

学会戦略ワークショップ
「21 世紀の生物研究における知識の共有と理解」
報告書

平成 17 年 7 月
独立行政法人科学技術振興機構
研究開発戦略センター
江口グループ

Executive Summary

研究開発戦略センター江口グループは、ライフサイエンス分野を中心に据えて科学技術分野を俯瞰し、社会ビジョンや社会ニーズの実現に向けて、今後重要となる研究開発分野、領域、課題を明らかにし、当該分野の専門家によるワークショップや海外調査などの議論や調査を経て、研究開発戦略の立案を行っている。

本報告書は当センターが平成15年3月に開催した「戦略ワークショップ（ポストゲノム系）」において今後のライフサイエンスの重要分野の一つとして抽出された「統合的な生命の理解に向けてこれまでや今後の研究で蓄積され、得られる実験データや知識を総合的に活用する」ために重要な研究分野に関して、その現状と今後の展開について第27回日本分子生物学会（平成16年12月8日～11日）において開催した学会戦略ワークショップ「21世紀の生物研究における知識の共有と理解の展開」の結果を取りまとめたものである。

ヒト・ゲノム計画が今世紀初頭に終了した生命科学分野では爆発的なデータの増加と蓄積の時代に入った。一方で、研究分野（知識）の細分化が進み、生命科学分野の専門家においても分野の俯瞰ができにくくなりつつある現在、実験研究の対象自体に注目していたのでは、知識を体系化し、知識の再生産を図ることは不可能である。

学会戦略ワークショップでは分野の俯瞰や巨大知識を体系的に整理し、知識化する試みなど最新の研究成果を紹介し、これからの生物研究を支える知識環境にどのような革新が必要かを統合データベース、オントロジー、言語処理、生命誌などの視点から、生命知識自体を研究の対象としている専門家の発表を中心に検討した。その結果、これからの科学の発展のためには生命科学分野において爆発的に蓄積されつつある実験データや論文知識を当該分野の研究者のみならず今後重要となる融合研究での生命科学分野以外の研究者や一般の人々が知識として有効に活用できる環境を構築する必要があることが指摘された。そのためには、まず、計算機利用を前提に、計算機側で知識の処理や知識に基づいた処理が可能なシステム（知的なシステム）を構築する必要があることが指摘された。その上で、研究者や研究分野の知識、論理体系を辞書やオントロジーなどの形で表現し、計算機の持つ巨大な計算能力と膨大な記憶能力を組み合わせる新しい科学技術領域の研究開発を推進する必要があることが指摘された。

研究開発戦略センターでは、本学会戦略ワークショップの議論を受けて、生命科学研究の成果を研究開発のみならず、科学技術の理解増進にも、有効に活用、展開を図るための研究開発戦略の立案を進めることとしている。

目 次

[1] はじめに	1
[2] 学会戦略ワークショップの概要	2
[3] 21世紀の生物研究における知識の共有と理解	3
3.1 背景	3
3.2 現状：爆発的に増加し、蓄積されるライフサイエンスの情報や知識とその利用	4
3.2.1 DNAやたんぱく質などの配列や分子構造、遺伝子やたんぱく質の発現状態や変異及びそれら分子の結合定数など事実（データ）に基づく情報	4
3.2.2 遺伝子の制御関係などのアノテーション（注釈）、ネットワークやパスウェイ及び生物種別の統合データのような相互の関係性に関する情報や知識	6
3.2.3 文献やバイオオントロジー（生命現象の概念をまとめた辞書など）など意味に関する知識	7
[4] 今後重要となる研究分野	12
4.1 知識の表現法（形式化など）	12
4.2 様々なデータや知識を統合したり、操作したりするための技術開発	14
4.3 今後の課題	15
[5] まとめ	19

学会戦略ワークショップ

「21世紀の生物研究における知識の共有と理解の展開」

【1】はじめに

研究開発戦略センター江口グループは、ライフサイエンス分野を中心に据えて科学技術分野を俯瞰し、社会ビジョンや社会ニーズの実現に向けて、今後重要となる研究開発分野、領域、課題を明らかにし、当該分野の専門家によるワークショップや海外調査などの議論や調査を経て、研究開発戦略の立案を行っている。

本報告書は当センターが平成15年3月に開催した「戦略ワークショップ（ポストゲノム系）」において今後のライフサイエンスの重要分野の一つとして抽出された「統合的な生命の理解に向けてこれまでや今後の研究で蓄積され、得られる実験データや知識を総合的に活用する」ために重要な研究分野に関して、その現状と今後の展開について第27回日本分子生物学会（平成16年12月8日～11日）において開催した学会戦略ワークショップ「21世紀の生物研究における知識の共有と理解の展開」の結果を取りまとめたものである。

江口グループでは、本学会戦略ワークショップの議論によって指摘された問題点や研究課題に関する調査を進め、研究開発戦略の立案に活用することとしている。

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

【2】学会戦略ワークショップの概要

- ① 開催日時：平成16年12月11日（土） 13:15～15:30
- ② 会場：第27回日本分子生物学会 神戸国際会議場
- ③ プログラム：学会と社会との接点等に関するワークショップ
「21世紀の生物研究における知識の共有と理解の展開」
- ④ 参加者数：約80名
- ⑤ 要旨：

大量で多様な知識を整理し、体系化することは知識の共有と更なる理解のために重要な作業である。ことに、生命科学においては、知識を具体的な生命現象と関連づけながら整理し、生命の理解につながるように体系化（例えば、生命誌のような視点で）することが理解の発展の要件でもある。重要であるにもかかわらず従来は軽視されがちであった知識、特に生命科学における言語で記述される知識を整理する研究は世界的にも注目され始め、統合データベース、オントロジー、言語処理などの工学的な思考や技術の盛んな応用が展開している。

本ワークショップではこれらの研究に関わる研究者から、分野の俯瞰や巨大知識を体系的に整理し、知識化する試みなど最新の研究成果を紹介し、これからの生物研究を支える知識環境にどのような革新が必要かについて幅広い議論を喚起したい。更に、知識の時代といわれる21世紀の生命科学の発展に向けて具体的な研究戦略や戦略的ファンディングの提言につなげていく。

