

CRDS-FY2009-SP-08

戦略プログラム

生命・医学・医療・健康をつなぐ情報を
循環させる技術と基盤の構築と活用

～トランスレーショナル・ヘルスインフォマティクス・ベースの展開～

Strategic Program

Development of Translational Health Informatics Base



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

本戦略プログラムは、医学・医療における「知の循環」の確立に必要なとなる情報を循環させる技術と基盤（ここではT H I B：トランスレーショナル・ヘルスインフォマティクス・ベースと名づけた）の構築と活用を提案するものである。

T H I Bを構築することで、ライフサイエンス研究の成果と個々人の医療情報・診療データの相互利用が可能となり、基礎医学の進展や医療の高度化（個別化医療の実現など）に加えて、疫学研究の進展、医療プロセスの高品質化、さらには経済面からの医療の効率化といった効果が期待できる。

優れたライフサイエンスの成果を活用し、新しい治療法や医薬品・医療機器等として社会に還元していくための「健康研究（Health Research）」（橋渡し研究・臨床研究）の必要性から、わが国にも2008年より内閣府の下に健康研究推進会議が設立されている。その健康研究推進会議より提案された「健康研究推進戦略」の中では、基礎研究により生命現象と疾病のメカニズムを解明し、それを診断や治療法に転換して医療を実践し、その医療効果を評価して新たな課題を設定し、それを再び基礎研究につなげていくという、医学・医療における知の循環の確立を目指すべきとされている。

こうした状況にかんがみ、我が国のライフサイエンス研究の成果が円滑に医療に活用されるために、情報基盤の整備という観点から、ライフサイエンス分野のデータベースを十分に活用できる基盤を作り、ライフサイエンス研究の効率や質を向上させることを目的に、いくつかのプロジェクトが推進されてきた。しかしながら、患者の個人データを含む診療データを扱うためのセキュリティ、特にプライバシー保護の課題、診療データにおける病名を始めとする医療用語の相違によるデータ収集の困難さといった課題があり、医療側のデータ整備が十分でなく、ライフサイエンス研究と医療側の研究を両輪として発展させていく余地は十分残されている。今後これらを両輪として発展させていくためには、医療側データの整備を進めるための医療データの標準化や、プライバシー保護技術、認証・アクセス制御技術、さらに各種データベースを連携させるための技術を開発する必要がある。

ライフサイエンスの最先端の知と医療を結びつけるための情報基盤、さらには臨床データを有効に利用するための情報基盤、診療データ、健康診断データ、疫学調査データなどの活用も視野にいた情報基盤を総合的にT H I Bとして構築する。T H I Bの構築・活用に当たっては、インフォマティクスに関連する先進的研究への投資だけでなく、標準化活動や共用ツールなどのインフラ構築部分への投資と恒久的運用が必要である。

さらに、T H I B構築の具体的推進方法としては、当初から統一的大規

模システム開発を目指すのではなく、拠点を決めスモールプロジェクトでスタートさせ、インフラ構築を進めつつコミュニティを拡げていくこととし、既存システムの連携（フェデレーション）に重点を置いたシステム構築を目指す。

（注）本提案は、拠点でのインフラ構築整備に関わる資金と、個別研究課題を公募で実施する競争的資金を含む。

Executive Summary

This strategic program proposes the development and utilization of the informatics base (we termed as translational health informatics base (THIB)) which enables “the circulation of knowledge” in the medical science and clinical services.

By developing THIB, the fruits of life science research, personal medical information and clinical care data become interoperative. As a result of this, we can expect the following effects: the progress of epidemiological studies, improvement of clinical services in their qualities and efficiency.

It has been discussed that we should perform “Health Research” which enables us to make use of the fruits of latest life science research and to return it to the society as a more advanced treatment method, medicine or medical instrument. To this end, the Health Research Promotion Council was established in Japan under the control of the Cabinet Office in 2008. In the “Health Research Promotion Strategy” proposed by the Council, it is recommended that we establish “the circulation of knowledge” loop in the medical science and clinical services by clarifying the mechanism of life phenomenon and diseases, converting the clarified mechanism into diagnostic and treatment methods, identifying new research subjects through assessing the effectiveness of the methods, and carrying out research on the identified subjects.

