独立行政法人　科学技術振興機構（JST）
「ライフサイエンス分野統合データベースセンター設置準備の検討とりまとめ」
（2010 年 6 月 14 日）

目次

１．わが国の統合データベース・センターの創立へ向けての基本的な考え方

昨年、専門家による三回にわたる委員会、データ作成者、データ利用者からの意見聞き取り、NCBI においてリップマン所長及び関係者との意見交換などを行った。これらを通じて得られた、今後の統合データベース・センター創立のために特に留意すべき点は以下の通りである。

（１） 背景

最近におけるライフサイエンスの急速な進歩によって、生命の実体及びその機能を担う DNA、RNA、タンパク質、代謝産物（分子）などに関する情報が飛躍的に増大し、微生物、植物、動物、ヒトなどの生命体を研究する分野が共通の基盤で情報の交換が行われ、またそれを支えるテクノロジーも共通化、一体化して研究を推進するようになった。更にはそれらの応用分野である我々の生活に密接に関係する医学、農学などの諸分野においても、これらの基礎科学の知見及びそれに伴うデータを利用することがきわめて重要になり、今やデータベースの利用はライフサイエンスの進歩発展にとって必須条件となっている。従って、そこに蓄積された膨大な情報をいかに有機的に統合し、その利用者に効率的に提供するシステムの構築をするかきわめて重要であり、今後の日本のライフサイエンスの進歩及びその産業への波及にとっての欠かすことが出来ない条件になった。しかしながら我が国の現状は、これらの膨大な情報に関する集積及びその利用に関する統一的な場がなく、加えて各省庁間におけるデータの共有の重要性に関する認識が不十分であるがために、データベースは多くの機関に分散し、従ってその国際的地位は、NCBI に代表される世界のデータベースの中でも必ずしも高くない。従って今後、我が国のライフサイエンスとその応用分野の発展を図る上で、従来、ばらばらで相互連関が不完全であった多くのデータベースを統合し、フォーマット、相互リンク、データの質などを一元化した統合データベース・センター（仮称）の創立が焦眉の課題である。また、アジアにおいて、特に中国などきわめて急速、かつ革新的なデータベース整備の動きが見られるので、本センターの創立は、アジアでの科学技術をリードすべき我が国の国策及び国益にとってもきわめて重要なものである。更に各データベース間の有機的連携の欠如からその運営に関しての費用対効果が低くなっていることも昨近の政治的課題となりうるかも知れない。

具体的には、統合データベースは以下のようなものであるべきものと考える。

（２） 統合データベース・センターの機構

統合データベース・センターには実務的なデータを取り扱う部門と、絶えず変化しているデータベースに関する様々な課題を研究、検証するための部門から成り立つことが望ましい。前者においては統一的な言語、フォーマット、データの質の確保を目指し、データ作成者の意見はもちろん、従来、あまり関心が払われなかったデータに対するユーザの意見を常に反映させ、世界に誇るべき利便性の高いデータベースを目指す。後者については将来ますますデータの複雑化、データ量の膨大化に対応して、どのような形で我が国のデータベースを対応、構築していくかについての研究、検証を行い、それを実際の統合データベースの運営に反映させるようにする。この場合、若い意欲のある研究者の参加が欠かせない。一方、ライフサイエンス分野のデータベース構築や統合化のための人材は必ずしも十分とはいえず、統合データベース・センターを中心とした人材育成やキャリアパスの構築も必要である。また、上記機関の運営は、データ生産者の協力や利用者の意見を集約しながら進めることが重要であり、そのため各機関の代表者も含めたオープンな運営に努める必要がある。さらに、我が国のデータベースの世界的に見た独自性及びその存在意義を高めるために努力することも極めて重要である。より具体的には、アジア人のゲノム・データベース、イネなどのアジア固有の農業植物のデータベース、我が国のもつ豊富な産業用微生物に関するデータベースなどを充実する。

（３） 既存のデータベースとの関係

現在の公的及び私的研究機関にある既存のデータベース及び将来作成、構築したデータまたはデータベースを本データベース・センターに提供してもらえるよう努力する。この際の原則として、①各機関の意向を尊重し、ヴォランタリーな提供を基本とする。②それらのデータの質、フォーマットなどが統合データベースのスタンダードに合致するように協力を求め、場合によっては技術支援を行う。③これらのために必要に応じて資金の援助を行う。また、データベース構築の早い段階からデータベース・センターより統合化を支援する。

一方、本データベースにデータ提供者がそのデータを提供することによって、提供者のデータベースの価値（プレステージ）がより高く認められるような国際的なデータベースにすることが必須である。これによってデータ作成者にとっても、データを提供することが、強いインセンティブになることを期待する。

（４） データ公開基準

本データベース・センターを通したデータの公開に関しては、全面無条件公開が望

ましいが、必ずしもこれに拘泥しない。特に各省庁、各データ作成機関の政策、ポリシーとの整合性が問われる場合は、柔軟に対応し、データ作成機関の意向を十分に尊重する。一方、オープンイノベーションという言葉に代表されるように、これからの学問、産業の発展には多くの分野にまたがる集合知の結集が重要であり、そのための情報共有の土壌作りも合わせて進めていく必要がある。

（５）統合データベースの運営資金と他機関への資金提供

ライフサイエンス分野のデータは、種類、量とも急速に増大している。そのため、統合データベースの将来的な持続可能性を担保するための仕組みを検討する必要がある。今回行った議論は、おもに JST の予算による統合データベースの運営に関するものであったが、統合データベースの性格上、近い将来には各省全体で必要な予算措置がとられることが望ましい。それにより名実共に、我が国におけるライフサイエンスの統合データベースになるように目指す。なお、この予算は、あくまで統合データベースの運営及びその発展、向上のために使われるべきであり、他の研究機関への資金援助は、本センターの運営、本データベースの質的向上等に必要とされるものに限る。一方、統合データベースは質の高い個別データベースの上に成立するものであり、個別データベースを構築、維持するための仕組みは別途考慮されることが望ましい。

（６）名称とその意義

現在のところ「統合データベース・センター」は仮称であるが、統合が完成した将来は、我が国のナショナルデータベース・センターという形で名称の変更を考えるべきであろう。統合データベースはライフサイエンス分野の国力の象徴であり、米国の NCBI、欧州の EBI に相当する機関の設立が望まれる。そのために第一段階(H23-H25)として JST に作るセンターもそのような方向を指向する組織としてふさわしい体制と機能をもったものでなければならない。

（７）JST ライフサイエンス分野統合データベース・センターの当面の具体像

①利用者：当面、大学や企業の研究者、技術者といった専門家を対象とする

②対応すべきニーズ：

・データベースや解析ツールの所在情報、利用法の提供

・散在するデータベースやデータのシームレスな利用環境の提供

・大量データや維持できないデータベースに対する継続的公開基盤の提供

・データベースの活用によるイノベーションとそのための情報環境整備

③持つべき機能：

・戦略立案機能：統合すべき対象と統合に至る手順、問題点、データ共有ガイドラ

イン等の調査・検討、関連府省を含めた国内外関連機関との連携調整およびアジア諸国も含めた国際連携の実施等

・ポータルサイト：データベース等カタログの整備、横断検索の高機能化および知識発見につながる高度な検索・解析機能の提供、ならびに各省データベースとのネットワーク構築

・統合データベース構築：統合化に必要な標準化への取り組み、目的、用途ごとのデータベース統合化および高品質なアノテーションの実施等のデータベース品質管理

・データベース公開基盤：計測技術の進歩に応じた新規データ公開基盤の構築ならびに維持困難な、あるいは、共有可能なデータベースの受け入れとアーカイブサイト、レポジトリサイトの維持・更新

・研究開発機能：高度な検索等の統合的利用を実現するためのシームレスなデータベース構築・利用技術の開発、最先端の情報環境の研究、分子データ、文献データ活用のためのインフォマティクス技術の開発およびインデックス、辞書、データフォーマットなどの開発

・システムの構築・維持・管理：ポータルサイトや統合データベース、アーカイブサイト等の構築によるサービス提供ならびに統合化に必要な研究開発を支障なく行うための基盤システム構築

・ファンディング機能：現 JST バイオインフォマティクス推進センターのデータベース高度化・標準化プログラムなど統合データベースに寄与する事業への資金提供

・人材育成：統合データベース・センターの自律的発展を担保できる、あるいは個々のデータベース構築を支援できる人材育成機能と育成された人材のキャリアパスの構築

・広報・普及啓発：データベースや解析ツールの利用法に関するメディアの作成ならびに講習会等による提供サービスの周知、普及

・評価機能：アクセス数など客観的・定量的なデータ提供、ならびにデータベースの有効な評価方法についての検討

（８）JST（第一段階 H23-H25）のセンターを作るための議論を超えた部分の検討

・本委員会で議論しまとめたことの中には、JST（第一段階 H23-H25）のセンターを作るための議論を超えたものも多々含まれる。これらに関しては、また、JST のセンターのあり方そのものについても、より上位の委員会で各府省の方々および関係者の方の参加を得て更なる検討を加えることとする。