演者と演題：

高木 利久 東京大学大学院 新領域創成科学研究科 教授

「知識の共有化とさらなる理解のための生物論理の形式化」

村田 英克 JT生命誌研究館 サイエンス・コミュニケーション&プロダクションセクター チーフ代行

「語る科学に向けて 一体知識は増えているのか？」

菅原 秀明 情報・システム研究機構 国立遺伝学研究所 教授

「配列からみる生命科学 配列以外のフィーチャーは役に立つのか？」

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

「機械にデータを読ませたら」

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

川本 祥子 情報・システム研究機構 国立情報学研究所 研究員

「バイオポータル」

辻井 潤一 東京大学大学院 情報学環 情報理工学系研究科 教授

「バイオNLP 機械に生物学言語は扱えるのか」

野田 正彦 科学技術振興機構 研究開発戦略センター フェロー

「科学知識とファンディング」

[3] 21世紀の生物研究における知識の共有と理解

3.1 背景

大量で多様な知識を整理し、体系化することは知識の伝搬の促進に重要な作業である。ことに、ライフサイエンスにおいては、知識を具体的な生命現象と関連づけながら整理し、生命の理解につながるように体系化することが目的であると同時に発展の為の要件でもある。重要であるにもかかわらず従来は軽視されがちであった知識、特にライフサイエンスにおける言語で記述される知識を整理する研究が、統合データベース、オントロジー、言語処理、生命誌、エビデンスに基づく戦略的ファンディングなど伝搬される知識を様々な形で利用する分野の拡大により、ますます重要になってきている。

本研究分野の重要性は研究開発戦略センターの前身である研究開発（R&D）戦略室が実施した戦略ワークショップ（ポストゲノム）において指摘された（図1）。このような背景をもとに、研究開発戦略センターでは本分野の現状と今後重要となる研究分野がどのように展開すべきか、などについて検討を深めるために、生命科学関係の研究者が最も参加する日本分子生物学会において当該分野の研究者の協力を得てワークショップを開催した。本報告書はワークショップでの発表や議論をもとに取りまとめたものである。

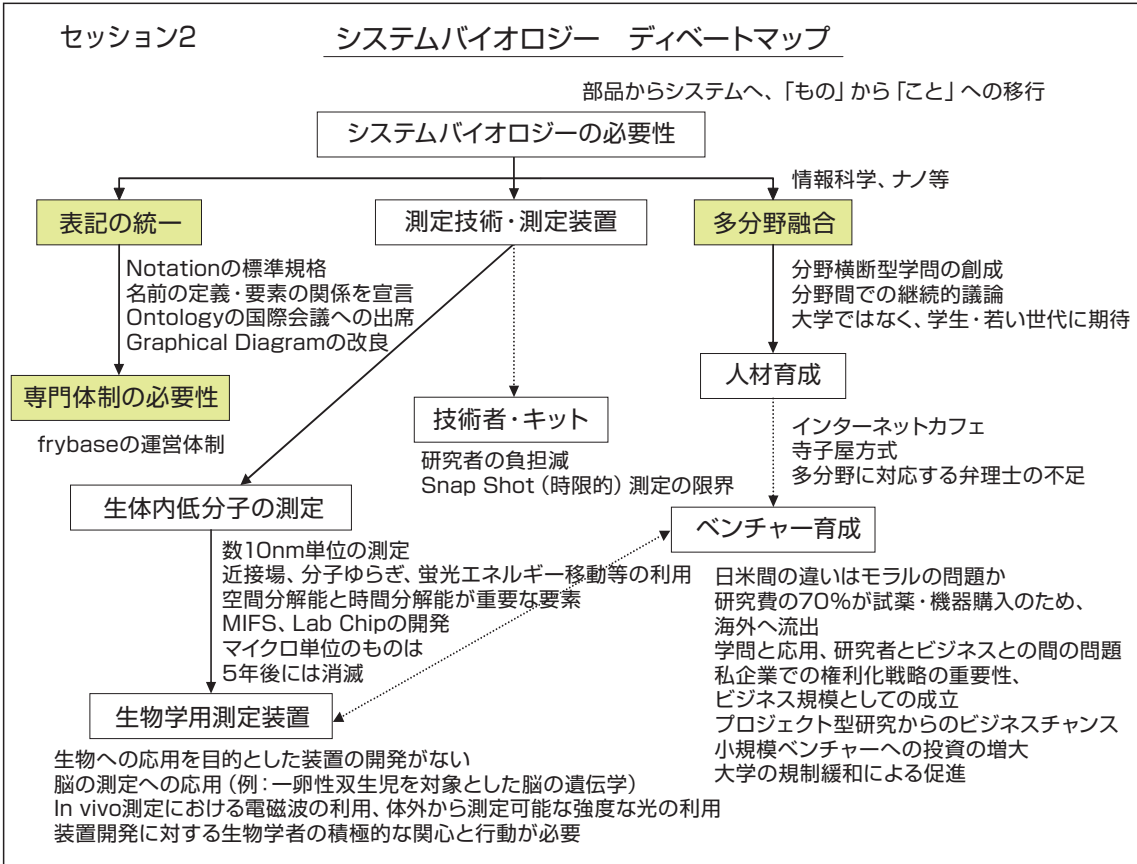


図1 戦略ワークショップ「ポストゲノム（平成15年3月開催）」の議論より

本図は戦略ワークショップでの議論の展開を図示したものであるが、システムバイオロジーの必要性の議論の過程において、生物現象の測定や観測技術の重要性とともに、その結果得られる情報や知識に関する表記やオントロジーの統一の重要性が議論されている。また、多分野融合における重要な課題であることも議論された。

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

3.2 現状：爆発的に増加し、蓄積されるライフサイエンスの情報や知識とその利用

生命科学研究における情報や知識は以下のように大まかに分類することができる。以下項目毎に現状を説明する。

- DNAやたんぱく質などの配列や分子構造、遺伝子やたんぱく質の発現状態や変異及びそれら分子の結合定数など事実（データ）に基づく情報
- 遺伝子の制御関係などのアノテーション（注釈）、ネットワークやパスウェイ及び生物種別の統合データのような相互の関係性に関わる情報や知識
- 文献やバイオオントロジー（生命現象の概念をまとめた辞書など）など意味に関する知識

