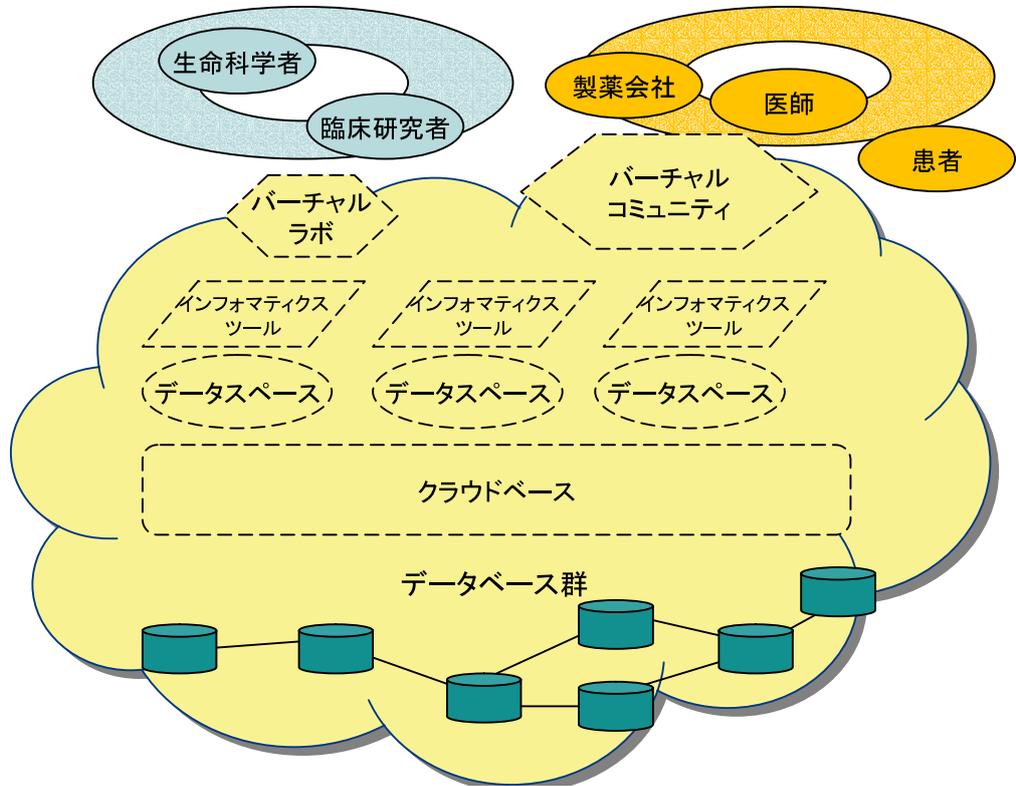


図2 THIB構成概要図

図中点線で囲まれた部分がTHIBの研究開発要素であり、インフラ整備、ツール間連携およびデータ標準化、そしてインフォマティクスツールの3つに大別される。

既存のデータベース群は、ネットワークで接続され、クラウドベースと呼ばれるミドルウェアにより連携可能となる。生命学者、臨床研究者、製薬会社、医師といった利用者は、データベースの位置などを意識することなく、ツールを利用し研究活動などを行う。

インフラ整備の観点から必要となるものとして、システムアーキテクチャ、クラウドベース、データスペース、仮想ラボ、仮想コミュニティがある。さらにツール間の連携やデータ標準化のための開発要素も存在する。こうしたインフラ整備は主に既存技術の集積によってなされるが、システムアーキテクチャなど新規研究要素として重要なものを含み、また、インフラを整備・運用していく中で新たな研究課題が生じてくる。

先端的な研究開発要素としては、3.3で述べるインフォマティクスツールがある。

### 3.1 インフラ整備に関する技術開発課題

#### (1) システムアーキテクチャ

技術の進歩に耐えうる柔軟な設計が求められ、規模の拡大を考慮した最初のアーキテクチャ設計が非常に重要となる。ここでは、システム間のあまり緊密な連携を想定せず、ゆるやかな連携が可能となるような設計（フェデレーション技術）が適切である。