Several projects have been promoted for developing information infrastructure to smoothly turn the fruits of the life science research into clinical services. These projects are aiming at creating the basis to fully utilize data base in the life science fields and improve the efficiency and quality of life science research. However, these projects did not have enough coordination with the studies performed in medical institutions and there is much room left to enhance “the circulation of knowledge” loop in the life science research and studies performed in the medical institutions. From this point of view, we propose comprehensive development of those information infrastructures as THIB which allows us to link cutting-edge knowledge in the life science, clinical data, medical data, health examination data, and epidemiological study data. When developing and utilizing THIB, we need to fund not only in informatics studies, but also infrastructure build-up, standardization activity, common tools, and sustainable operation of THIB.

Rather than trying to develop a full-scale THIB system from the very beginning, we should start from a small trial-phase project and gradually expand the institutional community. From a practical standpoint, an emphasis should be placed on the federation of the existing systems throughout the development.

目 次

エグゼクティブサマリー Executive Summary

1. 提案の内容	1
2. THIBの構築と活用に投資する意義	3
3. 具体的な研究開発課題	5
4. 研究開発の推進方法	11
5. 科学技術上の効果	15
6. 社会・経済的效果	17
7. 時間軸に関する考察	19
8. 検討の経緯	21
9. 国内外の状況	22
付録1 参考文献	29
付録2 専門用語説明	30

1. 提案の内容

本戦略プログラムは、健康で安心な社会の確立、医学分野での技術立国を目標として、医学研究のサイクルと医療提供のサイクルをなめらかに連携してまわし、医学研究から医療・健康といった社会実装への加速化を実現する情報の循環のための技術と基盤（ここでは、THIB：Translational Health Informatics Baseと名づける）を構築・活用することを提案する（図1参照）。

我が国では、ライフサイエンスに関する基礎研究の分野での成果は出ているが、実際の医療への適用という観点からは、出ている成果が十分に活かしていない状況にある¹。JST-研究開発戦略センター（CRDS）は、ライフサイエンス分野で得られた知識を、臨床的課題の解決に結びつけるための転換、橋渡しを行う研究開発分野を発達させる必要があるとの観点から、戦略イニシアティブ「統合的迅速臨床研究（ICR）の推進—健康・医療イノベーション」²、戦略イニシアティブ「医工融合によるイノベーションの推進—医工融合研究のグランドデザイン」³を提言してきた。

これまでライフサイエンスの分野では、データベースを利用した研究が細分化され閉じた世界で個別に実施されてきている。こうした分散した状況を打破するために、情報基盤の整備という観点から、データベースを統合する動きがある。文部科学省ではライフサイエンス統合データベースプロジェクトを、データベースを十分に活用できる基盤を作り、ライフサイエンス研究の効率や質を向上させることを目的に推進している。しかしながら、患者の個人データを含む診療データを扱うためのセキュリティ特にプライバシー保護の課題、診療データにおける病名を始めとする医療用語の相違によるデータ収集の困難さといった課題があり、医療側データの整備が必ずしも十分ではない。また、医療データを収集する意義に関する合意形成がいろいろなステークホルダーの間で必ずしも十分ではない。このような状況を打破するためには、制度面、技術面、倫理面など多角的な検討が必要であるが、使いやすい情報基盤の整備、それを可能とする情報技術研究の推進、医療側データの整備を進めるための医療データの標準化や、プライバシー保護技術、認証・アクセス制御技術、さらにライフサイエンス側のデータベースとの連携を促進するための技術などが重要である。また、この情報基盤の構築と運用をどのような体制によって推進するのも重要なポイントである。

本戦略プログラムではライフサイエンスの最先端の知と医療の現場を結びつけ、さらには臨床データをライフサイエンスの研究者が有効に利用するための情報基盤をTHIBとして構築することを提案する。情報