3.2.1 DNAやたんぱく質などの配列や分子構造、遺伝子やたんぱく質の発現状態や変異及びそれら分子の結合定数など事実（データ）に基づく情報

ゲノム (DNA) やたんぱく質など分子に関するデータは近年、爆発的に増加している (図2)。また、それらのデータ資源は様々なデータベース (表1) に格納されて、公開されており、インターネットから容易に検索・閲覧・解析できるようになってきている。しかしながら、そのようなデータベースやツールは分散しているために、そこから必要な情報や分散している情報を組合せて必要な知識を的確に抽出するためには、利用者に一定の知識や技術が要求されるのが現状である。従って、分子生物学などの知識を持つ利用者以外の研究者についてはこれらの情報や知識はまだ十分に活かされていないと考えられる。また、一般の消費者においても生物関係のニュースや話題をもっとよく知りたいというニーズが高いと思われるアンケート調査結果も得られている (表2)。

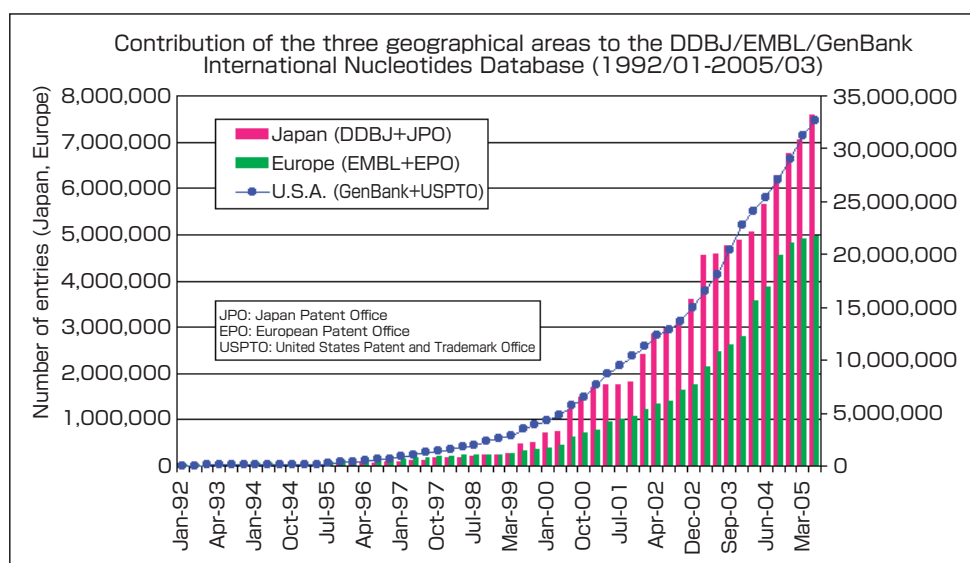


図2 急速に増加するゲノムデータ

出典：DDBJホームページ <http://www.ddbj.nig.ac.jp/images/ddbjnew/percentage-e.gif>

- ◆ Nucleotide Sequence Databases (48)
 - ◆ RNA sequence databases (30)
 - ◆ Protein sequence databases (94)
 - ◆ Structure Databases (58)
 - ◆ Genomics Databases (non-human) (128)
 - ◆ Metabolic Enzymes and Pathways; Signaling Pathways (24)
 - ◆ Human and other Vertebrate Genomes (57)
 - ◆ Human Genes and Diseases (58)
 - ◆ Microarray Data and other Gene Expression Databases (31)
 - ◆ Proteomics Resources (5)
 - ◆ Other Molecular Biology Databases (14)
- (Nucleic Acids Research 2004 Vol32 データベース特集号より)

表1 データ資源の広がり

出典：学会戦略ワークショップ 国立遺伝学研究所生命情報・DDBJ研究センター教授 菅原秀明先生
発表スライドより

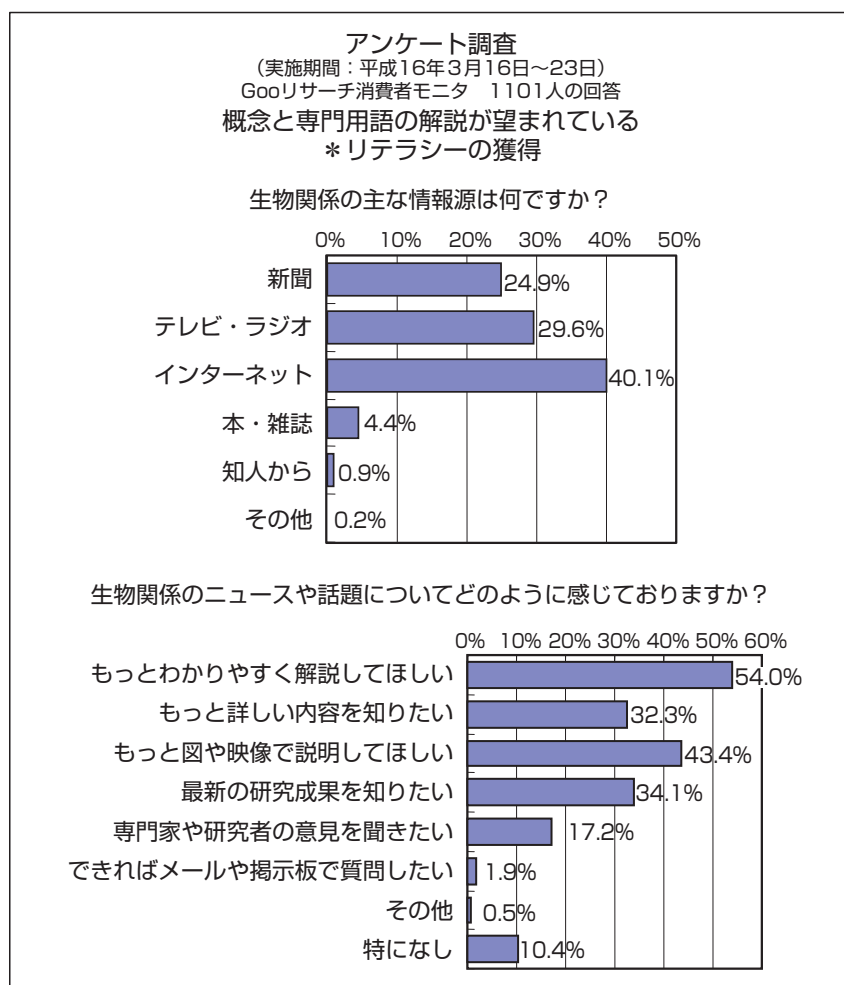


表2 一般の消費者意識

出典：学会戦略ワークショップ 国立情報学研究所研究員 川本祥子先生 発表スライドより

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

従って、生命科学研究の情報から知識へ、そして、それを一般的な生命の理解へとつなげることによって、専門家はその知識を病気の治療や作物の育種につなげ、一般の人々は健康や食生活、環境や教育などで知識を利活用することが可能となる（図3）。このためには、生命科学の専門家から一般の人々まで簡便に生命科学の情報や知識が取り扱える環境を構築する必要がある。