クラウドコンピューティングの概念をベースとしたシステムアーキテクチャを検討する。クラウドに着目する理由は、以下のものがある。

- ① 研究者にとっては本来データや計算機の所有ではなくその利用が主たる目的であり、「所有から利用へ」というクラウドの概念にそったものである。
- ② 計算機資源の有効利用、スペース的な課題、電力消費などのエネルギー的課題についてもクラウド化により、省スペース、省エネが実現できる。

#### (2) クラウドベース（Cloud Base）

クラウドベースは、各種データベースを緩やかに連携するための仕組みとして構築する。その中で検討されるべき項目としては以下のようなものがある。

##### a) フェデレーション技術

独立したデータベースシステムの自治を尊重しつつ、クラウドとして各種データベースへのアクセスを実現するための仕組みとして、ゆるやかな連携をすすめていくためのフェデレーション技術を開発する。

##### b) スーパーコンピュータを含む分散コンピューティング

シミュレーションやデータの解析を行うスーパーコンピュータもクラウドの中に位置づけられ、スーパーコンピュータへのアクセスを容易にする技術を開発する。

##### c) 大規模分散ファイルシステム構築技術

単なるデータベースとしてのアクセスだけではなく、大規模データのファイル転送やバックアップなどを含む、高速計算に対応できる高速なファイル入出力を行うための技術を開発する。

格納されているデータ形式、ファイルのサイズが多様であるため、ネットワーク上に分散している物理的なディスクや物理的なサーバを仮想化して論理的に管理する。稼動していないものはリソース・プールに登録して、変動する要求に応じて動的にそこからリソースをわりあてて、スケーラブルなサービス提供をおこなうためのファイル仮想化技術を開発する。

## d) セキュリティ技術

クラウド化の中で最も課題となるのは、セキュリティである。クラウド内のデータの機密性や完全性を保障するための技術、およびデータへのアクセス制御とそのためユーザ認証技術を開発する。

## e) ワークフロー技術

時間のかかるデータ解析や、実験のデータ処理の手順をあらかじめワークフローとして実現する技術を開発する。共同研究や、臨床試験など複数の人が関わる作業の順序（業務フロー）を柔軟に設定し、連携を容易にする技術を開発する。

## f) 大規模運用管理技術

かならずしも中央集権的管理とはならないため、分散協調型の管理技術、およびユーザ登録や、アプリケーションの登録管理技術を開発する。

## (3) データスペース (Data Space)

多様なデータ（論文、シーケンスデータ、画像、各種診療情報など）を総合して扱うために、データベースよりさらに上位の概念として、ツールが扱う各種データを取り込み関係させる仕組みを開発する。XML (Extended Markup Language) を基準としたセマンティックWeb技術を中心に各種のデータをどうリンクさせて構造化させるか、データの標準化、辞書の整備、オントロジーなどの開発もあわせて行う。

## (4) Virtual LabおよびVirtual Community

セキュリティが保たれた空間として、研究者がツールやデータを扱えるネットワーク上の研究室を提供する。あわせて、研究者コミュニティが議論や成果を公開できるネットワーク上の場所を提供する。Virtual Labのメンバーであることの認証や、メンバー間のセキュアな通信、共有データや個別データへのアクセス制御などのセキュリティ技術、およびVirtual Communityに関してもクローズドなコミュニティの形成を可能とする仕組みや、共同作業空間を構築する技術を開発する。

### 3.2 ツール連携およびデータ標準化に関する技術開発課題

インフォマティクス関連でいろいろなツールが作成される。ツール間でのデータ受け渡しや、ツールのユーザインタフェースの統一感など、研究者がいろいろなツールを組み合わせ利用することが可能となるような仕組み、そのためのツール作成のガイドラインあるいはツール間インタ