¹ 政策研ニュース No.25 2008年7月

² CRDS-FY2006-SP-08

³ CRDS-FY2006-SP-14

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。

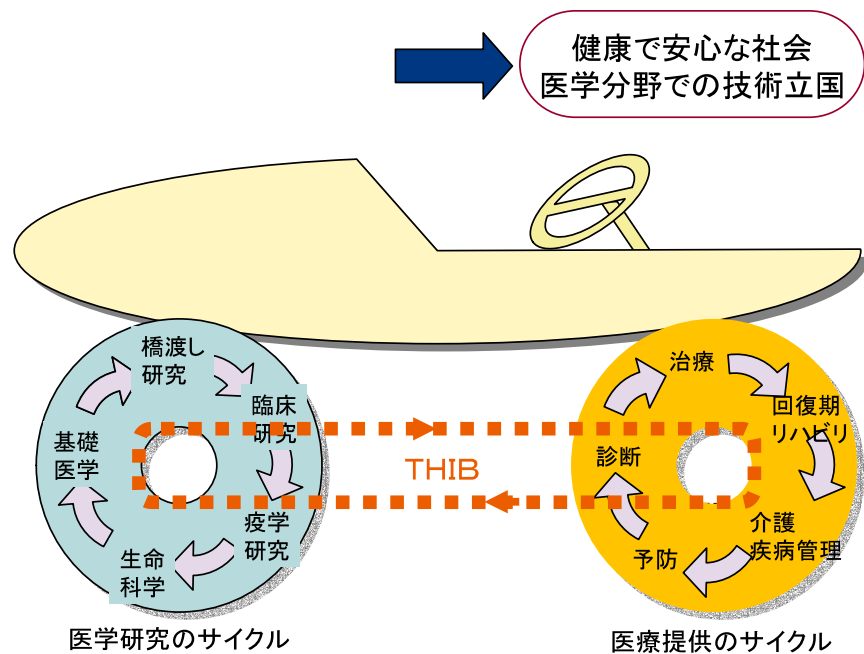


図 1. 概念図

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。

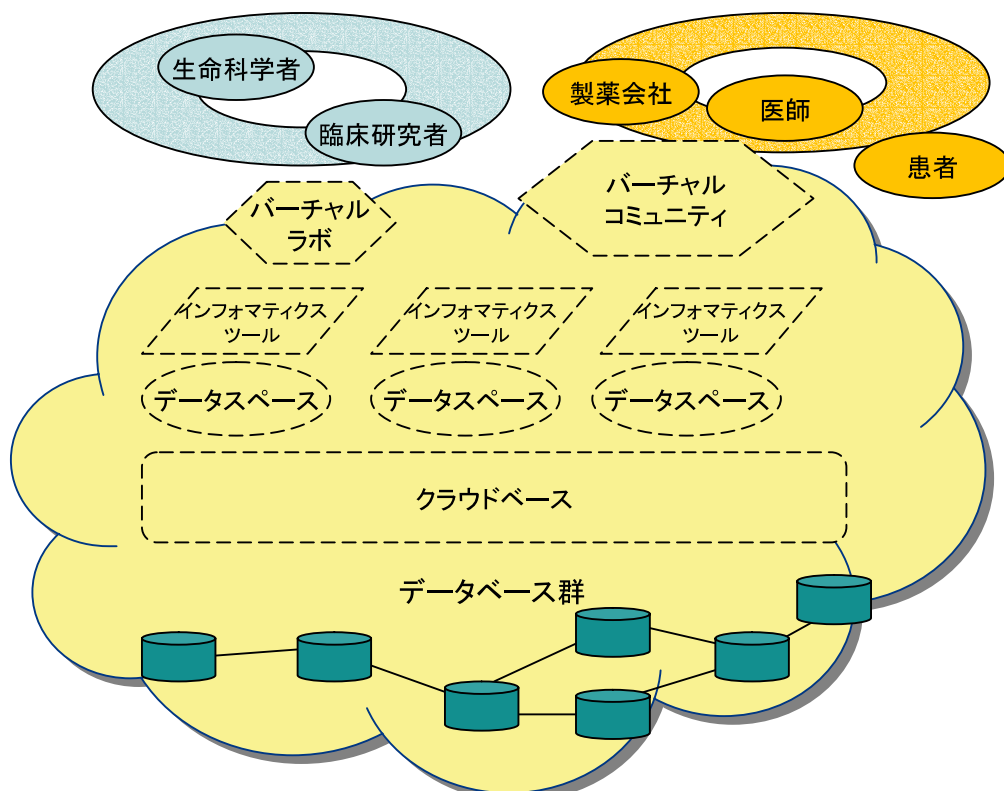


図2 THIB構成概要図

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。生命科学者、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、バーチャルラボ、バーチャル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。