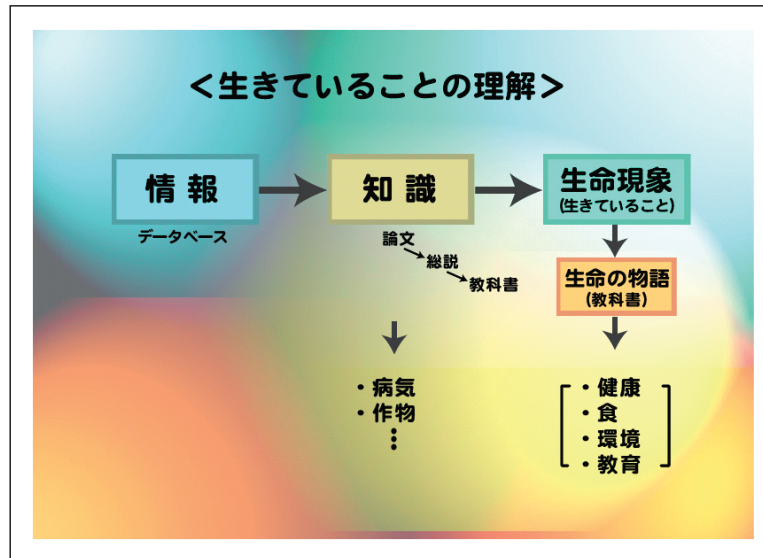


図3 生きていることの理解

出典：学会戦略ワークショップ JT生命誌館サイエンス・コミュニケーション&プロダクションチーム代行 村田英克先生 発表スライドより

3.2.2 遺伝子の制御関係などのアノテーション（注釈）、ネットワークやパスウェイ及び生物種別の統合データのような相互の関係性に関わる情報や知識

生物研究に限ってもデータの爆発的増大と研究分野（知識）の細分化により、統一的・俯瞰的な生命理解が難しくなりつつある。しかしながら、必要な知識を得ようとする、結局は大量の、しかも専門分野以外の論文を読む以外の適当な方法は見あたらない。さらに、その内容を研究者が再構築して新たな知識にすることが不可能になりつつあるのが現状である。

例えば、パスウェイデータベースは基本的には遺伝子やたんぱく質などの分子間相互作用ネットワークの図的表現で分子間相互作用や制御関係などが対象となっているが、何処まで詳しく記述するか、何処まで操作可能にするかなどは各データベース間でまちまちである（図4）。また、論文に書かれた知識の一部が取り扱われていることがほとんどで、分子間相互作用のつながりのみであったり、実験情報等の内容は分類タグ付け、もしくはコメント扱いであったり、解釈された結果のみが取り扱われていたりというのが現状である。

今後はこれまでの情報中心の取扱いから一歩進んで、実験の組み立て、実験データからの結論を導く論理や生命理解の枠組みなどの知識を簡便に取り扱える環境を構築する必要がある。

ある。

また、生命科学研究における知識の共有と理解の展開を進展させるためにデータベースの高度化や計算機を利用した利用者の情報・知識利用のための支援技術やツールの開発が必要となってくる。

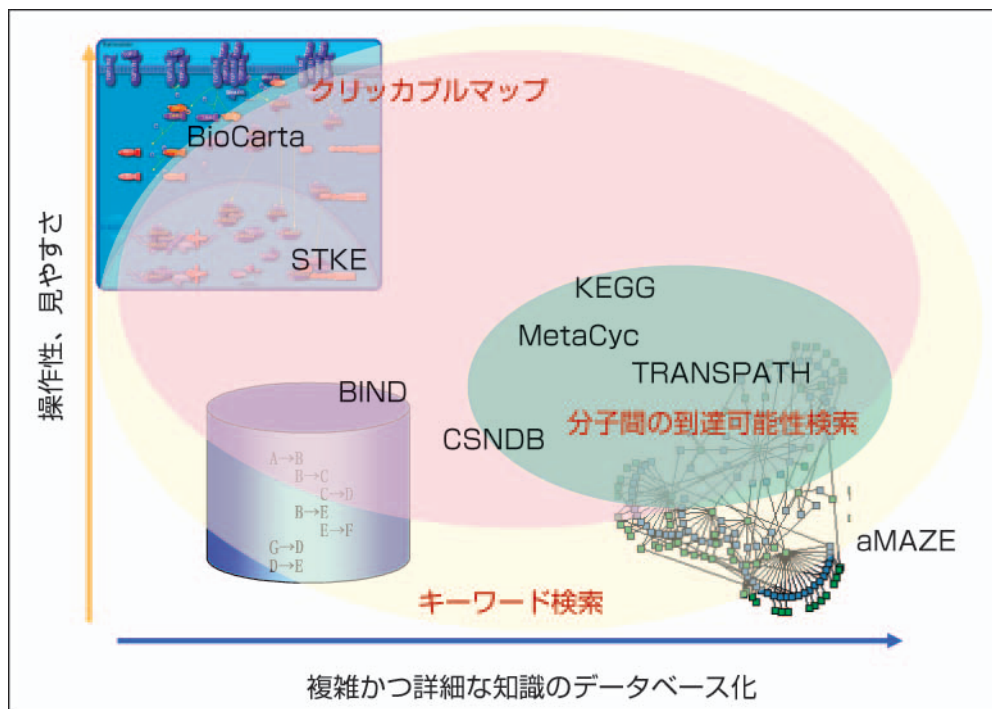


図4 各種パスイデータベース

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

3.2.3 文献やバイオオントロジー（生命現象の概念をまとめた辞書など）など意味に関する知識

生命科学研究における研究成果や得られた知識はこれまで主として論文として記載され、学術雑誌に掲載されてきた。このような論文の知識を利活用するために我が国ではJOISなどの日本語の要約が付されたデータベースが開発され、公開されている。一方、ゲノムなどのデータは無償で公開されているが、国の資金で行われた研究から得られる論文なども同様に無償で公開しようとする動きが米国NIH（国立保健衛生研究所）で始まっている。このような活動はインターネットを通じて生命科学研究の成果に無料でアクセスできる機会（オープン・アクセス：図8）を拡大するものである。オープンアクセスは生物・医学研究に関わる専門家のみならず、生命科学分野以外の研究者が分野融合研究を進める場合や一般の人々への科学技術知識の普及という観点などからも重要である。