フェース仕様などを開発する。世界での取り組みを見据え、個別独自のものとならないように留意しつつ、国際的なインタフェース仕様をリードあるいはその策定に貢献する技術を開発する。

医療データを始めとする、様々なデータのメタデータおよびオントロジーの標準化が進められており、国際的活動と整合をとりつつ、我が国での標準化を推進する。

### 3.3 インフォマティクスツールに関する研究開発課題

#### (1) Privacy preserved mining

患者の個人情報を含む診療データを扱うため、プライバシーを保ちながらいろいろな検索を可能とする技術を開発する。アクセス制御や暗号化といった通常の情報セキュリティ技術に加えて、暗号化されたデータを復号しないで検索するなどの技術を開発する。

#### (2) 多様なデータの処理・検索技術（論文、画像、パスウェイ、ネットワーク図、動画など）

テキストベースの研究論文とあわせて、その論文のためのデータやプログラム、実験の動画など様々なメディアへのリンクや、横断的検索を可能とする技術を開発する。単なるキーワード（文字列）検索にとどまらず、例えば化合物構造や塩基配列などによる検索など、多様な検索技術を開発する。階層構造をもつデータの一括検索技術や推論検索技術といった特徴ある検索技術を開発する。診療情報については、個人の時系列データの検索と医療の高度化、標準化に向けた横断的な検索が可能となる技術を開発する。

#### (3) 高速シーケンサからの大量データを高速に処理するためのストリーミング技術

非常に大量のデータが高速シーケンサから出力される。これを一旦蓄積してから解析しては処理時間がかかってしまう。高速にデータ処理しながら同時に蓄積するなどのストリーミング技術を開発する。

#### (4) 可視化技術

ゲノムをはじめとする多量な情報の理解、対象の全体像の理解、解析の補助としての可視化技術を開発する。診断の際にゲノムを解析し個別化医療を行うことを想定した場合には、医師にわかる可視化、さらに患者に説明するための可視化といった観点も重要となる。

臓器シミュレータなどでの画像提示のための3次元グラフィックス

技術など、いろいろなシミュレータの画像処理技術や入出力技術を開発する。

(5) ヒューマンインタフェース技術

医療従事者にとってデータ入力しやすいインタフェースを開発する。これまでのキーボードによる入力だけでなく、ICタグや、センサーなどを利用した行動入力による方法、あるいは音声による入力方法などマルチモーダルなインタフェース技術を開発する。

提案の内容

THIBの構築と  
活用に投資する意義

具体的な研究開発課題

研究開発の推進方法

科学技術上の効果

社会・経済的効果

時間軸に関する考察

検討の経緯

国内外の状況

付録・専門用語説明



## 4. 研究開発の推進方法

ここでは、研究開発を推進する体制、推進するにあたっての課題、および人材育成に関して述べる。

### 4.1 研究開発推進体制

3章でも述べたように、取り組む研究開発課題としては、既存技術のインテグレーション主体のインフラ構築部分と、先端研究としてのインフォマティクスツールの研究開発、個別研究としてのクリニカルインフォマティクス、バイオインフォマティクスがある。プロジェクトのキーとなるデータ収集を実際に行う方法として以下のような方法が考えられる。

