

ライフサイエンス分野の統合データベース整備事業

ライフサイエンス統合データベース開発運用

(戦略立案・実行評価／統合データベース開発／統合データベース支援)

21年度 研究成果報告書

平成22年3月

大学共同利用機関法人 情報・システム研究機構	高木利久
独立行政法人 産業技術総合研究所生命情報工学研究センター	浅井 潔
財団法人 かずさディー・エヌ・エー研究所	中村保一
国立大学法人 九州大学	林 健志
国立大学法人 奈良先端科学技術大学院大学	松本裕治
国立大学法人 東京大学	森下真一
国立大学法人 お茶の水女子大学	瀬々 潤
学校法人 関西文理総合学園長浜バイオ大学	池村淑道

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、大学共同利用機関法人 情報・システム研究機構、独立行政法人 産業技術総合研究所生命情報工学研究センター、財団法人 かずさ DNA 研究所、国立大学法人 九州大学、国立大学法人 奈良先端科学技術大学院大学、国立大学法人 東京大学、国立大学法人 お茶の水女子大学及び学校法人 関西文理総合学園長浜バイオ大学が共同で実施した、平成21年度の「ライフサイエンス統合データベース開発運用」の成果を取りまとめたものです。

目 次

1. プロジェクトの目的	4
2. 戦略立案・実行評価（情報・システム研究機構）	4
2. 1 戦略立案・実行評価の実施計画	4
2. 2 戦略立案・実行評価の実施内容	5
3. 統合データベース開発	15
3. 1 共通基盤技術開発（情報・システム研究機構、生命情報工学研究センター、 奈良先端科学技術大学院大学）	15
3. 1. 1 共通基盤技術開発の実施計画	15
3. 1. 2 共通基盤技術開発の実施内容	16
(1) 共通基盤技術開発（情報・システム研究機構）	16
(2) ワークフロー技術を用いた統合 DB 環境構築 （生命情報工学研究センター）	20
(3) 専門用語辞書管理システムと専門用語解析技術の開発 （奈良先端科学技術大学院大学）	22
3. 2 ヒト統合 DB の開発・運用（情報・システム研究機構、九州大学）	26
3. 2. 1 ヒト統合 DB の開発・運用の実施計画	26
3. 2. 2 ヒト統合 DB の開発・運用の実施内容	27
(1) ヒト統合 DB の開発・運用（情報・システム研究機構）	27
(2) 多型知識表現技術の開発（九州大学）	42
3. 3 モデル生物・産業応用生物統合 DB の開発・運用 （情報・システム研究機構、かずさ DNA 研究所）	44
3. 3. 1 モデル生物・産業応用生物統合 DB の開発・運用の実施計画	44
3. 3. 2 モデル生物・産業応用生物統合 DB の開発・運用の実施内容	45
(1) モデル生物・産業応用生物統合 DB の開発・運用 （情報・システム研究機構）	45
(2) 植物及び植物関連微生物のゲノム情報データベース統合と高度化 （かずさ DNA 研究所）	46
4. 統合データベース支援（情報・システム研究機構）	48
4. 1 ポータル整備・運用、広報、普及啓発	48
4. 1. 1 ポータル整備・運用、広報、普及啓発の実施計画	48
4. 1. 2 ポータル整備・運用、広報、普及啓発の実施内容	49
(1) 統合 DB プロジェクトポータルサイトの整備・運用	49
(2) 広報・普及啓発	53
4. 2 データベース受入	55

4.2.1	データベース受入の実施計画	55
4.2.2	データベース受入の実施内容	55
4.3	人材育成(東京大学、お茶の水女子大学、長浜バイオ大学)	63
4.3.1	人材育成の実施計画	63
4.3.2	人材育成の実施内容	64
(1)	DB構築者の養成(東京大学)	64
(2)	DB高度利用者の養成(お茶の水女子大学)	66
(3)	アノテータ・キュレータの教育(長浜バイオ大学)	67
5.	プロジェクトの総合的推進(情報・システム研究機構)	68
5.1	プロジェクトの総合的推進の実施計画	68
5.2	プロジェクトの総合的推進の実施内容	68
5.3	その他シンポジウム、セミナー等	70
(1)	統合データベース講習会(情報・システム研究機構、長浜バイオ大学、 かずさDNA研究所)	70
(2)	BioHackathon2009(情報・システム研究機構)	75
(3)	バイオインフォマティクスへの招待(お茶の水女子大学)	77
(4)	アノテータ、キュレータ養成のための運営委員会と公開イベント (長浜バイオ大学)	78
6.	成果の外部への発表	80
7.	特許出願状況	80
8.	実施体制	80
9.	整備実績一覧	80