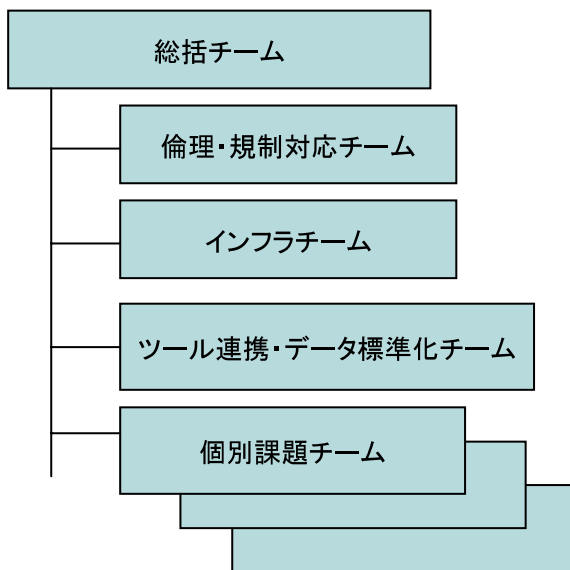


図3 研究開発推進体制の例

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。

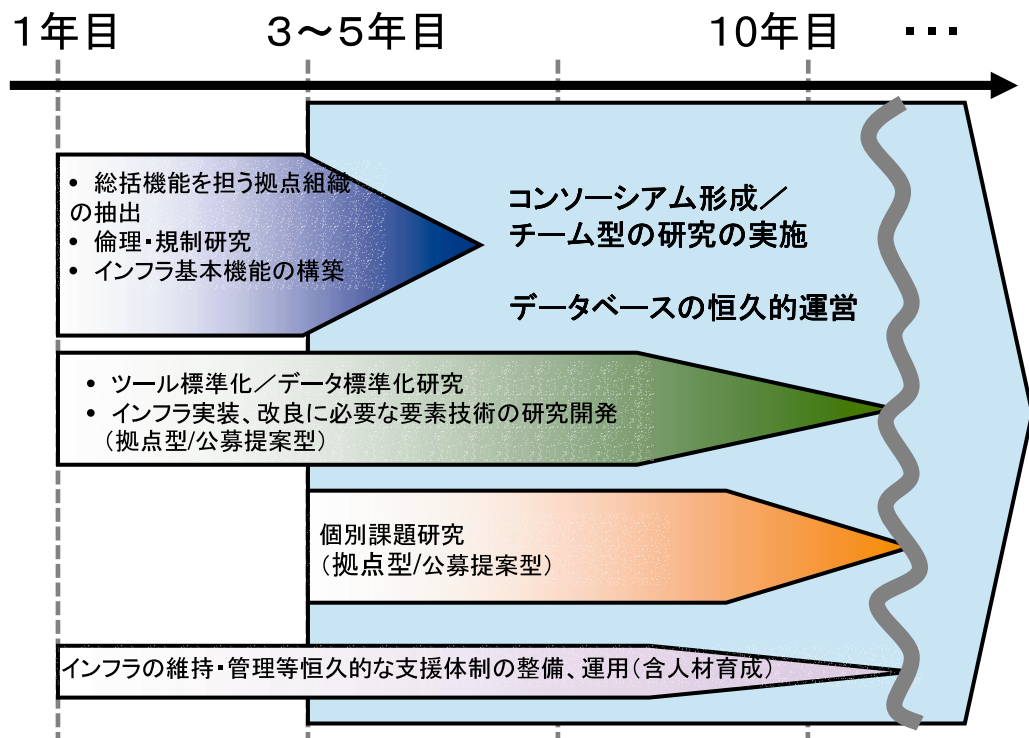


図4 時間軸

8. 検討の経緯

8.1 有識者インタビュー

以下の有識者から提案の構想段階で有益なご意見を伺った。

- ・開原成允（国際医療福祉大学 大学院長）
- ・池田俊也（国際医療福祉大学 教授）
- ・堀田凱樹（情報・システム研究機構 機構長）
- ・高木利久（ライフサイエンス統合DBセンター センター長）
- ・井原茂男（東京大学 先端科学技術研究センター システム生物学ラボラトリ 教授）
- ・田中博（東京医科歯科大学 大学院生命科学研究部 教授）
- ・大江和彦（東京大学 大学院医学系研究科 医療情報経済分野 教授）
- ・真鍋一郎（東京大学 大学院医学系研究科 疾患生命工学センター 特任講師）
- ・興紹貴英（東京大学 大学院医学系研究科 22世紀医療センター 助教）
- ・中村祐輔（東京大学 医科学研究所 ヒトゲノム解析センター長 教授）
- ・豊田哲郎（理化学研究所 生命情報基盤研究部門 部門長）
- ・安達淳（国立情報学研究所 学術基盤推進部長 教授）

8.2 ワークショップ

科学技術未来戦略ワークショップ「トランスレーショナル・インフォマティクス・ベースの展開」を2010年1月18日（月）13:30～18:30 研究開発戦略センター2F大会議室にて開催した。少人数の有識者を招聘し、提案の内容を補強する目的で実施した。

<ワークショップ参加者>

- ・安達淳 国立情報学研究所 学術基盤推進部長 教授
- ・大江和彦 東京大学 大学院医学系研究科 医療情報経済分野 教授
- ・田中博 東京医科歯科大学 大学院生命科学研究部 教授
- ・豊田哲郎 理化学研究所 生命情報基盤研究部門 部門長
- ・永井良三 東京大学 大学院医学系研究科 教授

<WSプログラム>

13:30～14:00 オープニング

- 連絡事項・資料確認など（CRDS）

- 提案趣旨説明 (CRDS)

14:00~15:00 セッション1 TIB⁴に求められること (情報循環に関して、医療の現場から)

- 永井良三 (医学・医療における知の循環)
- 大江和彦 (医療データの標準化)
- 田中博 (オミックス医療とデータベース)

(休憩)

15:20~16:00 セッション2 TIBの方向性

- 安達淳 (学術共有クラウド)