疾病を決め、それに関するデータを複数の病院でまず収集し、データ収集上の課題などをクリアにした上で、できるだけ多くの病院からのデータを収集する仕組みを構築する。

非常にまれな疾病に関しては、専門医のネットワークを構築し、個別データを収集する仕組みを構築する。

それぞれ拠点を決め、病院のネットワークあるいは専門医のネットワークを構築する。

実際に研究開発を推進する場合には以下のようなチーム構成が考えられる（図3）。各チームの役割を以下に述べる。

#### (1) 総括チーム

プロジェクト推進上の課題解決、進捗管理を実施する。

#### (2) 倫理・規制対応チーム

研究を進めるにあたっての倫理的課題および規制に関わる検討とその対応方針の策定を実施する。

医療研究者、バイオインフォマティクス研究者をはじめ、法律家、倫理問題研究者、政策担当者など多様な人の参加が望まれる。

#### (3) インフラチーム

システムの全体設計として、システムアーキテクチャを決め、インフラのソフトウェアを開発し、インフラとして実装する。

ここには、構築者としての情報技術の研究者、利用者として要件を提示できる医療研究者、バイオインフォマティクス研究者の参加が望まれる。

#### (4) ツール連携・データ標準化チーム

個別課題チームで作成されるツール群が相互接続性をもつように、ツールのインタフェース等の標準化および、マッシュアップのための

仕組みを構築する。またデータの標準化についてもあわせて実施する。

情報技術の研究者、バイオインフォマティクスの研究者、医療研究者の参加が望まれる。

#### (5) 個別課題チーム

バイオインフォマティクスや、クリニカルインフォマティクスなど、例えば疾病の特定に関わる遺伝子の探索や、疾病発症に関わる代謝メカニズムなどの研究課題を実施する。

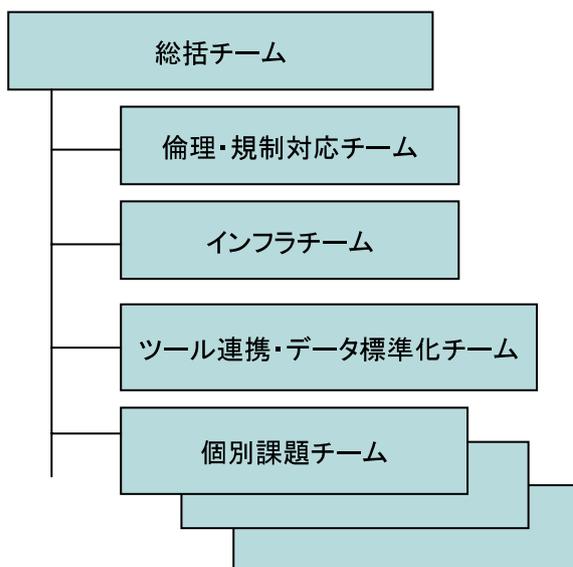


図3 研究開発推進体制の例

## 4.2 推進上の課題

### (1) 倫理的・社会的課題

研究の目的や手法、技術の進展によって、倫理問題への対応も変化しつつある。単純に提供者と試料・解析データの連結可能匿名化にとどまらない。インフォームドコンセント手続き、個人遺伝情報の保護、サンプルとデータの厳格管理のそれぞれの段階で、多様な法的・倫理的課題を時流に合わせて抽出・検討し、適切な対応を行う必要がある。

倫理・規制対応チームにて取り組む。

### (2) セキュリティおよびプライバシー

患者の個人情報を含む診療データを扱うため、セキュリティ特にプライバシー保護の観点が必要な課題となる。情報セキュリティ技術の適用のみならず、運用のガイドラインなどの整備と、それを個人病院のような末端まで広げていく取り組みが必要となる。

総括チームおよび倫理・規制チームにて取り組む。

### (3) 収集データの信頼性、継続性

診療データに関する現状の課題としては、①病名が医療費請求のために便宜的な病名となっている、②病院や医師によって病名の付け方に相違がある、といった課題がある。医療用語をはじめとする標準化が必要であり、その前段階として、医療辞書の整備やオントロジーの整備を加速化する。また、収集したデータがプロジェクトの終了によって保管・維持に必要な予算措置が停止し、散逸してしまうようなことがないように、国としてデータを恒久的に維持する仕組みを整備する。例えば、省庁横断でTHIB運営センター的組織を設けるといったことも考えられる。

総括チームにて取り組む。

### (4) インセンティブ設計

データ提供者、病院およびそのデータを利用する研究者、研究機関、保険者、医療・健康産業、行政など多様なステークホルダーのインセンティブ（メリット、デメリットなど）を整理し、課題を見つけ出し、そこを解決するシステム設計をする必要がある。