このため、大量に公開されつつある研究論文から研究者が必要とする論文だけでなく、一般の人々まで必要な知識を簡便に抽出する技術が可能となれば生命科学研究の情報や知識がより一層利活用されることは疑いがない。そのためには、大量の情報や知識から

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

必要な情報や知識を簡便に抽出する技術が必要となる。

- ・背景1（科学の発展のあり方）
研究者は論文をインターネット上にセルフアーカイブすることにより、自分の考えの影響力を最大化し、考えを「世界中の同僚や専門分野の科学者や学者」に効果的に届けるべき。
(Stevan Harnad 1994)
- ・背景2（最初の例）
インターネット上の最初のプレプリントサーバ (arXiv... Paul Ginsparg 1991)
- ・背景3（必要性と費用の増大の板挟み）
図書館員の嘆き「商業出版社の雑誌の購読費が高額であると言いつけてきた」ことの指摘
(Charles Oppenheim ラフバラ大)
- ・背景4（公的助成側の権利の主張）
公的助成を受けた研究成果を（無償で）提供した研究者やその所属機関に売ることによって出版社が利益を得ていることが容認できるか？→エルゼビア社への批判

図8 オープンアクセスの背景

Harold Varmus (ノーベル賞、NIH前所長) の活動とその後	
1999 E-Biomed 生物医学研究論文サーバの立ち上げ	
2000 PubMed Central 医学及び生命科学分野の無料の電子公共図書館 …但し、保管される雑誌はまだ僅か	
2000 PLoS: Public Library of Science NPO出版社 (Godon財団等からの寄付金で創設)。出版後6ヶ月したらPubMed等に収録。 設立と同時に、オープンアクセス化の推進とそれを拒む雑誌に論文を投稿しないことを誓う公開書簡に署名するよう科学者仲間を説得する檄文を公開。著名な研究者のほかWellcome Foundation (英国)、Max Planck (独国) など欧米の機関も賛同。	
2005 NIH職員は論文投稿後1年以内に編集されていない論文原稿を全てPubMedに搭載する。5月より、NIHの資金で研究した研究者にも適用。	

図9 ライフサイエンス分野におけるオープンアクセスの動き

大量の知識から必要な知識を抽出するためには、統制された知識の体系が必要である。オントロジーとはもとは存在論の意味である (図10-1)。工学領域でのオントロジーは用語や概念の体系化のことである。生命科学では遺伝子やたんぱく質の名前、性質、機能などを治めた辞書やそれらを表現するための統制された語彙や記述言語、種々の生命階層における概念や機能の関係や分類体系がこれに相当する (図10-2)。また、大量の知識を処理するために計算機の使用が前提となるため、ここでは計算機が扱えるテキストや図表での辞書や語彙、記述言語、あるいは、概念や機能の関係性の分類体系をさす。これまで、生命科学研究では様々な側面から研究された生命現象が異なる用語で記述されることが多く、統一的な理解を妨げる要因ともなってきた。このような背景から生命科学のオントロジーを体系化する研究が始まってきている (図10-3)。生命科学の知識を簡便に取扱う仕組みを構築するには、狭義のオントロジーに限定せず、生命科学分野の知的内容を用語の定義や概念間関係などを通して明示的に形あるものとして表現し、体系化すると共に、明示された知識を計算機が利用して、大量のデータや論文等から必要な情報や知識を抽出する技術確立が必要がある。

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

オントロジーとは

1. a branch of metaphysics concerned with nature and relations of being
2. a particular theory about the nature of being or the kinds of existents

哲学：Systematic account for Existence.
知識工学：An explicit specification of a conceptualization

遺伝子やタンパク質の名前・性質・機能を収めた辞書
上記を表現するための統制された語彙や記述言語
種々の生命階層における概念や機能の関係や分類

図10-1 オントロジーとは

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

生物学・医学分野におけるオントロジー

- Gene Ontology
ー遺伝子の生体機能の定義・用語を整理分類
- Disease Ontology
ーヒトの病気関連のオントロジー
- UMLS
ー生命科学を中心とした文献を分類するためのキーワードを階層化
ーMeSH：UMLSのサブセットオントロジー
- TAMBIS Ontology
ー分子生物学、バイオインフォマティクス分野の用語体系
- MGED Ontology
ーマイクロアレイデータのアノテーションを目的に開発
- BioCyc
ー代謝パスウェイを中心に構築した語彙体系
- ...

図10-2 生物学・医学分野におけるオントロジー

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

ゲノムオントロジー研究の背景

これまで生物学においては、生物種や研究対象によって言葉遣いが異なっており、したがって同じ現象が異なる用語で呼ばれることがしばしば起きてきた。そこで共通の言葉遣いを提唱し、既存のさまざまな言葉遣いの対応関係を記載した、統制された知識体系の構築が必要である。

“生命科学の用語は意味が曖昧であり、データベース統合化や比較ゲノムの障害になる” (Schulz-Kremer '97)

- ・例：遺伝子 (gene) の定義
 - －GDB：転写翻訳されてタンパク質となるDNA上の領域
 - －GenBank：遺伝的な特徴（表現型）の制御に関連するDNA上の領域

図10-3 ゲノムオントロジー研究の背景

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

はじめに

1

学会戦略ワークショップの概要

2

理解 21世紀の生物研究における知識の共有と

3

分野 今後重要となる研究

4

まとめ

5

[4] 今後重要となる研究分野

4.1 知識の表現法（形式化など）

今後重要となる研究分野は生命科学研究によって得られる情報や知識を計算機によって操作可能な知識の形式で蓄積し、計算機と人との間で知識を共有できる枠組みを提供する技術やツールの研究や開発を包含する研究分野である。具体的には統合データベース、オントロジー、言語処理、生命誌などの分野で分野の俯瞰や巨大知識を体系的に整理し、知識化するような研究である。これまでバイオインフォマティクスと呼ばれる研究分野は、配列や立体構造など比較的単純なデータや実験データなどの解析データを中心に取扱っていた。今後は分子間相互作用、機能など複雑なデータや結果の解釈、整理された知識などへ対象を移す必要がある（図11）。