- 豊田哲郎 (理研サイネス)

16:00~17:30 セッション3 総合討論 (CRDS)

- データ収集、解析に関する課題
- 制度的課題
- 人材育成
- ファンディング
- 研究推進の方法、時間軸
- etc.

17:30~18:30 まとめ

<WSで主な議論>

1. TIBの目指す方向

- 所有するITから利用するITへという観点からクラウド技術が注目を集めている。研究者はデータを所有したいわけではなく、データを利用して研究したい訳であるから、研究に専念できる環境という意味でもクラウド化していく方向であろう。
- バーチャルラボやバーチャルコミュニティの実現により、グローバルな研究環境を構築でき、世界の知を集めることも可能となる。

2. より効率的、持続的なデータ収集、保管、活用のあり方

- 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。
- データの標準化に関しては、メタデータの標準化がむしろ重要である。

3. 研究課題の推進体制ならびに資金配分のあり方

- インフラとしての研究環境へのファンディングのあり方に関して、期

⁴本文中では、THIBとなっているが、ワークショップ開催時点では、TIB (Translational Informatics Base) という呼称で議論した。

限のある競争的資金での継続的運用は難しく、競争的資金ではないファンディングの枠組みを検討する必要がある。

- 推進体制に関しては、疾患別プロジェクト、病院での疾患収集など、いくつかの方法が考えられる。
4. 研究課題の実施に必要な人材育成のあり方
- トランスレーショナル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い。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。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る
5. T I Bを恒久的に維持、発展させるための国民の理解、評価のあり方
- データベースがどのように役立っているか、ということはある程度定量的に呈示しないと、T I Bを維持管理する妥当性や資金の必要性が説明しにくいと考えられる。
 - T I Bに必要な研究領域を包括して学術領域（新学術領域）として確立させることが重要である。メカニズム解明にとどまらない研究領域としての設定が必要。この分野はうまくいけば「自己収入」が可能であることもアピールのポイントなる。