例えば、どこまでのデータを収集するかという観点から、疾患別のデータベースとした場合には、一定条件を満たした疾患群の患者に関して、インセンティブやコストパフォーマンスの説明をし易いといったことがある。また、収集されたデータに対する受益者（データを利用して研究する研究者等）がいる大学病院、医療機関と、当面はデータを提供するだけの医療機関（個人病院等）の違いを考慮しなければならない。

総括チームにて取り組む。

### (5) データベースの著作権、利用規約

データを収集し公開するにあたって、その著作権が課題となる。公開されているデータを安心して利用するためにも著作権上の課題をクリアしておかなければならない。

せっかく収集したデータが、特定目的のための利用に限られる現状を変えていく必要があり、倫理の面も含め議論を深めていく。

総括チームにて取り組む。

### (6) データの公開

データを公開することに関しては、共有するデータに対してレベル設定を行い、レベルに応じたアクセス制御の実施などを検討する。米国NIHでは、研究データの公開をグラントの要件とするような取り組みもあり、こうした取り組みに関して検討を進める。

総括チームにて取り組む。

#### (7) 医師会や学協会との連携

THIBへのデータ提供に関する啓蒙や推進に関して、医師会や学協会での推進を実施する。また、学協会などでのデータの標準化に関する活動に対して、ファンディングを実施し、標準化を強力に推進する。ツール連携・データ標準化チーム、総括チームにて取り組む。

### 4.3 人材育成

#### (1) 情報研究者の誘導

インフラ系の構築はもちろんのこと、バイオインフォマティクスやクリニカルインフォマティクス分野での情報関連の研究者が必要とされる場面がますます増えてきており、ファンディングにより研究者を誘導し育成を図っていくシステムづくりが重要である。

#### (2) 研究支援人材の育成

医師に代わって収集データに関する説明を行い、インフォームドコンセントを得る作業に従事するメディカルコーディネーター、収集したデータを整理しデータベースとして利用しやすいものにしていくアナレーター、データベースの運用管理者など、実際の研究に必要な環境を整備・維持していく人材の育成もあわせて実施する。診療情報管理士といった公的資格に準ずる研究情報管理士といった資格制度も検討すべき課題である。

#### (3) 人材の継続性

トランスレーショナルヘルスインフォマティクスの人材育成は既存の大学組織の中ではサポートを受けにくく、現状では競争的資金に依存せざるを得ない。境界領域的人材の養成、雇用、活動を支援し、講座に定着させる仕組みが不十分である。プロジェクトの切れ目におきるキャリアパスの中断は大きな問題である。米国の大学では競争的資金が切れても一定期間研究者をPI (Principal Investigator) が雇うことができる資金制度があり、我が国でも検討すべき課題である。

## 5. 科学技術上の効果

医学ならびにライフサイエンス分野において期待される効果としては、以下がある。

- (1) 関連した研究分野間の情報連係がスムーズになることによる新しい知識の発見
 

大規模なコホート研究や、臨床データの蓄積をベースに総合的研究を実施することで、疾患理解や予防に関する新しい知見が得られることが期待できる。これらは新たな医薬品等の開発の促進、効率化にもつながる。
- (2) パーソナルゲノム解読を用いた個別化医療の実現に向けた技術の進展
 

ゲノム解析技術において、高速シーケンサが開発され、解析チップも従来に比して高性能化がすすんでおり、近い将来には1000ドル程度で個人の全ゲノム解析が可能になる時代がくる。パーソナルゲノム情報にもとづいた個別化医療（オーダーメイド医療）への適用が可能となる。疾病と遺伝子の関連性解析に基づいた薬剤の特定、投与方法などの解明が進展する。
- (3) EGM（Evidence Generating Medicine）の推進
 