・とくに国のファンディングによるデータ等についての公開、アーカイブなどの原則（プリンシプル）、持続可能性のための予算の確保等については、別途設置され

るところで、より詳細な議論を行う。

２．主な論点と意見

以下、平成２１年９月に（独）科学技術振興機構に設置された「ライフサイエンス分野統合データベースセンター設置準備委員会」での主な論点と出された意見についてまとめる。

（１）目標・理想

日本の府省に亘って提供されているデータベースを統合し、日本で生産するデータの価値を最大化することにより、日本のユーザ、更には世界のユーザに貢献できる日本が誇るべきデータベースを作っていくのが目標となる。最も重要なことは、ユーザの立場に立った使い勝手のよいものを作ることである。すなわち、制約のないデータ共有と操作感の統一を実現し、ライフサイエンス分野の知識発見を支援できる環境の構築である。また、参加することにより個々のデータベースの利用価値が上がり、多くの人から認められるデータベースを作ることである。

（２）施策として必要な理由

ライフサイエンス研究の成果を最大限利用するためにはデータを共有化し、多くの人が利用できるようにすることが必要である。情報だけではなく、バイオリソースなどの基盤部分を最大限活用できるようにすることは、公的資金の投入を最大の成果へとつなぐ方策であることから、必要な施策である。これにより、ライフサイエンスの新たな展開とバイオ産業の国際的競争力向上に貢献できる。また、世界の先進国として自国で生産したデータには最後まで責任を持つべきであること、さらには現在のライフサイエンスにおけるデータベースの重要性に鑑み、統合データベースは府省にとらわれない存在として国が継続的に支援すべきものである。

（３）データやデータベースの公開や共有に関する事項

データ公開基準については前記のとおりであるが、データやデータベースの公開や共有に関して以下の議論が行われた。

①研究資金による違い：マッチングファンドやコンソーシアムにより作られたデータやデータベースでは、一定期間会員に対する優先的利用権を付与すべきだが、その後の公開は可能。

②研究のタイプによる違い：データ産出型のプロジェクト研究においては、プロジェクトに積極的に関与することによりデータ共有が実現しやすくなる。また、データの早期公開を巡っては海外でも問題がないわけではなく、難しい面もあるが、知財獲得という国益と学問、産業の総体的な発展という国益のバランスをよく考える必要があ

る。個別研究においては、データを生産する研究者の一次利用に配慮を十分したうえで、皆で共有できる二次利用に至る手順を検討すべきである。

③データの種類による違い：配列データや蛋白質立体構造データについては、出版社との連携のもとに公開体制が整備されている。臨床データは、個人情報が含まれるなどの特異性があるが、その利用に関するニーズは高く、総合科学技術会議での検討が待たれる。また、疾患研究に向けた統合データベースの運営については、個別構築サイトとデータベース・センターの役割分担に関する検討が必要である。

④新旧データおよびデータベース資産への対応：過去のデータベース資産の共有が生きる分野は、データ生産効率の上がりにくい画像や論文である。一方で、次世代シーケンサーなど測定技術の発達の著しい分野では、新世代のデータ集積、共有が、旧世代データの統合に優先する。

（４）JST ライフサイエンス分野統合データベース・センターの性格

平成23年度に設立される予定の上記センターが、当面どのような利用者による、どのようなニーズに対しての、どのような目的やレベルでの活用を主として想定しているのかを明確にするという観点で、以下の議論がなされた。

①利用者：大学や企業の研究者、技術者といった専門家向けか、マスメディアなども含めた一般向けかという議論がなされ、後者もある面で重要であるものの、当面、専門家を対象とすべきとの結論になった。

②ニーズ：データベースや解析ツールの存在が分かりにくい。データベースやデータが散在していて利用しにくい。大量の未公開データや維持できないデータベースに対する継続的な公開基盤が必要など。

③目的：大量のデータが産出されるプロジェクトやデータから知見を見出す研究へと変化する中で、データ共有による公的資金の有効性の向上、あるいは散在するデータベース及び解析ツールの利用の利便性向上といった点が挙げられた。

④必要性：上記の大きな観点に加え、可能な限り重複を排除するとともに、散在する多様なデータベースをシームレスに利用できる環境は必須、あるいは整備されていないデータ共有の受け皿を整備することによりデータの散逸を防ぎ、また、利便性を向上することにより研究の無駄を省くことが必要等の議論があった。

（５）JST ライフサイエンス分野統合データベース・センターの機能

①データトレーサビリティの確保の重要性に関して以下の指摘がなされた。

・他のデータと併せて利用することを可能とする。

・利用者は、どこのデータベースのどういうエントリーから得ているかの情報を必要としている。

・データをセンターに寄託するインセンティブとして、論文の根拠としてのデータトレーサビリティを確保することにより、論文発表後の検証の際に研究者や法人を保護できる。
・データ利用に関する取り扱いを明示してあることが利用の可否判断に重要な事項である。

②サービスとして必要な機能として以下の指摘がなされた。
・どこにどのようなデータベースやツールがあるかわからないという問題点はポータルサイトを整備することにより解決する。
・各大学、各機関に散在しているデータベースや解析ソフトウエアを1ヶ所で案内・管理する one-stop-shop のような機能が望まれる。
・データベースカタログや横断検索は便利であるが、簡単な操作で柔軟に情報を一覧表にまとめられる機能が提供されると利用度が上がる。
・データ一括ダウンロードサービスは、産業界等の知的所有権の保護等に対応した利用者の所内利用への対応や、大量データを組み合わせて研究する研究者への対応として有用である。
・利用者が探したいデータについて登録しておくと、該当する新しいデータについて通知してくれるサービスも便利。
・次世代シーケンサーのデータなど、海外からデータダウンロードできないほどのデータ量の存在がある。こうした大量データを提供することも重要である。

③センターがもつべき研究機能についての考え方は、絶えず新しいものに対応するためにも、また、若手研究者の積極的な参加のためにも必要であるとの議論の中で、以下の指摘がなされた。
・データベースサービスの向上に特化した研究を主体とすることが望ましい。
・研究機能は、若い研究者にインセンティブを与えるものであるが、その研究の対象はバイオインフォマティクス全般ではなく限られるべきである。その対象は、サービス機関としてのセンターのアップグレードや、世界の最先端的な方法でのデータの供給や収集・統合等である。
・しかし、日本のバイオインフォマティクスの現状を考えると、プロジェクト等との共同研究や研究に使えるあるいは研究に使われた解析ツールを生み出すようなバイオインフォマティクスの研究が同時に行われることは重要ではないか。

④データベースの統合化については、以下のような議論が行われた。
・個々のデータベースを作った後、統合するということが主な機能。
・どういうものを作るのかという明確なイメージを作り上げる。例えば、中心となるデータベースを設定し、世界標準（NCBI など）との連続性を確保する。また、単純に個々のデータベースを集めてくるということではなく、それぞれのデータベースを相互にリンクしたり、リンク先に有用な情報があるというような情報を維持する

ことも重要

・NCBI の有用性は、NCBI を見に行けば、そこから全部データがつながるような立派なコンテンツがあること。

・対象とするデータやデータベースについては、現実的に、段階的に対象データを広げていくべきとの指摘がなされた。検討すべき対象は以下のとおり。

－分子から個体、生態とか環境まで
－実験データ、ファクトデータ
－文献、特許等
－文献の図表
－論文のサプリメンタルインフォメーション
－扱う生物種
－医薬品等の化合物、マテリアルのカタログ等

⑤収載・維持すべき情報として話題となった例

・センターとしてコアになるデータベースが必要
・日本固有のコンテンツ、世界がやっていないデータ
・充分に使い切れていない情報のデータベース化

⑥連携を検討すべき国内関係機関

・塩基配列やバイオリソースのカタログであれば国立遺伝学研究所
・文献や特許データであれば科学技術振興機構や国立情報学研究所
・多くのデータを有している大きな研究機関としては理化学研究所