9. 国内外の状況

9.1 国内の状況

ライフサイエンスデータの統合化の動きは、第3期科学技術基本計画の分野別推進戦略「ライフサイエンス研究全体を支える基礎・基盤課題」の中でも重要なテーマとして認識され、大量データ時代の到来にあわせ、施策として加速している。その中の主要プロジェクトである、統合データベースプロジェクト（平成18年度～平成22年度；平成21年度政府予算額9億円 平成20年度11億円）は、我が国のライフサイエンス関係のデータベースの利便性の向上を図るため、データベースの統合化及び利活用のための基盤技術開発、人材育成等を行い、データベースの統合的活用システムを構築することを目的とし、現在、(独)情報・システム機構を中心とした関連中核機関が積極的なデータ統合化を進めている。また、バイオインフォマティクス研究の推進（平成21年度政府予算額18億円、平成20年度17億円）が(独)科学技術振興機構で実施されており、解析の効率化・省力化、利用の高度化等を実現するための革新的なゲノム解析ツールの研究開発等、ファンディング機能も活用してバイオインフォマティクス研究を行っている。上記2つのプロジェクトは平成23年度より統合化されることが決まっており、現在その準備段階にある。

公共的なデータ提供ソースとしては、(独)理化学研究所を中心に、バイオリソース事業（平成21年度政府予算額32億円+補正33億円 平成20年度32+5億円）が我が国の中核的拠点として、生物遺伝資源（バイオリソース）の提供を行うとともに、基礎データの蓄積も含めた基盤技術の開発を進めている。データとそれに伴うバイオリソース両者を保有する機関である、ライフサイエンス研究全体の基盤拠点として期待される。

さらに医療分野に特化した大型プロジェクトとして、バイオバンクジャパン計画（個人の遺伝情報に応じた医療実現化プロジェクト）(第1期 2002-2007年度 2002年度は補正、計220億円 第2期 2008年—2012年（予定）平成21年度予算額27億円）が進捗している。通称「オーダーメイド医療実現化プロジェクト」として、2期目に入っており、東京大学医科学研究所に設置されたバイオバンクに収集された患者DNAサンプルを活用し、(独)理化学研究所がSNP（遺伝子の個人差）と薬剤の効果、副作用などの関係や疾患との関係を明らかにする研究等を実施、バイオバンクに収集された疾患について発症に関与する遺伝子や発症メカニズムの解明を目指し疾患関連遺伝子研究の加速を目指している。

いずれも、信頼性、継続性及び先導性の確保と共に戦略的かつ効率的な整備の促進が重要であり、データの活用、研究成果や基盤技術の普及、人材育成、国内外の関連機関との有機的な連携等が必須であり、このために

も、本戦略プログラムで提案のTHIBの構築が重要となる。

9.2 海外の状況

バイオインフォマティクス関連研究に対する公的資金サポートの状況について以下、米国、英国、欧州、アジアに関する調査結果を述べる。

<米国>

米国では、国立衛生研究所 (National Institutes of Health, NIH) 傘下の国立研究資源センター (National Center for Research Resources, NCRR)内において、2001年からBiomedical Informatics Research Network (BIRN)という研究プログラムが開始されている⁵。

BIRNは全米各地の提携機関の協力を得て医科学研究者向けにソフトウェアツールとインフラリソースを提供している。現在は脳科学に関連した、ヒト脳のMRI画像と医療情報の連携検索を可能にするFunctional Imaging BIRN Tools、マウス脳解剖図のデジタルアトラスを扱う Mouse BIRN Tools、そしてヒト脳の形態計測画像の管理、保管、検索を行うMorphometry BIRN Toolsを提供している。Functional Imaging BIRN Toolsにおいては、被験者(患者) 各種のMRI計測画像と医療情報をURLのリンクによって結びつけることによって、アルツハイマー病、躁鬱病、多動性注意欠陥障害、自閉症などの神経疾患をもつ脳の機能的な特徴の抽出、分析が可能になっている。これらのツールを利用するのは脳科学研究従事者に限らず、他の疾患研究に携わる研究者、技術者、医療従事者も多いことがわかっている。

米国に拠点を置くデータベース系企業では、診療データを元に患者の治療フローを分析し、実際の医療プロセスとそのアウトカムを比較することで医療の質の向上、医療経済的な評価に役立てている。