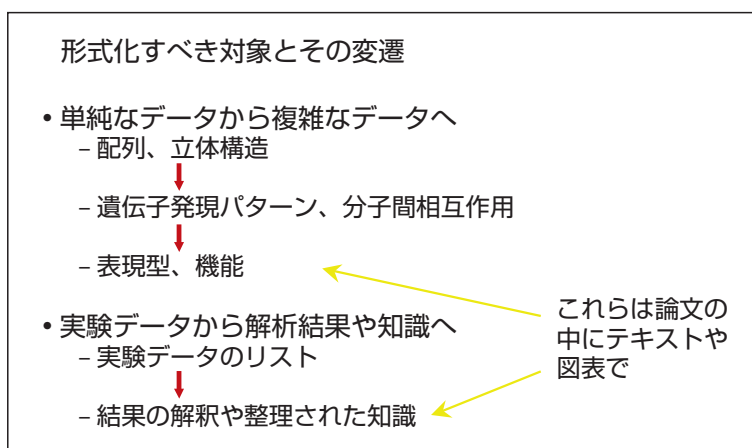


図11 生命科学研究で得られる情報や知識について形式化すべき対象とその変遷

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

生命科学研究から得られる知識の形式化に必要な要素には以下のものがある。

- 論文や教科書（テキストやポンチ絵）から様々な知識を効率よく取り出すこと。
論文（テキスト）から分子間相互作用、遺伝子発現制御、細胞内局在、パスウェイ、たんぱく質の機能、機能や概念の階層、実験条件などを自動的に抽出する。ポンチ絵から計算機による上記知識の自動抽出はまだ難しい。
- 複雑なデータや知識を計算機上に、人間も理解可能な形で、うまく表現すること。
複雑なデータや知識の表現方法としては分子レベルでの実体表現、実体間の関係表現・挙動の表現、機能、表現型の表現などがある（図12）。

複雑なデータや知識をどう表現するか

分子レベルでの実体 (entity) の表現

ゲノム配列、遺伝子 → ATGCの並び
 タンパク質 → アミノ酸配列
 原子の座標配置

実体間の関係表現、挙動の表現

分子間相互作用 → 2項関係
 遺伝子発現情報 → 発現強度 (数値リスト)

機能、表現型の表現

パスウェイ、ネットワーク → パスウェイデータベース
 概念、機能階層 → オントロジー (ex. 遺伝子辞書)
 (動的な挙動) シミュレータ)

図6 複雑なデータや知識の表現

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

はじめに

1

学会戦略ワークショップの概要

2

21世紀の生物研究における知識の共有と理解

3

今後重要となる研究分野

4

まとめ

5

4.2 様々なデータや知識を統合したり、操作したりするための技術開発

近年の大量に蓄積され、あるいは、蓄積されつつある生命科学研究の様々なデータや知識を統合したり、操作したりすることを可能にする技術は計算機を活用することによってのみ可能と考えられる。そのためには、計算機の持つ巨大な計算能力と膨大な記憶能力を組み合わせて利用する新しい科学技術領域の研究開発を推進する必要がある。