⑦国際連携に関する考え方

・日本のデータを公開するだけでも世界に対する窓口となる。
・アジアや中国に向けた方策あるいは視点をもって方針を定めるべきであるが、いずれにせよスピードが必要。

⑧ファンディングに関する考え方は上記のとおりであるが、以下のような議論がなされた。

・個々のデータベースは研究レベルであり、別のファンディングで支援すべきである。ただし、統合データベースに足りないところがあった場合、統合データベースに参画するという条件でファンディングを行う場合もあってよい。
・JST バイオインフォマティクス推進センターのデータベース高度化・標準化プログラムなど、JST バイオインフォマティクス推進センターと文科省統合データベースプロジェクトとの一本化という観点からのファンディングは機能の一部である。
・個々のデータベースであっても統合データベースであってもナショナルデータベースに相応しい、よいデータベースにはファンディングしてもよい。
・データを公開したい人がデータの公開技術を必ずしも持っているわけではないので、データを公開したい人とデータを利用したい人をつなぐ仕組みが必要。その

際、データ公開に携わる人（公開者と利用者をつなぐ人）がいるところへのファンディングをするようなデザインが必要。

⑨人材育成については、自前の人材育成機能と外部委託のバランスが重要との指摘がなされた。また、キャリアパスについては、非常に高い業務をやったからこそ、その人材へのニーズが生まれるので、そうしたニーズを生むようなレベルのセンターを作る必要があるとの指摘もなされた。

⑩統合データベース・センター評価について、以下のような議論がなされた。

・客観的・定量的なデータ（例えばデータベースへのアクセス数）を公表すると共に、有効な評価方法についても検討する必要がある。

・客観的な評価としては、利用率や論文の採択のされ方、ヒット数も候補。

・どれだけ利用されたか－利用回数、利用時間、ダウンロード回数や、利用されてこういうことが導き出された－論文等に引用された回数も考えられる。

３．付録

（１）NCBI や EBI などとの比較

NCBI は、DB の統合化を目標に作られたわけではない。ユーザやプロジェクトの要望を入れつつ機能を拡充した結果が現在の NCBI。EMBL/EBI は少し異なり、複数の DB を統合化させた様な趣があるが、Ensembl については、やはりゲノムコミュニティからの要望を反映したもの。UCSC はゲノムに特化しているが、これは自発的に作成したものが NIH に取り込まれたもの。

NCBI,EMBL/EBI ともにデータ生産拠点と密着している。PubMed も文献データ生産の拠点。

参考:

○ 国際的データベース

NCBI、EBI 共に核酸データベース（GeneBank、EMBL）開発提供を主務として、そこから派生する情報を加工し、独自のリソースを確立することで、国際的な評価を得ている。従って、主幹となるデータベースの有効利用を目指した「結果」。

○ NCBI が世界の主流

NCBI は、分子生物学分野に特化した情報機関であり、他分野における計算機資源及び技術の開発については、NCBC (The National Centers for Biomedical Computing)等において、分担して行われている。分子生物学者以外の研究者が NCBI を利用するとすれば、PubMed が目的であり、ソースであるデータベース Medline の開発は NLM が担当。→ NCBI のユーザは分子生物学データの利用者

○ 国家レベルの制度、方針

NCBI のデータベースやデータ利用に関わる権利や倫理等の政策的な問題は、連邦政府および NIH の基本方針によるものであり、NCBI が決定しているのではない。EBI においては、各プロジェクトがその協力機関、出資機関等と個別に検討している。（JST 追加）

○ データとデータベースの区別。

公的データベースにデポジットする情報は塩基配列データ等でフリーアクセスが実現している。米国では公的資金により助成された研究ではデータ、リソースの公開を要求している場合があり、その場合、フリーアクセスが実現されている。一方、データベースは各々異なるライセンスを提示しておりフリーアクセスとなっていないこともある。（JST 追加）

（２）米国でのインタビュー記録　（文責　科学技術振興機構）

米国訪問者

・大石道夫（かずさ DNA 研理事長、科学技術振興機構 ライフサイエンス分野統合デ

ータベースセンター設置準備委員会委員長)、
・・大濱隆司(科学技術振興機構ワシントン事務所長)
・・平川美夏(科学技術振興機構　研究員)
・高木利久(情報・システム研究機構ライフサイエンス統合データベースセンター　センター　センター長)
・大久保公策(情報・システム研究機構　国立遺伝学研究所教授)

1)ナショナルアカデミーズ　研究データ・情報委員会　ポール・ウーリア氏訪問
場所　ナショナルアカデミーズビルディング
日時　11 月 16 日 10:00ー11:30
面会者　Paul F. Uhlir　(Director, Board on Research Data and Information)

- ナショナルアカデミーズは、全米科学アカデミー(The National Academy of Science (NAS))、全米工学アカデミー (The National Academy of Engineering(NAE))、医学機構(The Institute of Medicine(IOM))から構成され、各々に個人の研究者が所属する。これらについてThe National Research Council(NRC)が管理や政策決定を行う。
- 研究成果の公開に関しては、個々の研究者の自主性に任せられており、アカデミーが強制することはない。
- 公開に関するポリシーの違いは、ファンドの提供機関やファンドの質でも違うし、提供機関内部で実施するか、資金を提供して外部で実施するかによっても異なる。一般に、外部に提供する場合の方が、強制力は弱い。
- 宇宙や気候、環境などのデータは、測定を行う機関や研究者が解析を進める以前に、迅速に公開することが求められるが、一般には論文発表が優先されるのではないか。
- 研究者が研究データを公開したがらない傾向は、ライフサイエンス分野で見受けられるが、ファンドはオープンになっているので、周囲の目や圧力によって隠してはおけないだろう。
- 公的機関と個人の両方からファンドを得ている場合は、公開が制限されることが多い。
- 公的機関自身の情報公開については、オバマ政権になってより促進している。
- 米国は、データの公開と共有によるメリットを認めているが、世界的な傾向ではなく、欧州はむしろ国益や国策による保護を優先し、国外にデータを公開しないし、中国はトップダウンでコントロールされているだろう。
- インターネットの普及によって急速にデジタルデータ共有は進んでおり、量効果と処

理技術開発によって一層利益をもたらすと考えられる。

[参考文献]

Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation (2008).
http://brtf.sdsc.edu/

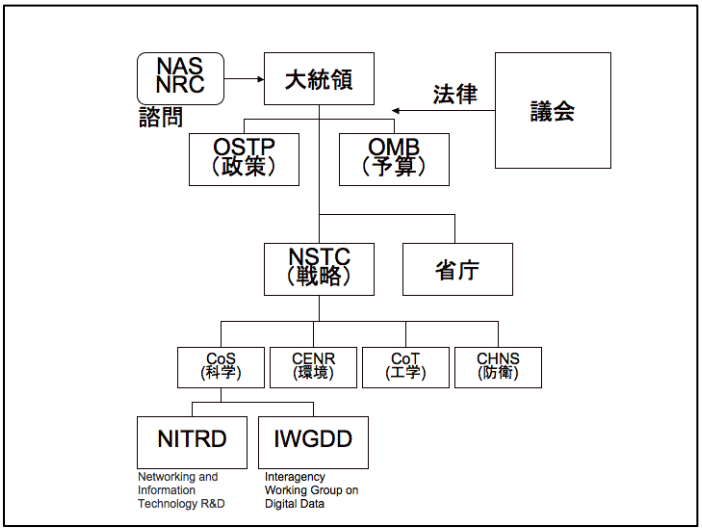
The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies (2009)
http://www.nap.edu/openbook.php?record_id=12687&page=R1

米国動向報告: The National Academies(ナショナルアカデミーズ)の概要
科学技術振興機構 研究開発戦略センター海外動向グループ (2007)
http://crds.jst.go.jp/kaigai/report/TR/AM/US20071001.pdf

2.)国立医学図書館健康情報プログラム展開室 エリオット・シーゲル博士、CENDI ボニー・キャロル エグゼクティブディレクター訪問
場所 国立医学図書館(NLM)
日時 11 月 16 日 14:00〜15:30
面会者 Elliot R. Siegel, Ph.D.(Associate Director for Health Information Programs Development)
Bonnie C. Carroll (Executive Director, CENDI Federal STI Manager's Group)

- 科学技術情報の公開促進について、ナショナルアカデミーズはボトムアップであるが、ここではトップダウンの協議機関について組織図(下図)を元に説明があった。
- IWGDD は、省庁など 22 機関が参加するデジタルデータの公開と共有に関するワーキンググループであり、データベース化等のインフラ整備に関わる検討がなされているが、省庁間の違いを容認しており強制力はなく、調査報告とアドバイスを主に行っている。
- CENDI(注)は、13 の機関から科学技術担当者がボランティアで参加する中立的な組織であり、国から予算措置はされておらず、参加機関が分担で運営している。
- 現状では公開を前提としているが必ずしも進んではおらず、むしろ商業ベースの方が柔軟な対応を取ることもあるので、正しいデータの所有権を認め、共有することが必要である。
- データベースセンターを作ることに関しては、保管機能ならば可能性はあるが、一カ