1. プロジェクトの目的

ライフサイエンスやバイオ産業に従事する研究者や技術者がいわゆるゲノムプロジェクト・ポストゲノムプロジェクトの成果や多様な DB や解析ツールをストレスなく利用してより高度な研究開発が効率よくできる環境（統合 DB）を実現することを目的とする。このため、情報・システム研究機構、科学技術振興機構、産業技術総合研究所生命情報工学研究センター、かずさ DNA 研究所、九州大学、奈良先端科学技術大学、東京大学、お茶の水女子大学、長浜バイオ大学は共同して、以下の3つの業務を行う。

(1) 戦略立案・実行評価

ライフサイエンス（基礎生物学および応用生物学（医・薬・農））、情報処理技術、ライフサイエンス DB の3分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した我が国の DB 整備戦略を立案する。また、3分野の動向に関する情報の網羅的収集・分析を行い、戦略立案や実行評価を支援するチームを組織する。

(2) 統合データベース開発

ライフサイエンス、バイオ産業にかかわる情報へのアクセスと利用に関する格段の利便性向上とそれによる研究開発の飛躍的な効率化と質的向上を目指して、ヒト統合化 DB およびモデル生物・産業応用生物統合 DB の開発・運用と、それに必要な統合 DB 構築技術、DB の標準化、文献等からの知識獲得技術等の共通基盤技術開発を実施する。

(3) 統合データベース支援

個々の公開 DB や利用可能な DB 解析ツールなどを網羅的に調査し、提示するポータルサイトを構築する。運用にあたって専門家及び利用者の意見集約システムを開発し、意見をフィードバックして運用する。また日本語による研究情報の流通促進に注力する。プロジェクトデータベースを受入れ、相互運用可能な DB へ変換した上で公開し、運用する。また、広報・普及啓発の活動を行う。人材育成は実績のある機関と連携し階層別に、アノテータ・キュレータ教育、DB 構築者養成、及び DB 高度利用者養成を行う。

2. 戦略立案・実行評価（情報・システム研究機構）

2. 1 戦略立案・実行評価の実施計画

ライフサイエンス（基礎生物学および応用生物学（医・薬・農））、情報処理技術、ライフサイエンス DB の3分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した我が国の DB 整備戦略を立案する。上記委員会に対し、3分野の動向に関する情報の網羅的収集・分析を行い、戦略立案や実行評価を支援するチームを組織する。このよう

な体制のもとで、戦略立案、実行評価の支援業務とそのための情報収集・分析だけでなく、関係府省、利用者、産業界、学会・出版社、国内外の研究機関等関係者との連絡調整等の活動もあわせて行う。

1) 戦略立案

平成 20 年度までに、統合データベースプロジェクトの目標や年次計画の具体化、国内外の動向調査も踏まえたデータベース統合化のあるべき姿の検討を進めてきた。併せてデータ共有化や公開促進の観点から、国家プロジェクト成果に関する権利関係や著作権、個人情報保護に関する調査を行い、プロジェクト成果であるデータの提供を促進するための取り組みや提供いただいたデータの運用方針の立案、個人情報に関わるデータの公開方針の策定などを行ってきた。さらに、総説誌など日本語文献の公開に関する取り組みも進めてきた。平成 21 年度は、これまでの成果をベースに、特に本プロジェクト終了後の進め方も念頭においた、我が国におけるライフサイエンスデータベース統合の在り方の検討を進めると同時に、引き続き国家プロジェクト等により産生されるデータおよびその他有用な関連情報の共有および公開を実現するための方策を推進する。また、国内外の研究機関等関係者との連絡調整という観点から、平成 19 年度、20 年度に、世界的に分散しているデータの統合利用に関する国際ワークショップを開催し、この分野の世界的なハブとしての役割を務めた。平成 21 年度はこれまでの活動の成果も踏まえ、データの共有、利用を促進するための世界的な発信源となるための活動を継続する。

2) 実行評価

平成 20 年度は、想定されるユーザドメインから、有用な回答が期待できる有識者を選定し、本プロジェクト成果のユーザ評価を実施した。また出展した学会等の展示会でのアンケート結果も踏まえ、逐次サービス内容の改善を行った。併せて、サービスサイトへのアクセスログを持続的に収集する体制を整え、ユーザからの利用状況の把握に努めた。平成 21 年度は、より評価者を充実した上でのユーザ評価の実施、また展示会等でのアンケートを含めて、さらにユーザの生の声の把握に努め、引き続きサービス内容のユーザニーズへの適合に努める。また、アクセスログについてはプロジェクト参加機関からの情報を入手する体制を整え、プロジェクト全体としての利用状況の把握に努める。