また、2009年のAmerica Recovery and Reinvestment Act (アメリカ再生・再投資法) のなかにはHealth Information Technology for Economic and Clinical Health (HITECH) Actが含まれ、約20億ドルが医療情報のインフラ整備や導入に費やされることが決定している。

⁵ http://www.nbirn.net/index_ie6.shtml

新規プログラム	内容	投資額 (百万ドル)
地域普及センター	医療情報技術の取り扱い等を医療提供者に普及させるための機関を全米で70機関設立	643
健康情報交流 人材育成プログラム	地域間での健康情報のやり取りを支援 健康情報の専門化の育成など。4万5千人の育成を目標としている	564 118
健康情報技術の高度化に向けた研究	情報の機密性、ネットワークプラットフォームの構築など健康情報技術の高度化を目指した研究開発の支	60
全米健康情報ネットワークなど	健康情報基盤の構築、整備のための投資	64
その他含めた総計		2000

(New England Journal of Medicine 362:382, February 4, 2010より抜粋)

この法律によって、健康情報の電子化がなされ、アメリカ国民の健康やアメリカの医療制度を改善することが目標とされている。また、こうした目標の達成には国民の参加や医療従事者による情報の適切な活用が必要とされており、これに向けた技術開発やインセンティブの付与の仕組みづくりが進められている。

我が国でもEHR (Electronic Health Record) の整備など、THIB構築とともに実施していく必要がある。

<欧州>

欧州の主な資金配分組織は欧州委員会のSeventh Framework Program for Research and Technological Development (FP7)である。FP7Infrastructure枠において、FP6期間中に推進された欧州分子生物学研究所 (European Molecular Biology Laboratory : EMBL) のバイオインフォマティクス事業など、国・地域縦断的に関連した事業を統合・拡充して、インフォマティクス研究プロジェクトのうち急速に増加、拡大する生物医学情報を活用する次世代のバイオインフォマティクス構築の計画を纏めるために、欧州ライフサイエンス情報基盤計画コンソーシアム (European life-science infrastructure for biological information ELIXIR) が設立されている⁶。ELIXIRは2011年からの本格稼働を目指した基盤整備のために、2007年から2010年の間に€4.5millionが投資されている。2011年以降も総額€470millionの基盤整備費用と、毎年€100millionのコンソーシアム運営費用が投じられる予定であり、欧州のバイオインフォマティクスの中核を担う組織になると予想される。その他、欧州委員会の科学技術政策に関する情報検索サービスCORDIS (Community Research and Development Information Service) によると、医科学研究、健康科学など、本戦略プログラムに関連したFP7Cooperation (EU加盟国以外の国の研究者も参加可能な目標達成型国際共同研究) Health枠

⁶ <http://www.elixir-europe.org/page.php?page=home>

において11件採択されている。実施期間は3～5年、配分額は実施期間が3年のものは€2.8 million前後、5年のものは€10 million前後である。研究代表者の所属国を見ると英国が4件と最も多く、次いでイタリア、スペインが2件ずつ、残りは、オランダ、デンマーク、ハンガリーの研究者を代表とするチームに配分されている。

EU各国でも医療情報とライフサイエンス研究の連係を推進している。デンマークでは、病院における患者情報と、分子レベルのオミックスデータを統合する事によって、疾患の病態と遺伝子発現の関係解明を進めている。

特に、欧州での基盤整備プロジェクトは、本プロポーザルのTHIBと関連しており、国際連携を視野に入れたTHIBの構築を目指す必要がある。

付録1 参考文献等

- (1) 内閣府 健康研究推進会議 「健康研究推進戦略」 平成21年7月
- (2) ELIXIR <http://www.elixir-europe.org/page.php>
- (3) 統合データベースプロジェクト <http://lifesciencedb.mext.go.jp/>
- (4) 中村裕輔著 「これからのゲノム医療を知る」 羊土社 2009年
- (5) 森下真一、阿久津達也編、「バイオデータベースとソフトウェア最前線」 実験科学増刊 Vol.26-No7 2008 羊土社

付録2 専門用語説明

アクセス制御

情報セキュリティの用語で、主体（Subject）が対象（Object）に対してアクセス（読み／書き／実行）することを許可したり、拒否したりすること

XML (Extended markup Language)