所で所有するという考え方は受け入れられにくい。むしろ平等な立場で、データを公開、共有し、相互に支援しあう形が望ましいだろう。



[参考文献]

Harnessing the Power of Digital Data for Science and Society （January 2009）
http://www.nitrd.gov/about/harnessing_power_web.pdf
CENDI 米国連邦政府 科学技術情報シニア・マネージャーによる省庁間グループ
情報管理 Vol. 48 (2005) , No. 3 p.144-148 前田知子 （科学技術振興機構）

〔CENDI（注）〕
CENDI の名称の由来は当初より参画している4機関の頭文字を組み合わせたもの
COMMERCE － National Technical Information Service (NTIS)
ENERGY – Office of Scientific and Technical Information (OSTI)
NASA － Scientific and Technical Information Program (STI)
DEFENSE － Defense Technical Information Center (DTIC)
I NFORMATION

3）国立バイオテクノロジー情報センター デビット・リップマン所長訪問
場所 国立バイオテクノロジー情報センター (NCBI)
日時 11 月 17 日 13:00～15:00
面会者 David J. Lipman, M.D. (Director, NCBI, NLM, NIH)

・ NCBI のデータ利用率について、分子生物学データベースでは、米国（39%）、中国

（11%）、日本（8%）、ドイツ（7%）、英国、インド（5%）の順であり、PubMed 文献のページ数では、米国（36%）、日本(7%)、英国、ドイツ（5%）、中国、カナダ（4%）と続いており、日本はデータベース活用に関心が高い。

- NCBI は、配列データの収集が中心であり、将来それ以外のデータも対象とするかはわからないが、共通の形式で組織化できるデータを対象と考えている。イメージデータは対象にしてこなかったが、テキストのデータでは実績がある。新規データや他機関のデータについては、任されれば引き受けざるえないので要求に応じて努力するが、政府内では頼まれやすい位置にいると認識している。
- すでにデータベースとして流通し、確立しているデータを一カ所に集中する目的で集めるつもりはない。
- データの統合、集約を行うときは、共通のキーに注目して統合し、共通でないデータを失わないことが必要である。
- 日本のデータベースの発展のためには、他にはないユニークなリソース（例えば KEGG）で必要性を示す、日本語文献など日本固有のデジタルデータをもっと組織的に扱い公開する、データベースを利用した研究を生物学者にもっと推奨し、コストパフォーマンスの高い研究を行う基盤を作る、などが考えられる。
- データベースの善し悪しの評価は、全ての Web log（5TB/日）を解析する、数十箇所の実際にヘビーユーザである研究室に使い方、意見などをインタビューして分析する。個人の意見は、偏りがちであるが、突っ込んだ内容で、なるべく数多く、何年にも亘って続けることで、方針が明らかになり改善につながる。
- NCBI の雇員のうち 20%が基礎研究者で、残りは開発者である。いずれも 4 年ごとにレビューして継続を判定するが、人材がリソースであり、長期に勤めることを推奨している。
- データの共有については、NCBI では判断せず資金提供機関の判断に任せている。公開は進んでいるが、伝統的に論文が出るまではデータを公開しない機関や分野も残っている。
- データベースは組織化されるべきであり、コンテキスト（実験条件など）に依存するデータは、測定の基準が異なるなど、集めただけでは価値が活かされない。個別研究のデータ公開や共有とは異なる面がある。
- 今後の高速シークエンサーのデータについては、病原菌の株による配列の違いの比較などが、データベースとしては意味があると考える。
- 最近の興味は、Open publishing で Google を活用したオンラインジャーナル作成システムを開発した。小さなコミュニティでもコストがかからずにジャーナルが発行でき、その分野の専門家が集まってレビューすることで、迅速に信頼性の高い文献が蓄積できる。また Google の仕組みで検索も可能であり、メリットが大きい。
  日本語の解説: http://johokanri.jp/stiupdates/education/2009/08/003517.html

4）国立医学図書館政策展開室　ジェリー・シーハン氏訪問
場所　国立医学図書館（NLM）
日時　11 月 17 日 15:30～17:30
面会者　Jerry R. Sheehan （Assistant Director of Policy Development）

- Policy Development は NLM の一部署であるが、NIH 全体の政策を検討し、表現を決定している。
- データの公開については規則を決める方法もあるが、データの性質（個人情報等）や技術的な問題から自ずと決まることも多い。
- NIH ではデータ公開ポリシーを規定しているが、実際はペナルティが課されることはなく、論文が出せない研究は自然と予算が減ることになるだけである。
- GenBank などのデータ効果ポリシーと連動したデータベースも存在するが、一般には文献が発表媒体であり、それ以上は求めていない。
- 省庁など特定の目的があり、利害が異なるところでは、公開に関する判断も当然異なり、米国でもそれでまかり通っている。
- 機関のデータは、オバマ政権になって公開が促進されているが、一般の研究データについてはケースバイケースである。
- 臨床データなどは公開されることで結果の信頼性が増すが、資金提供機関や受給者も多様であり一定の判断ではない。
- データを公開しない理由はいろいろあり、国防や安全保障、個人情報などは隠すことを容認できるデータである。公開基準については、データの種類ではなく、データを利用する際の意味に注目して分類することが有効であろう。

[参考サイト]
NIH Data Sharing Policy
http://grants.nih.gov/grants/policy/data_sharing/

National Institute of Heath Public Access
http://publicaccess.nih.gov/

Genome-Wide Association Studies (GWAS)
http://grants.nih.gov/grants/gwas/

ClinicalTrials.gov
http://www.clinicaltrials.gov/

（３）ライフサイエンス分野統合データベースセンター設置準備委員会の目的

（JSTライフサイエンス分野統合データベースセンター設置準備委員会規則より抜粋）　（平成 21 年6月 24 日　平成 21 年規則第 109 号）

（目的及び設置）

第1条　文部科学省が進めるデータベースの統合・維持・運用を図るため、我が国として目指すべき統合データベースに相応しいセンター機能を有する組織として整備予定のライフサイエンス分野統合データベースセンターの設置準備に関し調査審議するため、組織規程（平成15年規程第2号）第9条の規定に基づき、機構にライフサイエンス分野統合データベースセンター設置準備委員会（以下「委員会」という。）を置く。