2. 2 戦略立案・実行評価の実施内容

1) 戦略立案

戦略立案については、これまで進めてきた統合データベースプロジェクトの目標や年次計画の具体化、国内外の動向調査をベースに、データベース統合化のあるべき姿の検討を進めてきた。併せてデータ共有化や公開促進の観点から、著作権、個人情報保護に関する調査、検討、並びに今後のデータ産出の中心となることが想定される次世代シーケンサの生み出すデータに関する調査を平成20年度に引き続き行った。また、日本語文献の公開についても、コンテンツを充実すべ学協会等との交渉を積極的に行い、日本溶

媒抽出学会 (http://lifesciencedb.jp/lbdb.cgi?gg=tool_academy) や歯科基礎医学会 (http://lifesciencedb.jp/lbdb.cgi?gg=tool_academy) の学協会カタログ、日本における生物多様性関連Webサイト一覧 (<http://lifesciencedb.jp/lbdb.cgi?pg=1>) や日本分類学連合 (<http://lifesciencedb.jp/lbdb.cgi?pg=1>) のカタログ、そして原生生物情報サーバー (<http://lifesciencedb.jp/lbdb.cgi?pg=1>)、日本産糸状菌類図鑑 (<http://lifesciencedb.jp/lbdb.cgi?pg=1>)、病原真菌データベース (<http://lifesciencedb.jp/lbdb.cgi?pg=1>) 等を追加し日本語コンテンツの充実を図った。さらに、国内外の研究機関等関係者との連絡調整という観点から、昨年平成20年度に引き続き、世界的に分散しているデータの統合利用に関する国際ワークショップである **BioHackathon**を開催し(参加者は海外32名、国内36名)、今回はデータベースのより高度な収集と活用を目的にセマンティックウェブについて国内外の主要な研究者が集まった(3. 1. 2項項参照)。

データベース統合化の今後の進め方については、平成21年1月の文部科学省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の「ライフサイエンスデータベースの統合・維持・運用の在り方」に関する報告書、平成21年5月の総合科学技術会議ライフサイエンスPTにおける統合データベースタスクフォースの報告書で、本プロジェクト終了後も、本プロジェクトの成果である統合データベースを恒久的に維持、発展させる必要があること、また、そのための具体策として、本プロジェクトと科学技術振興機構(JST)のBIRD事業を一本化するための組織をJST内に構築するべきである、との提言がなされた。そこで、本プロジェクトでも、平成20年度に松原謙一研究運営委員長のリーダーシップの下、運営委員の先生方から収集したライフサイエンス統合データベースのあるべき姿に関するご意見をベースに、本庶 佑総合科学技術会議議員をはじめ、井村裕夫京都大学名誉教授、平田 正協和発酵キリン名誉相談役、水上政之科学技術振興機構理事をお招きして開いた研究運営委員会拡大懇談会や通常の研究運営委員会の場での議論などを通じて、上記のライフサイエンスPTタスクフォースの報告書で設立が提言されている統合データベース推進本部(仮称)への申し入れ事項の案文として、以下の9項目がまとめられた。

1. これからのライフサイエンスのブレークスルーは、存在する莫大なデータから仮説を創生してそれを検証する、データドリブン型の研究により達成される。
2. 広くデータを共有してこそイノベーションの加速化が実現できる。従って、我が国バイオ産業の発展には、網羅的に効率よくデータにアクセスできる環境が不可欠である。
3. データ共有できる環境を構築するためには、多岐にわたる対象や階層のデータベースの統合化とこれを構築、維持する機関の存在が必須である。
4. 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。
5. 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。

6. 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べきである。

7. 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。

8. 統合データベースの有用性を維持するためには、恒常的に使い勝手やデータの付加価値向上を図る必要があり、そのための研究開発機能を備える必要がある。

9. 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不十分であり、統合データベース構築と連携した人材育成やキャリアパスの構築が必要である。

データ共有や個人情報保護については、分担機関の一つである東京大学の「疾患解析から医療応用を実現する DB 開発」において、共有方針の作成検討にオブザーバーとして参加し、共有方針や登録・利用申請に関する書面の作成を担当した。書面は国外からの申請も考慮して、日本語とともに英語版も作製した。これらの文書は、12月に「疾患解析から医療応用を実現する DB 開発」ウェブサイトから公開された。

当プロジェクトの実施期間も半ばを過ぎたことから、中間での成果発表の位置づけとして、また同時に、データベースを取り巻く様々なコミュニティメンバーが一堂に会する機会として、平成 21 年 6 月 12 日に東京大学本郷キャンパス浅野地区内武田ホールにおいて、統合データベースプロジェクトシンポジウム 2009「データベースが拓くこれからのライフサイエンス」を開催した。講演 7 件、パネル討論、ポスター発表（プロジェクト内外問わず募集）85 件からなるシンポジウムに、当日登録者も含め、のべ 393 名の参加者（含む関係者）があった。発表者に後日依頼してポスター PDF を提出してもらい、ウェブサイトで公開した（<https://symposium.lifesciencedb.jp/2009/>）。