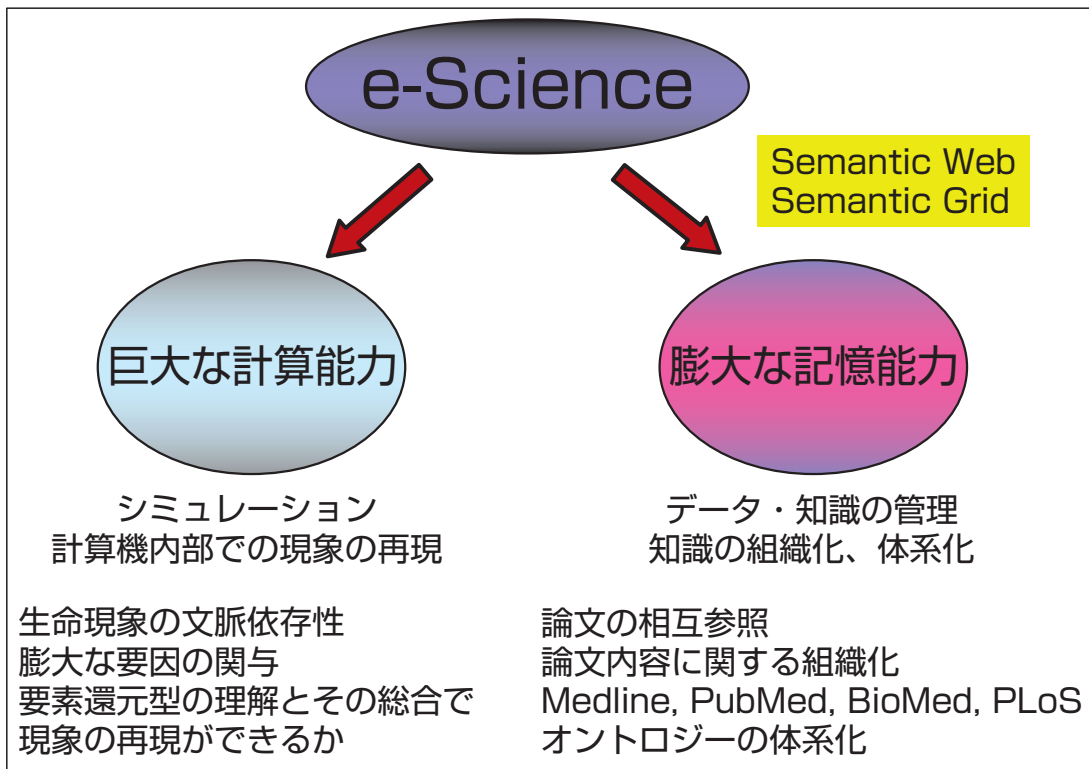


図7 新しい科学技術分野 (e-Science) の創成

出典：学会戦略ワークショップ 東京大学大学院教授 辻井潤一先生 発表スライドより

4.3 今後の課題

今回のワークショップにおいて、これまでのバイオインフォマティクス研究は主として生命科学の情報を取り扱ってきたが、今後は情報から知識へと主体が移ることが指摘された。従って、これからの課題は研究者の論理や生命科学の論理を形式化することであり(図8)、そのために、オントロジーの整備(図9)と、それらを計算機が取扱えるよう辞書化する必要性(図10)が指摘された。また、インターネットを通じて情報や知識を簡便に抽出するには、インターネット環境についても提供者側に自由度の高い従来のWeb servicesから計算機がより扱いやすいSemantic Webへの移行と計算機の利用を前提としてオントロジーや辞書を取りまとめる必要性が指摘された(図11)。そして、このような研究が進展すると疾患候補遺伝子の提供(図12)や生命科学研究の展開から一般の人々への知識の広がりを実現するなど生物研究における知識の共有と理解の展開が同時に実現できる(図13)ことが示された。

展望

まとめとこれからの課題

- ・ 現在の開発動向
 - －生物知識の検索や抽出
 - －生物知識の表現
 - －パスイェデータベース
 - －オントロジー
- ・ これからの課題
 - －研究者の論理、生物論理の形式化
 - －仮説生成や検証の論理の抽出と表現
 - －機能や概念の間の関係の計算方法の開発
 - －パスイェの比較、統合などの推論技術の開発
 - －より高次の知識(組織レベルや個体レベル)への対応

図8 複雑なデータや知識の表現

出典：学会戦略ワークショップ 東京大学大学院新領域創成科学研究科教授 高木利久先生 発表スライドより

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

バイオポータルでのオントロジーの構築と利用

- ・利用者側（一般、専門外入門者）の観点：
 - －できるだけ単純化された、分かりやすいものが良い
 - －さまざまな分野から、分野横断的にアクセスできるのが良い
- ・利用者側（研究者）の観点：
 - －既知の知識だけでは不十分。新規の情報を見つけるすが欲しい
- ・構築者側の観点：
 - －構築にできるだけ手間がかからないのが良い



- ・手作業で大まかな構造を構築し、詳細構造の構築は用語間の共起性などを利用して自動化する
- ・さまざまな分野のオントロジーを用意し、共通する用語間をリンクし、分野横断的に利用できるようにする

図9 データ資源の広がり

出典：学会戦略ワークショップ 国立遺伝学研究所生命情報・DDBJ研究センター教授 菅原秀明先生
発表スライドより

生物分野専門用語（日本語）の辞書化と課題

- ・表記ゆれ・同義語・類義語を幅広く集めたシソーラスの整備
- ・専門用語のカタカナ表記や、複合語、新語の整理、収集
- ・辞書作製各ステップの自動・半自動化による作製支援法の整備
- ・既存の辞書や分子名データベースとのリンクや統合
- ・関連語の収集とオントロジ化の検討、また概念の視覚化の検討
- ・実用化にむけての課題（ウェブ、検索や自然言語処理に利用するために）

(同義語・類義語の接頭・接尾語などによる整理例)

- ・同義語タイプ ○○レセプター=○○受容体、XX遺伝子=XXタンパク質
- ・類義語タイプ ○○細胞≒○○細胞株、XX病=XX症、症候群
- ・関連語タイプ ○○タンパク質≠○○欠損症、転写XX≠転写△△

(よく用いられるカタカナ語の例)

- ・プロテオーム、発現プロファイル、キナーゼ（研究者はさらにカイネースと発音する）

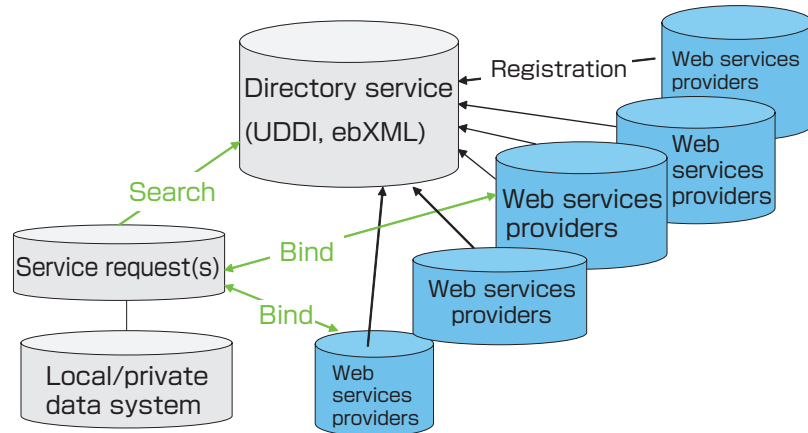
(典型的な複合語の例)