（任務）

第2条　委員会は、次に掲げる事項について調査審議し、理事長に対して提言する。

（1）　ライフサイエンス分野統合データベースセンターの機能に関する事項

（2）　その他設置準備に関する事項

（４）ライフサイエンス分野統合データベースセンター設置準備委員会委員名簿

◎大石　道夫　　（財）かずさディー・エヌ・エー研究所　理事長

浅井　潔　　東京大学 大学院新領域創成科学研究科　教授

有田　正規　　東京大学 大学院新領域創成科学研究科　准教授

今西　規　　（独）産業技術総合研究所 バイオメディシナル情報研究センター 研究チーム長

大久保 公策　情報・システム研究機構 国立遺伝学研究所　教授

五斗　進　　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　准教授

高木　利久　　情報・システム研究機構 ライフサイエンス統合データベースセンターセンター センター長・教授

田畑　哲之　　（財）かずさディー・エヌ・エー研究所　副所長

豊田　哲郎　　（独）理化学研究所横浜研究所 生命情報基盤研究部門　部門長

中村　春木　　大阪大学 蛋白質研究所附属プロテオミクス総合研究センター　教授

中村　保一　　情報・システム研究機構　国立遺伝学研究所　教授

林　哲也　　宮崎大学　フロンティア科学実験総合センター　教授

藤山　秋佐夫　情報・システム研究機構　国立情報学研究所　教授

蓑島　伸生　　浜松医科大学　光量子医学研究センター　教授

山崎　由紀子　情報・システム研究機構　国立遺伝学研究所　准教授

計：15名（敬称略　委員五十音順）

◎は委員長

オブザーバー

内閣府

文部科学省

厚生労働省

農林水産省

経済産業省

（５）分科会委員名簿

1）データベース利用者分科会　委員名簿

◎大石　道夫　　（財）かずさディー・エヌ・エー研究所　理事長

浅井　潔　　東京大学　大学院新領域創成科学研究科　教授

今西　規　　（独）産業技術総合研究所　バイオメディシナル情報研究センター
研究チーム長

五斗　進　　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　准教授

高木　利久　情報・システム研究機構　ライフサイエンス統合データベースセンター
センター長・教授

中村　保一　　情報・システム研究機構　国立遺伝学研究所　教授

林　哲也　　宮崎大学 フロンティア科学実験総合センター　教授

蓑島　伸生　　浜松医科大学 光量子医学研究センター　教授

山崎 由紀子　情報・システム研究機構 国立遺伝学研究所　准教授

磯野　克己　　(財)かずさディー・エヌ・エー研究所　常務理事

伊藤　武彦　　東京工業大学　大学院生命理工学研究科　教授

黒川　顕　　東京工業大学　大学院生命理工学研究科　教授

小谷　秀仁　万有製薬株式会社　執行役員

角田　達彦　(独)理化学研究所　ゲノム医科学研究センター　チームリーダー

中井　謙太　東京大学　医科学研究所　教授

中川　博之　大日本住友製薬株式会社　ゲノム科学研究所　スペシャリスト

長洲　毅志　エーザイ株式会社　理事・研究開発担当付　部長

菱垣　晴次　大塚製薬株式会社　基盤技術研究所　所長（ご欠席）

（敬称略　本委員会委員からの分科会委員、分科会委員それぞれ五十音順）

◎は主査

2)データ生産者分科会　委員名簿

有田　正規　　東京大学 大学院新領域創成科学研究科　准教授

大石　道夫　　(財)かずさディー・エヌ・エー研究所　理事長

大久保 公策　　情報・システム研究機構 国立遺伝学研究所　教授

◎高木　利久　　情報・システム研究機構 ライフサイエンス統合データベースセンター センター長・教授

田畑　哲之　(財)かずさディー・エヌ・エー研究所　副所長

豊田　哲郎　(独)理化学研究所横浜研究所 生命情報基盤研究部門　部門長

中村　春木　　大阪大学 蛋白質研究所附属プロテオミクス総合研究センター　教授

藤山 秋佐夫　情報・システム研究機構 国立情報学研究所　教授

伊藤　隆司　　東京大学　大学院理学系研究科生物化学専攻　教授

河合　純　　（独）理化学研究所　オミックス基盤研究領域　プロジェクトディレクター

古崎　晃司　大阪大学　産業科学研究所　　准教授

後藤　信哉　東海大学　医学部内科学系　教授

菅野　純夫　東京大学　大学院新領域創成科学研究科　教授

夏目　徹　　（独）産業技術総合研究所　バイオメディシナル情報研究センター　細胞システム制御解析チーム　チーム長

若槻　壮市　高エネルギー加速器研究機構

物質構造科学研究所　構造生物学研究センター　センター長

（敬称略　本委員会委員からの分科会委員、分科会委員それぞれ五十音順）

◎は主査

オブザーバー

吉田　輝彦　　国立がんセンター研究所　腫瘍ゲノム解析・情報研究部　部長

（６）委員会開催記録

2009年　9月15日（火）　第1回委員会
10月15日（木）　第2回委員会
11月13日（金）　第3回委員会

12月　4日（金）　データベース利用者分科会
データ生産者分科会

2010年　4月23日（金）　第4回委員会

4．参考資料

（1）データベース利用者分科会概要

ライフサイエンス分野統合データベースセンター設置準備委員会

データベース利用者分科会　概要

1．　日時　　平成21年12月4日（金曜日） 10時00分～12時00分

2．　場所　　東京グリーンパレス内小会議室 「さくら」

3．　出席者

（委　員）　本委員会委員

大石主査、今西委員、高木委員、中村保一委員、蓑島委員、

（ご欠席：浅井委員、五斗委員、林委員、山崎委員）

分科会委員

磯野委員、伊藤委員、黒川委員、小谷委員、角田委員、中井委員、中川委員、長洲委員、（ご欠席：菱垣委員）

（事務局） 広瀬理事、門田審議役、大倉部長、菊池次長、黒田課長

4．　議事

（1）本分科会について

（2）データベース利用者分科会 論点

（3）意見交換

（4）その他

5．　配布資料

資料1 ライフサイエンス分野統合データベースセンター設置準備委員会

データベース利用者分科会委員名簿

資料2 ライフサイエンス分野統合データベースセンター設置準備委員会規則

資料3 JST統合データベースセンター（仮称）設置準備委員会分科会（DB利用者）論点

（ROIS・DBCLS　高木センター長 作成資料）

資料4 ライフサイエンス分野統合データベースセンター設置準備委員会

第1回、第2回、第3回 議論概要（案）（2009.11.24 JST作成）

参考資料①

「ライフサイエンスデータベースの統合・維持・運用のあり方」

（H21.1文部科学省科学技術・学術審議会研究計画・評価分科会ライフサイエンス委員会ライフサイエンス情報基盤整備作業部会）

参考資料②

「統合データベースタスクフォース報告書」
（H21.4総合科学技術会議　基本政策推進専門調査会 ライフサイエンス統合データベースタスクフォース）
参考資料③

ライフサイエンス分野統合データベースセンター設置準備委員会委員名簿

〇高木委員より下記の論点（データベース利用者分科会向け）を事前に提示し、分科会委員より意見をいただく形式で委員会は進められた。

- どういうサービスを提供すべきか？
    - どういう検索サービス、解析サービスが望まれるか？
- どういう種類のデータまでを対象とすべきか？
    - 分子から個体、生態、環境まで？微生物からヒトまで？
    - 生データだけでなく、文献、特許、マテリアルなども？
- センターでの研究開発、人材育成のあり方は？
    - DB 構築に特化したものだけか？バイオインフォ全般か？
    - 知財をとるような共同研究（DB 利用者との）もすべきか？
    - センターでどういう人材を育ててほしいか？
- データ共有のあり方、進め方？
    - どのような共有ポリシーが望まれるか？
- 組織、体制のあり方は？
    - 永続的で開かれた運営にはどういう点に配慮すべきか？
    - 利用者の意見を反映する仕組みは？DB の評価はどうあるべきか？

〇製薬企業
【どういうサービスを提供すべきか？】
・データベースの利用の仕方
・・基本は社内でデータあるいはデータベースをダウンロードし、社内データベースを構築し、それを検索に使用する。
・・専門家には日本語は不要。きちっと使えるようなデータがあればよい。クエリーを投げるようなことは稀であり、こういう部分についてダウンロードをできるシステムが必要。
・・特許を出していないものを外に投げて解析していただくことは難しい。
・・どこの誰がどういうデータを検索したという記録が残ることは望んでいない。
・・文献や特許は既に専門の有料データベースが存在し、それを利用する。
・・ヒトに関するデータベースも有償のサービスがかなり整備されており、それを利用する。
・サービスで期待するもの
・・基本的には社内で対応がしにくい検索、もしくは解析

・・どうやってそこに行けばよいか。ワンストップショップ、どこか1つで管理すること。
・・各大学、各機関で少し毛色の変わった、例えばアルゴリズムを使ったデータを使った解析のソフトについて、どこか1つの場所で同じ結果の解析を違う方法で行って、その結果を比べることができること。
・・データベースカタログや横断検索によって、知らなかった情報のありかがわかるようになった。所在がわかると、直接そこにその情報を見にいく。必ずしも統合データベースを経由しない。
・・個別データベースを直接参照するよりも、統合データベースを参照、経由したほうが非常に便利であれば（例えば情報の取り出し方が便利である、データを統合しているデータベースとなれば）、統合データベースを使用することになるのではないか。
・・簡単な操作でいろいろなデータベースについて、このデータベースのここ、このデータベースのここ、という形で所在を一覧表にまとめられる機能。
・権利関係について
・・権利関係は、明確に記載。日本のデータベースはどちらかというと何も書いていないことが多いので、明確に記載してほしい。有料か無料かは企業にとって問題はない。使用して良いか否かがわからないことが、企業としては困る。

【どういう種類のデータまでを対象とすべきか？】
・一番の関心は、分子、生体、ヒト、それから動物実験。
・ヒトのデータベースに関しては、有料のもの、海外のベンチャーが提供しているものも存在するため、かなり充実している。
・日本の独自性があるものとして考えられるものは、実験動物のデータ、臨床データ、生体の医療データ、疾病データなど。これらを充実していただけると、かなりありがたい。
・生データ、文献、特許、マテリアルについて、企業内特許のデータベース、日本の国から提供されているもの、海外のベンダーから提供されているもの、すべて社内で利用できる。
・文献とか特許は、真剣に検索するならそれぞれの専門のサイトで漏れのないような検索をしないといけない。有料でかまわない。
・分子から個体、ヒトを最初の視点にして微生物まで。利用者によって必要とするデータは異なり、「ここまで」ということを言うのは難しい。
・とりあえずバイオのデータだけでよいということで、バイオに関しては、可能な限りすべてのものが含まれるべき。

【センターでの研究開発、人材育成のあり方は？】
・人材育成
・・データベースの整理や管理は、内部の人材を使っていない。日本もアメリカも外注業者に委託。
・・人材は、データベースだけに特化した方よりは、バイオインフォマティクス全般に精通した知識