XMLは文書やデータの意味や構造を記述するためのマークアップ言語の一つ。マークアップ言語とは、「タグ」と呼ばれる特定の文字列で地の文に情報の意味や構造、装飾などを埋め込んでいく言語である。

オントロジー

文書の内容を説明する意味情報（メタデータ）を各文書に付加し、メタデータを記述する用語を定義する構造を構築する。この構造がオントロジーとなる。オントロジーを導入することにより、検索対象となる文書が単なる単語の集まりとしてではなく、文書全体で大きな意味を持ったデータとして扱われ、各文書について統一的な付加情報をもたせることができる

コホート研究

コホート研究は、疫学におけるひとつの研究手法である。特定の因子に暴露した集団と暴露していない集団について、研究対象となる疾患への罹患率を調査し比較することで、因子と疾患の関連を検討する研究手法。

クラウドコンピューティング

クラウドは、ネットワーク（インターネット）をさし、ネットワーク上のサーバで各種の処理が行われる形態をさす。

ゲノム疫学

ゲノムなどの遺伝子情報をもとに、集団における疾患と特定の集団における健康に関連する状況あるいは事象の、分布あるいは規定因子に関する研究を実施すること。

ストリーミング処理

並列処理の1つで、入力したデータに対して、処理を実行しながら、出力を行う。

セマンティックWeb

Webページの意味を扱うことを可能とするために考案され、W3Cにおいて現在標準化が進められている。データの意味を記述したタグが文書の含む意味を形式化し、コンピュータによる自動的な情報の収集や分析へのアプローチが可能となると期待されて

いる。

パスウェイ

パスウェイとは細胞の機能を実現する分子メカニズムを書き下した知識のことで、ヘテロな構成要素からなり、強く（ネットワーク状に）構造化されている特徴がある。

マッシュアップ

Web上に提供されている情報やサービスなどを組み合わせて、新しいソフトウェアやサービス、データベースなどを作ること

メタデータ

データについての情報を記述したデータ。あるデータそのものではなく、そのデータに関する情報のことである。例えば、データの作成日時や、作成者、データ形式などである。

■戦略プログラム作成メンバー■

勝山光太郎	フェロー	(電子情報通信ユニット)
高野 守	フェロー	(ライフサイエンスユニット)
福士 珠美	フェロー	(ライフサイエンスユニット)
山本 雄士	フェロー	(臨床医学ユニット)
丹羽 邦彦	上席フェロー	(電子情報通信ユニット)

※お問い合わせ等は下記ユニットまでお願いします。

CRDS-FY2010-SP-08

戦略プログラム

生命・医学・医療・健康をつなぐ情報を 循環させる技術と基盤の構築と活用

～トランスレーショナル・ヘルスインフォマティクス・ベースの展開～

Strategic Program

Development of Translational Health Informatics Base

平成22年3月 March 2010

独立行政法人科学技術振興機構 研究開発戦略センター 電子情報通信ユニット
Center for Research and Development Strategy
Japan Science and Technology Agency

〒102-0084 東京都千代田区二番町3番地

電話 03-5214-7484

ファックス 03-5214-7385

<http://crds.jst.go.jp/>

© 2010 JST/CRDS

許可無く複写／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.

ATTAATC A AAGA CCTAACT CTCAGACC
CT CTCGCC AATTAATA
TAA TAATC
TTGCAATTGGA CCCC
AATTCC AAAA GGCCTTAA CCTAC
ATAAGA CTCTAACT CTCGCC
AA TAATC

AAT A TCTATAAGA CTCTAACT CTAAT A TCTAT
CTCGCC AATTAATA
ATTAATC A AAGA CCTAACT CTCAGACC
AAT A TCTATAAGA CTCTAACT
CTCGCC AATTAATA

TTAATC A AAGA CCTAACT CTCAGACC
AAT A TCTATAAGA CTCTAACT
ATTAATC A AAGA CCT
GA CCTAACT CTCAGACC

0011 1110 000
00 11 001010 1
0011 1110 000
0100 11100 11100 101010000111
001100 110010
0001 0011 11110 000101