情報基盤の構築によって、基礎研究領域での発見、エビデンスをベースとした医療、すなわちEBM（Evidence Based Medicine）を補完することができる。それだけでなく、情報を循環させるという観点から、医療側でのデータ（診療プロセスや結果のデータ）をエビデンスとして生成し、研究に活用するという考え方EGM（Evidence Generating Medicine）をライフサイエンスの現場で実践することも可能となる。

また、情報技術分野において期待される効果としては、

- (4) 情報技術フロンティアの開拓
 

シーケンサのデータ出力の高速・大容量化は、計算機処理の高速化を越えたスピードで進展しており、高速シーケンサが生成する大量のデータをいかに高速に処理するかが、情報処理技術としても大きな課題となり、ストリーミング処理技術などで新しい展開が期待できる。

また、Privacy Preserved Miningや多様なデータの処理検索技術、可視化技術、ヒューマンインタフェース技術などにも新しい展開が期待できる。



## 6. 社会・経済的効果

### (1) 医療の質の向上、医療経済的効果

THIBを整備し、収集する診療データのメタデータを標準化するなどし、大規模にデータを収集することを可能とすることにより、医療プロセスとその効果（アウトカム）の定量評価と比較が可能となり、標準的かつ効果的な医療プロセスの構築への足がかりとすることができる。具体的には、診療ガイドラインの整備等により医師の診療の負担軽減が図れ、医療プロセス全体の底上げが可能となり、医療の質の向上につながる。また、そこで得られる診療結果などのアウトカムデータにより、医療の経済的効果の測定も可能となり、医療の効率化につながることを期待される。

### (2) 医療産業競争力・研究力の強化

日本の医療・健康産業の産業基盤は強固とは言いがたく、同産業の基礎体力ともいふべき基礎研究から、その発展に必須である橋渡し研究および臨床研究への連携がよりスムーズになり、実際の創薬や診療における研究効率や生産性が向上することが期待される。さらに、こうした情報基盤によって、規制科学（レギュラトリーサイエンス）などへの適用を見据えた研究が充実し、広く医療産業におけるわが国の競争力の強化につながる。

### (3) 疫学の進展により最適な治療や予防が可能

情報基盤により信頼性の高いデータを効率よく集めることが可能となり、データの蓄積によりゲノム疫学といった分野を進展させることにより、最適な治療や予防が可能となる。



## 7. 時間軸に関する考察

4章でも述べたが、医療データ収集の観点から、プロジェクトの形態として以下のようなものが考えられる。

- 疾病を決め、それに関するデータを複数の病院でまず収集し、データ収集上の課題などをクリアにした上で、できるだけ多くの病院からのデータを収集する仕組みを構築する。
- 非常にまれな疾病に関しては、専門医のネットワークを構築し、個別データを収集する仕組みを構築する。

上記観点から疾病を決め、総括機能を担う拠点組織を決め、スモールプロジェクトを立ち上げ、3年～5年間実施し、インフラの設計と基本機能の構築するとともに、データ標準化や共用ツールに関する成功事例を積み上げる。データ標準化のガイドライン、ツール作成のガイドライン、倫理ガイドラインなどが作成される。あわせて、活動をネットワーク化し拡げていくコンソーシアム（コミュニティ）のような仕組みを構築する。また、THIBをクラウド的運用にもっていくための、仕組み（マネージメント）を構築する。

次に、基本機能として構築されたインフラを利用した数個の個別プロジェクトを3～5年間実施し、インフラの拡張・強化を図るとともに個別課題研究に取り組む。さらにコンソーシアム（コミュニティ）を拡充する。個別課題研究に関しては、拠点にて実施する課題と公募による課題推進の2本立てで実施し、拠点に成果が蓄積されるよう配慮する。

こうした国としての情報資源を恒久的に維持管理していく仕組みとして本提案のクラウドベースによる緩やかな連携整備を位置づけ、各省庁の所管するデータベースを含め国として整備していくものとする。

THIBを広め、世界の注目を集める高度なものとすることによって、日本の情報基盤から世界に発信すると同時に、世界の情報基盤の標準化をリードし、世界に貢献することができる。