を持たれた方がよい。ITの技術だけではなくて、医療とか、それから生物一般の知識を持った方。

・・データベースの構築だけではなくて、データを使って何かを導き出すという研究も一緒に行われていれば、よりよいデータベースの提供につながる。

・・企業は、バイオインフォマティシャンとかドライの研究者がいないということで外に投げているということが、逆にバイオインフォマティクスの発展を阻害している。

・・人材が育ち、そういう人はぜひ企業に欲しくて、中で解析してほしいというようなポジティブな螺旋が回れば、変わってくる。

・・バイオインフォマティシャンは、モレキュラー・バイオロジーとか薬理とかというウエットの人間と、完全にドライな、かなり正しいアルゴリズムを作れるようなSE、プログラマーといったところを橋渡しする人間、いわゆるトランスレーターみたいな、そういう位置づけであって、両方にすごく優れているということではないのではないか。

・研究開発

・・共同研究に関して、働く人のモチベーションを高めるのであれば、推進してはどうか。

【データ共有のあり方、進め方？】

・企業で実験・研究したデータの公開は、難しい。

・国のプロジェクト、コンソーシアムでマッチングファンド（企業もお金を払って）共同で作ったデータは、ある程度、先行利用し、例えば3カ月、その後に公開というのは、可能な範囲。

・コンソーシアムなり、会員から費用を集めて作ったケースというのは、ある一定期間、会員に対してプライオリティが与えられるべきだが、その後は一般に公開すべき。

・国のプロジェクトは、データ・ドリブン・サイエンスという言葉もあるように、非常に多くのデータを使って成果を出すという形になっていて、そのデータ自身の 100%の利用はもともと不可能というようなことがあるので、そのデータを有効利用するという意味で、共有するというのは当然。

・データの出所を明確にするということを守る限り、できる限りオープンに使えるような形であれば、このデータだけ孤立するというものではなく、さらにほかのデータとも一緒にして使っていくということができるだろう。使う側としては、どこのデータベースのこういうエントリーから持ってきたのだということを、情報として持っておかないといけない。

【組織、体制のあり方は？】

・利用者の意見反映

・・１年ごとの参加の更新をするというコンソーシアムは、比較的意見が反映されやすい。なぜなら、１年間でその意見が反映されないと、その翌年に企業が参加しないため、そこの部分でメカニズム的によく意見が反映される。

・・外部への情報発信を継続的に行う。

・・データベースに関する要望は、即座にそこに何か書き込めるような仕掛けがあるとよい。「こう

いう要望に対してこうなりました」とか、そういったものを逆に公開していくことで、オープンであることを示せる。

・組織・体制

・・国の統合データベースに早くしてほしい。

・・運営としては、文科省、JSTの色がなるべくなく、内閣府とか現場、それから産業界、NCBI、そういうところの全くの第三者が運営に直接かかわれるような形で進めていただきたい。最終的に、省庁から独立したような基金で運営できるものに、早くしていただきたい。

・評価

・・利用率とか論文の採択のされ方とかヒット数とか、そういうもの以外では、なかなか客観的な評価は難しい。

・・利用回数や利用時間、ダウンロード回数、論文等に引用された回数。

〇大学等研究機関

【問題提起】

・本当に統合データベースというのが、日本の風土なり文化なり今の行政体制の下に成立する可能性があるか。

・人をどうやって育てていくか。

・「使いやすい形であればそれでよい」。使いやすく魅力があって、かつ、データに信頼性が置かれるものを作るべき。

・最初はそれほど大きくなくてもいいから、モデルを作って外国に認められる、あるいは日本のいろいろな方に認められる非常によいものを作り、最終的には皆さんがそこに自発的に参加していただけるような形が、今の日本の一番現実的な方法ではないか。

・地球惑星科学の事例では、メタデータ（どんな機材で、いつ、どのように観測されたデータかというデータ）を、きちんと各自が配信して1カ所に集めた。メタデータが検索できれば、どこに何のデータがあるかわかるため、そこへ取りに行って、自分のところでインテグレートすればよいというシステムを作り上げている。

【どういうサービスを提供すべきか？】

・こういうものに対してこういうデータがないかとか、こういう検索をしたときに何か当たるデータがないかというような問い合わせみたいなものの登録を定期的にやっておいていただいて、出てきたらこちらに知らせてくれる機能。

（何か可能性がありそうなデータをすべて取り尽くすというものが、インフォマティクスをやっている人間でも、結構限界に達してきている）

・常に最新のデータにアクセスしたいという要求と、あるいはそういうオーソライズされて、もうあまり変わることのないデータにアクセスしたいという要求が、ある。

例えば、「自分たちが今まで研究していたこの遺伝子のオーソログは何なの？」ということ1つでも、なかなかとることができない。文献を読むと、「このヒトの今見ている遺伝子は、酵母の

これのオーソログなのだけれども」と書いてあっても、実はそういうデータがどこにあるのか、「では、ほかの遺伝子のオーソログは何なのだろう」とかと思ったときに、なかなかそこにたどり着くことができない。

・統合データベースセンターで検索すると、いろいろな情報を返すが、テキストデータだけではなくて、その解析の結果を見せるようなことを作るとよい。

・ソフトウェアのデータベースも必要。「利用方法（これをアプライするとこうなる）」、「ソフトウェア自体はこういうアルゴリズム」という内容が含まれるもの。

・微生物ゲノムのオートアノテーションとか、遺伝子の意味づけをするのはできるが、これまで個々の研究者の知識を注ぎ込んできたデータの集大成と関連付けるための活動を希望。

・大規模性、網羅性が必要になってきており、アノテーションが重要。質の高いアノテーションをまず付けることが必要で、検索というのはその後にそのアノテーションを使っていくため、まずそのアノテーションが大事。

・一般の研究者には、アノテーションばかり付けられていても、ダウンロードできたとしても、研究することというのは難しく、基本的な解析ツールというのは必要。

・利用実績、何か目玉となるようなデータベースがあって、それを世界の人の多くが使ってくれるとか、論文に引用されるとか、そういう目玉となるものを持つしかない。逆に、それを持てないような計画だったら、遅かれ早かれ立ち行かなくなるだろう。

・データを手間暇かけてキュレートするということが、そのデータの価値を高める、すごくよく整理されたデータベースというのが有用であるが、一方で、何でもかんでも整理して統合して使いやすくするという努力にかかる多大なエネルギーを考えると、要はインターネットでアクセスできればいいのだという考え方もあり得る。データをとにかく外から見えるようにできればそれでいいという考えも、ある。

【どういう種類のデータまでを対象とすべきか？】

・ヒトが中心で、マウスとか哺乳類からだんだん、どちらかというと、すそ野を広げていったときに微生物まで行き着く。もちろん、微生物でも常在菌とかがいるので、そこら辺は必要。

・日本語について、医学とか、あるいは薬のデータというのは、日本語のものというのは意外と多いのではないか。

【センターでの研究開発、人材育成のあり方は？】

・人材養成

・・テクニカルスタッフが雇われている大学の研究室があるが、情報のテクニカルスタッフのような方を雇えるような環境というのが、人材とともに整備されてくればいい。

・・すべてやるという人材。

【データ共有のあり方、進め方？】

・誰も自分の得になるということが明らかでない限りはなかなか出してくれないので、例えばパブリケーションが一気に増えるとか、自分の付加価値を高めるようなインセンティブを、示すべき。
・ビッグファンドのついたビッグプロジェクトというものに、常に関わってくる体制がよいのではないか。
・アメリカのデータが出てすぐ公開されるというのは事実ではあるが、その横で、ほとんどのそういう大プロジェクトの研究室等は、どこかの製薬会社が必ず事前にデータを見ている。データが出るときには、もう既に特許化されて、何らかのプロテクションした形で皆さんに公開されている。米国では政府のお金をビジネスにつなげていくというところは、存在する。
・NBRPというプロジェクトは、かなり下等な生物から高等の生物まで、物と実体と、それからその情報を一緒にインテグレートしようということで、日本の国家プロジェクトとして、有効に機能している一つのプロジェクトの例ではないか。こちらの統合データベースは、それも含めてさらに上の段階で統合するということだが、情報がなくなったらバイオロジーは、ある意味で成り立たないのではないかと考えるので、うまい統合の仕方を考えて欲しい。「統合」というのは、決して何かどこかに集めることだけではなくて、うまく繋ぐことだと思う。

【組織、体制のあり方は？】
・予算
・・お金のメリハリのある使い方というか、予算をどういうふうにうまく使っていくか、削れるところは削っていくかということを訴えていく。
・評価
・　・データベースの評価というのは、ビッグプロジェクトに関わりその引用回数によって評価されてよいのではないか。