- ・膜7回貫通Gタンパク共役型受容体、副腎皮質刺激ホルモン放出ホルモン（多数の表記が存在しうる）

図10 辞書化と課題

出典：学会戦略ワークショップ 国立情報学研究所研究員 川本祥子先生 発表スライドより

まとめ：諸問題の解決と平行して、Web services さらにはSemantic Webへと展開



Stein, L. (2002) Creating a bioinformatics nation, Nature, **417**, 119-120

図11 Web servicesからSemantic Webへの展開

出典：学会戦略ワークショップ 東京大学大学院教授 辻井潤一先生 発表スライドより

疾患候補遺伝子の提供

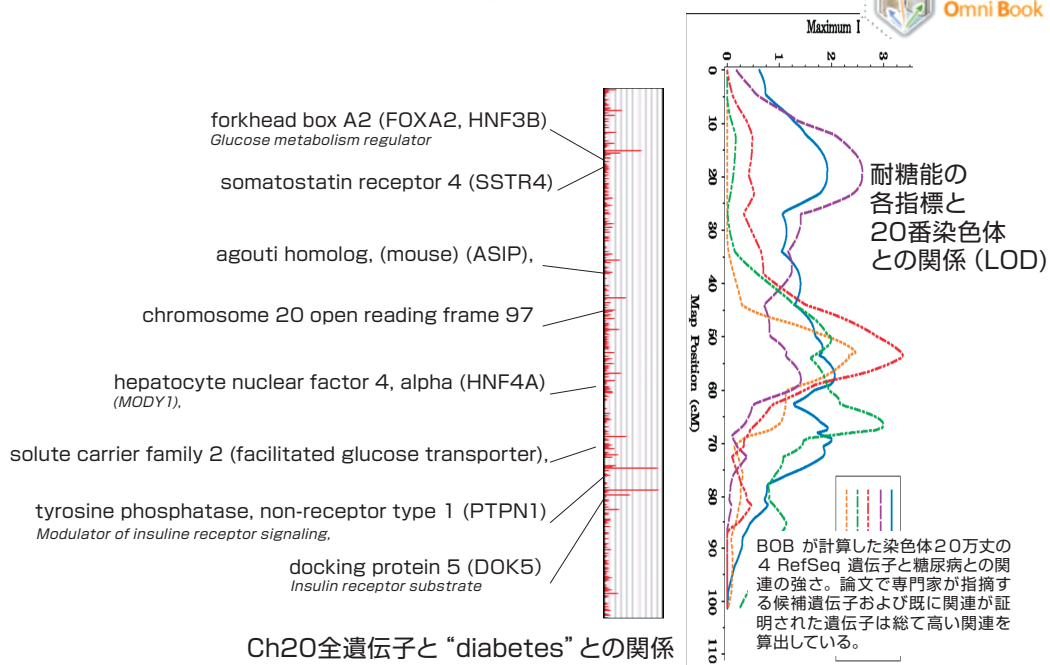


図12 疾患候補遺伝子の提供

出典：学会戦略ワークショップ 国立遺伝学研究所教授 大久保公策先生 発表スライドより

はじめに 1

学会戦略ワークショップの概要 2

21世紀の生物研究における知識の共有と理解 3

今後重要となる研究分野 4

まとめ 5

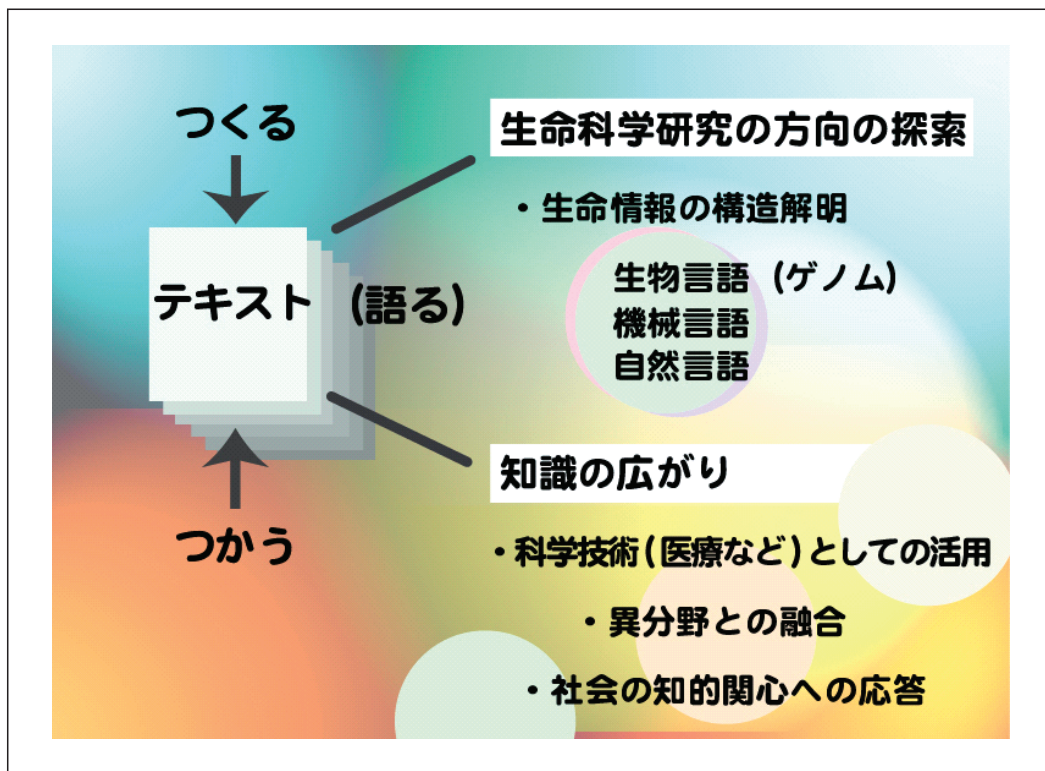


図13 生命科学研究の方向の探索と知識の広がり

出典：学会戦略ワークショップ JT生命誌館サイエンス・コミュニケーション&プロダクションチーフ代行 村田英克先生 発表スライドより

[5] まとめ

ヒト・ゲノム計画が今世紀初頭に終了した生命科学分野では爆発的なデータの増加と蓄積の時代に入った。一方で、研究分野（知識）の細分化が進み、生命科学分野の専門家においても分野の俯瞰ができにくくなりつつある現在、実験研究の対象自体に注目していたのでは、知識を体系化し、知識の再生産を図ることは不可能である。

学会戦略ワークショップでは分野の俯瞰や巨大知識を体系的に整理し、知識化する試みなど最新の研究成果を紹介し、これからの生物研究を支える知識環境にどのような革新が必要かを統合データベース、オントロジー、言語処理、生命誌などの視点から、生命知識自体を研究の対象としている専門家の発表を中心に検討した。その結果、これからの科学の発展のためには生命科学分野において爆発的に蓄積されつつある実験データや論文知識を当該分野の研究者のみならず今後重要となる融合研究での生命科学分野以外の研究者や一般の人々が知識として有効に活用できる環境を構築する必要があることが指摘された。そのためには、まず、計算機利用を前提に、計算機側で知識の処理や知識に基づいた処理が可能なシステム（知的なシステム）を構築する必要があることが指摘された。その上で、研究者や研究分野の知識、論理体系を辞書やオントロジーなどの形で表現し、計算機の持つ巨大な計算能力と膨大な記憶能力を組み合わせる新しい科学技術領域の研究開発を推進する必要があることが指摘された。

はじめに	1
学会戦略ワークショップの概要	2
21世紀の生物研究における知識の共有と理解	3
今後重要となる研究分野	4
まとめ	5

学会戦略ワークショップ
「21世紀の生物研究における知識の共有と理解」
報告書

平成17年7月

独立行政法人科学技術振興機構
研究開発戦略センター
江口グループ

Copyright 2005 by CRDS/JST

無断での転載・複写を禁じます。