以上

（2）データ生産者分科会概要

ライフサイエンス分野統合データベースセンター設置準備委員会
データ生産者分科会　概要

1.　日時　　平成21年12月4日（金曜日） 13時00分～15時00分
2.　場所　　東京グリーンパレス内小会議室 「さくら」
3.　出席者
（委　員） 大石委員長、高木委員、豊田委員、中村委員
（ご欠席　有田委員、大久保委員、田畑委員、藤山委員）

伊藤委員、河合委員、古崎委員、後藤委員、菅野委員、夏目委員、若槻委員

（ｵﾌﾞｻﾞｰﾊﾞ） 吉田部長

（事務局） 広瀬理事、門田審議役、大倉部長、菊池次長、黒田課長

4. 議事

（1）本分科会について

（2）データ生産者分科会 論点

（3）意見交換

（4）その他

5. 配布資料

資料1 ライフサイエンス分野統合データベースセンター設置準備委員会

データ生産者分科会委員名簿

資料2 ライフサイエンス分野統合データベースセンター設置準備委員会規則

資料3 JST統合データベースセンター（仮称）設置準備委員会分科会（データ生産者）論点

（ROIS・DBCLS 高木センター長 作成資料）

資料4 ライフサイエンス分野統合データベースセンター設置準備委員会

第1回、第2回、第3回 議論概要（案）（2009.11.24 JST作成）

参考資料①

「ライフサイエンスデータベースの統合・維持・運用のあり方」

（H21.1文部科学省科学技術・学術審議会研究計画・評価分科会ライフサイエンス委員会ライフサイエンス情報基盤整備作業部会）

参考資料②

「統合データベースタスクフォース報告書」

（H21.4総合科学技術会議 基本政策推進専門調査会 ライフサイエンス統合データベースタスクフォース）

参考資料③

ライフサイエンス分野統合データベースセンター設置準備委員会委員名簿

〇高木委員より下記の論点（データ生産者分科会向け）を事前に提示し、分科会委員より意見をいただく形式で委員会は進められた。

- センターの機能、位置付けは？
  - サービス、ツール、標準化、国際対応のあり方は？
  - データ生産者側での DB 構築とセンターとの関係は？
- センターでの研究開発のあり方は？
  - データ生産者との関係は？積極的に共同研究？

- DB 構築の高度化標準化の研究のみをすべきか？
- データ解析も含めたバイオインフォ全般か？

- 人材育成は？
  - センターでどういう人材を育ててほしいか？
- データ共有のあり方、進め方は？
  - どういう点に配慮して進めるべきか？
  - どのような共有ポリシーが望まれるか？
- 組織、体制のあり方は？
  - 永続的で開かれた運営にはどういう点に配慮すべきか？
  - データ生産者の意見を反映する仕組みは？

〇基礎的研究者の立場から

【センターの機能、位置付けは？】

・データ生産者側での DB 構築とセンターとの関係は？

・・メディエーター機能。データベースをやっている人たちが集まるフォーラム、生産者が集まるフォーラム、ユーザーが集まるフォーラムをメディエートして、ソリューションをやっていく。

・・個々のデータベースの構造はそれなりに認めつつ、それを統合する。全体とのインターフェースというのが、単に技術的なものではなくて、入りやすくなっているものがあるといい。

・・統合という言葉が、無駄なものを削るというようなニュアンスに一般的にはとらえられやすいが、ワークのシェアと、お互いの役割分担の中で絵をかいていくような形で、データ生産者から公開までつながっていくような形を描くとすっきりする

・・二度手間にならないように、最初からセンターに預けられる形がよい。その際、データ産出する研究者とコミュニケーションがとれる人材を育て上げてほしい。

・・普通のライフサイエンスの人に、余り「データ生産者」というと怒る人がいるのではないか。

・・恐らく、要らないデータはともかく共有に、一番大事なところは隠すという、インモラルなことになる。データ共有は難しく、多分二つのメンタリティーの人を相手にすることになるので、これはなかなか簡単にはいかない。

・・長期的な視野に基づいたファンディング：選ばれた機関がずっと実施できるということで、緊張感がなくなるということを起こさないための評価機構を持った上で、継続的なファンディングをするべき。

・国際化

・・日本の統合データベースとは言いながら、国際的に利用される開かれたものになっているということがとても重要。

・国際レベルでの標準化の必要性

・・配列情報であってもその特質にあわせたデータベースへの収録が必要。

・・日本発の「CAGE法」と呼んでいる方法を開発して、データを大量に出している。「CAGE法」は

よく使われる「SAGE法」と似てはいるけれども、質的に違うもの。このSAGE法やCAGE法のデータは、シーケンス情報として出てくるが、発現情報の色合いを持っているということで、データベース側としてどう位置づけるかというところが、まだ整理し切れていない。NCBIでは、SAGEデータはGEOというところに格納されているが、DDBJセンターにおいては、シーケンスのデータベースDDBJに格納されている。

・サービス

・・シーケンスベースのデータは三極に同じものがあるが、そのほかの関連機能（例えば（ゲノム）ブラウザ）を整備しているところがある（例えばNCBIとUCSC、EMBL）。さらに、ツールまで広げるとNCBIは充実しているということで、専らNCBIとUCSCを利用する。

【人材育成は？】

・技術者のキャリアディベロップメント、キャリアパスができていない。そういうものを作っていただきたい。

・研究開発をするときに、高度化、標準化の研究のみではなく、データ生産者、かつ、それを使う人たちが三つ巴になって、このセンターでの共同研究ができるという仕組みを作ると、いろいろな意味で使われる確率も高くなる。

【データ共有のあり方、進め方？】

・データ共有や公開の事例

・・ゲノムネットワークプロジェクトの例：データの共有あるいは外部公開に関して、契約においてその義務を設定する。データはプロジェクト内においては、すべて共有する。未発表の段階で共有する。共有後、6カ月後に一般公開をするというのを契約上で設定。

・・セルイノベーションプロジェクトの例：公募要領において、統合データベースにデータをつなげることが明記されている。今年度始まったということもあって、現時点においては、その具体的な手続は未定、その強制力みたいなものも現在はまだ不明瞭な段階。

・・タンパクの構造のデータの例：「タンパク3000」は基本的に6カ月後に外部公開するということにした。「ターゲットタンパク」は、知財のことがあるので、最初から6カ月ということを余り明確にしていない。むしろ知財をとりにいく方向、あるいは戦略としてとるべきだということが多分ある。

・データ公開と特許

・・基本的には、企業には自由に使っていただき、製薬業界等企業が隠しておきたい成果については容認して、売り上げが立ったら納税するという循環がある。

・・データ生産者側が公開したデータに対して、特許を取得していると、データを利用して産業界が商売をしようというときに、実はこれは特許が取得されており、このデータを利用して商売できないなど、産業界側から少し利用しにくい面と、知財はきちんと確保しなければいけない面という、バランスが常にある。その点についての非常に難しい問題なので、常に配慮されなが

ら運営されるということを強く望む。

・・データ生産者の側が、現実に即した範囲内で知財を出すというようなバランス感覚というものが、第三者が知財の専門家、それから知財をライセンスした経験のある人間、あるいは製薬業界の知財担当、経験者のような方が、公正なところで議論をしながら知財を出願していくような、コンサルティング機能があるべきではないか

・・プロテオムワイドに取得可能なデータの例：あるタンパク質のこの配列の部分が非常にイオン化しやすくて、定量計測するのが非常に適しているという情報をデポジットしたい。アカデミアが利用するのは問題ないが、データ生産者側からすると、ペプチドそのものが主役になり、これは特許をとってビジネスにしたいとすると、このデータが公開された後に、会社、企業がビジネスしたいということができないという状況になる。

【組織、体制のあり方は？】

・意見交換をする機会を継続的に持つ。

・データ共有、あるいはデータの統合ということが非常に大事であるということは、皆さんよく認識されていると思うが、データ統合ができるとライフサイエンスがガラッと変わる、あるいは医療費が削減される、のような説明ではなく、非常にインテリジェントな形で訴えていくのが大切。

・データ生産者の意見というのは、機会をなるたけ設けていただいて、声を拾い上げるということが大事。

・コアが要る。バーチャルでできるというのは非常に難しい。小さくてもいいから、ちゃんとやれる体制を作って、できたら運営費交付金みたいな形で持続できる部分を持ちたい。

<u>○疾患研究者の立場から</u>

- 自ら生産したデータを自ら利用することで研究を行う実験科学（wet）系疾患研究者の立場からの提示
- 要点は：
  - ✓ データ生産者をPIとする「一次利用*」。
  - ✓ データ生産者による一次利用あるいは一定期間の優先利用後の、統合データベースへのデータの「寄託」。
  - ✓ 統合データベースユーザーをPIとする「二次利用*」 。
  - ✓ 統合データベースセンターならではの研究開発と、それに伴う人材育成・供給。
  - ✓ 疾患研究用の統合データベースは連邦型（federation-type）で構築・運営。そのHQが統合データベースセンター。

  ＊　「一次利用」・「二次利用」の区別があまり意味を持たない、初めから皆で「一次利用」を目指す大型プロジェクト等の研究もあるが、bottom-up型の個人の発想の研究も重要で、そこでは「一次利用」の保証が必須。

【センターの機能、位置付けは？】

・サービス、ツール、標準化、国際対応のあり方は？

・データ生産者側でのDB構築とセンターとの関係は？

※データ生産者側でのDB構築はデータの一次利用、センターでのDB構築はデータの二次利用。

①データ生産者は従来通り、疾患専門研究者としての発想・提案に基づいて研究費を獲得し、その研究のための個別データベースを構築し、個別疾患研究と報告を行う。（データの一次利用）。このようなデータベースは「臨床に学び、臨床に還す」タイプの研究のツールとして必須であり、個人情報保護の観点から公開は不可能な詳細臨床情報・分子解析情報を含む。典型的には施設内データベース（単にその研究者が持つ一枚の症例のリストであることもある）。センターのDBとの重複はないと整理する。

②その後、公開可能な範囲で、可能な限り、元データを、連邦型・virtual統合データベースに登録する。疾患研究においては、全疾患で一つではなく、複数のDBが必要。疾患研究はその進歩に伴い、phenotype（疾患概念・分類を含む）の定義・解析項目等も変化するので、疾患研究専門家の意見が即座にかつ容易に反映できる、その疾患研究にある程度特化した複数のデータベースが必要（c.f. DDBJなどの分子視点のencyclopedia型DBは、中央集中型・central typeの統合DBであるべき）。

③連邦型統合データベースの「データ利用者」（dry系ユーザーの他、meta/pooled analysisなどを含めwet系ユーザーも多い）は、個々のデータの登録者（一次利用者）がやらない・できない、あるいは想定できない方法でデータを活用し、新しい知の発見の研究を行うための研究を提案し、研究費を獲得して、研究を行う（データの二次利用）。その際、データ登録者の協力・アドバイスを積極的に求め、彼等のcreditをあらかじめ決められた条件で適切に担保するが、知的活動としてのPIはデータ利用者。

【センターでの研究開発のあり方？】

・データ生産者との関係は？積極的に共同研究？

・DB構築の高度化標準化の研究のみをすべきか？

・データ解析も含めたバイオインフォ全般か？

④前項②の、二次利用のためのデータ登録とそのデータの公開までは「共同研究」とはせず、「データ生産者」はセンターにデータを「寄託する」（基本的に、双方とも対価を求めない）。

⑤前項③については、open access部分のデータについては共同研究を義務にはしない。Controlled access部分のデータについては、一定の基準に合致するデータ利用要請について、センターは利用承認の過程でデータ生産者の意見を聞くようにし（「寄託」なので）、必要に応じて積極的に共同研究を「仲介」する。

⑥センター自体もデータ利用者になりうる。その他、センターのmissionに基づき、「統合データベース」関連の研究を行う。統合データベースの構築・運営・活用技術、統合DBのデー

タを用いたbioinformatics研究等々。その過程で疾患研究者との共同研究も行う。センター職員がPIの研究も、疾患研究者（データ生産者を含む）がPIの研究もある。

なお、センター職員が統合DBの発展に資するデータ生産を目指して、データ生産者用の競争的資金にPIとして応募することは、制限した方がdemarcation及びデータ生産者との円滑な関係構築上、望ましいかも知れない。

【人材育成は？】

・センターでどういう人材を育ててほしいか？

⑦前々項①（データの一次利用）・③（二次利用）の研究の推進に資するdry系研究者。すなわち、疾患研究者は上記①の自らがPIとなる一次利用において、参画・協力してくれるbiostatistician/ bioinformaticianを探し求めている。また、自らが生産したデータについて、多くの研究者が活用し、③の研究が進むことも願っている（co-authorの論文が増える可能性もincentive）。

【データ共有のあり方、進め方？】

・どういう点に配慮して進めるべきか？

・どのような共有ポリシーが望まれるか？

⑧ゲノム関連のDBの多くが、個人を同定できる可能性と、同定された場合、その個人が被る不利益の程度、知財等への近さ等を考えて、open accessとcontrolled accessの2 tier（以上）にしている。そのモデルを踏襲し、この点においても将来の世界規模での連邦型統合データベース構築も視野に入れて、国際化・標準化に留意すべきであろう。

<u>〇臨床データの特性や経験の紹介</u>

・臨床のデータでは、どのデータから公開しようかというところで、もめる。データをどのような技術を使ってどう公開すれば、本当に自分が使いたい形で皆さんが共有できるのかという、そのデータベース、公開する技術的な面は、うまくやりとりできないというのは現実問題としてかなり起る。

・データベースを作る側と実際に公開する間の側のコミュニケーションを仲介するところが抜けてしまっており、ここをサポートする役割があるとよい。

・診療とリンクする臨床情報の扱いには倫理的な配慮が必須となる。臨床データベースに個別の患者さんの名前とかIDがリンクしている情報は、基本的に病院の外部に出すことはできない。

・個人を特定する情報に遡れないようにしつつ、患者さんの持っている臨床情報である血圧とか脈拍とか数値化できるデータ、血液検査データを科学研究に利用できるデータベースにできるか？という倫理の議論が非常に重要。

・倫理の問題さえクリアすれば、個人を特定できない状態において、ゲノムから臨床までのすべてのデータをデータベースとして組み込んでいくことは技術的には可能と思う。米国の一部の施設ではゲノム情報から臨床情報までを一括してデータベース化する試みがなされている。

・倫理面を仮にクリアできたとして、ではどのような形で公開するかという医療情報の標準化という面では、まだまだかなり厳しい状況が国内にも、海外にもある。病気の情報を一つ公開するのにどういう形でするかというのが、まだ延々と議論がされている状況。
・技術面とそういう人材の間の仲介面、その両方をうまくやっていく必要がある。
・新薬の臨床開発は、薬剤介入を受けた症例と受けない症例の臨床情報をeCRFというネットワークを用いたツールを用いて、有効性と安全性を検証するという方法にて行われている。
・個人の遺伝情報から様々な環境因子、リスク因子への曝露の過程を経て疾病の発症に至る経路をどこかの国が明確に解明してしまえば、医療介入の理想型が明確にされる可能性がある。
・疾病発症に至る地図というか、ストリーミングを日本以外の国から戦略的に押さえられてしまうと、日本の医師もプロフェッショナルフリーダム（自己裁量権）を失う可能性がある。
・この医療介入の理想型が真に科学的かつ日本人にも理想的な理想型なら患者さんにもメリットが大きいが、他国がデータベース作りに大きく先行すると、われわれは提案された理想的な医療介入が日本人に対する理想的な医療介入であるか否かを検証することもできなくなり、極めて危険であると認識。
・ヴァンダービルト大学とかバーバード大学のバイオバンクの目的は、ゲノムの情報から臨床情報までの道筋のモデルを作ることが目的。
・最大のハードルは、社会の皆さんに医療の質を向上させるために、採血を受けた時には検査情報とゲノム情報が公共のデータベースと使用されるということに、合意してもらえるかどうかなのである。
・日本人5,000人と、主に西欧人を中心とする約6万例の症例のコフォートを作って、1年、2年、3年と心血管イベントの発症を追跡する研究の結果、欧米人は心筋梗塞発症率が高く、日本人は脳梗塞発症リスクについては高いが、心筋梗塞率は低いことがわかった。病気の発症率も、それから死亡率も薬剤の使い方もきき方も、東洋人と西洋人には大きな違いがある。日本人の症例におけるデータベースが必須である。
・疾患研究においては、データベース構築自体が研究のツールであって、疾患研究の本当に根幹にかかわるところで、疾患概念も変わるし、疾患の名前も分類も定義も変わってくるというところがあるので、やはり一次利用としてのデータベース構築の研究というのは、そのまま温存し、統合データベースとは別なグラントとして進行するようにしていただきたい。

<u>〇臨床データベースの特異性</u>

遺伝子や蛋白のデータベース（DB)との違いは？

●情報が固定されておらず毎年のように変化する。
●個人情報が含まれる。珍しい疾患であれば病院名と入院日で個人が特定される？
●患者一名の臨床情報自体が統合DBの要素を含んでいる（複雑性）。
●臨床DB作成において、何をどのように記述するかスタート時に不明な点が多い。
（作成時に目的が特定されていないので何をどのように記述するか決まりがない）

●一般公開に馴染まない

以上