図 2.1. 6 月シンポジウムのポスターと会場の様子

データベースの受入・公開にともなう著作権問題の解決方法の一つとして当プロジェクトではクリエイティブ・コモンズ・ライセンスを採用している。同ライセンスの創始者である米国の法律学者が来日する機会をとらえ、この問題を広く一般に問いかけるために、国内の著作権法関係者、欧州での動向に詳しい関係者らを集めて、平成 21 年 10 月 5 日に東京大学本郷キャンパス弥生地区内弥生講堂において、統合データベースプロジェクトシンポジウム「科学における情報の上手な権利化と共有化」を開催した。参加者は 257 名（含む関係者）で、テーマ内容に関連してライフサイエンス関係者以外に法学や出版関係者も集うシンポジウムとなった。講演の内容は動画としてウェブサイトで公開した（<http://symposium.lifesciencedb.jp/IPDS/>）。



図 2.2. 10 月シンポジウムのウェブサイトと動画配信

2) 実行評価

実行評価については、平成 20 年度に依頼した有識者に、学会・展示会等でアンケートにより募集したユーザを加え、計 291 名に評価を依頼し、本プロジェクト成果のユーザ評価を実施した。また出展した学会等の展示会でのアンケート結果も踏まえ、逐次サービス内容の改善を行った。また、平成 20 年度に体制を整えたサービスサイトへのアクセスログ収集を持続的に実施し、ユーザからの利用状況の把握に努めた。

ユーザ評価については、平成 20 年度に評価を依頼した 92 名の有識者に加え、各学会や展示会でのアンケートの際に募集したユーザ評価への参加希望者を合わせて、計 291 名に依頼した。平成 20 年度に開発した本プロジェクトの公開サービスを中心に、中核機関以

外のサービスについてもほぼすべてを対象としたため多数となった評価対象サービスをデータベース統合化のステップごとに分類して、ステップごとに評価ページを作成し、各サービスごとに新たに開発された内容を記述するなどして、2度以上評価に参加しているユーザにもわかりやすい表記を心掛けた。また、平成20年度の評価の際に要望のあった、成果目標や予算配分の資料も併せて掲載した。各サービスへの評価は平成20年度の評価の際に作成したサービス紹介も兼ねたウェブサイト記入方式（5段階評価とフリーコメント）を応用した（図2.3）。回答数は項目ごとに異なるが、のべ117名（回答率40%、評価者の背景を図2.4.に示した）であった。ステップごとのフリーコメントは16~104件（合計213件）あり、全体に対するフリーコメントとして「今後当プロジェクトで提供すべきサービスについて」は29件、「当プロジェクトが行っている情報提供について」は11件（5段階評価も実施）、「ユーザー評価の方法/形式について」は20件あった。また、これらのフリーコメントに対して回答可能なものについては、担当者からの回答を併記して、ウェブサイト“平成20年度「統合データベースプロジェクト」事業成果に対するユーザー評価の結果”（<http://lifesciencedb.jp/feedback/hyouka20/?result>）から公開した（図2.5）。これらのコメントのうち、すぐに実現可能なものは実施し、次年度の開発に盛り込めるものは盛り込んだ。

The image shows a screenshot of the evaluation website. On the left is a sidebar with navigation links like '評価サイトトップ' and '観覧と専門分野について'. The main content area is titled 'サービスのステップごとの分類説明' and contains a flowchart of the project steps: Step 1 (Database, Tools, Project Catalog), Step 2 (Database and Tool Integration), Step 3 (Integration and Workflow), and Step 4 (Integration and Workflow). Below the flowchart is a detailed evaluation form for 'ステップ1. データベース、ツール、プロジェクトカタログ'. This form includes a 'TEST SEARCH' button, a table for '観覧と専門分野について' (Viewing and Specialties), and a table for '観覧数' (Viewing Count).

観覧数	件数	割合
学生	4	3.0%
ポストドク研究員	14	11.3%
教員(助研・助産)	20	15.3%
教員(准教授)	0	0.0%
少研研究員(研究員)	11	10.5%
大学院研究員(学生)	4	3.0%
民間企業 研究員	24	22.9%
民間企業 その他	10	9.5%
その他	2	1.0%

観覧数	件数	割合
かなり役に立ちそうなので、どんどん進めてほしい	33	28.2%
それなりに役に立ちそうなので、できる範囲で進めてほしい	28	42.4%
責任を持って評価できない	5	7.6%
あまり役に立たなそうなので、進めなくてもよい	0	0.0%
まったく役に立たなそうなので、やめたほうがよい	0	0.0%
投票総数	66	

図 2.3. 21 年度評価記入サイトの一部

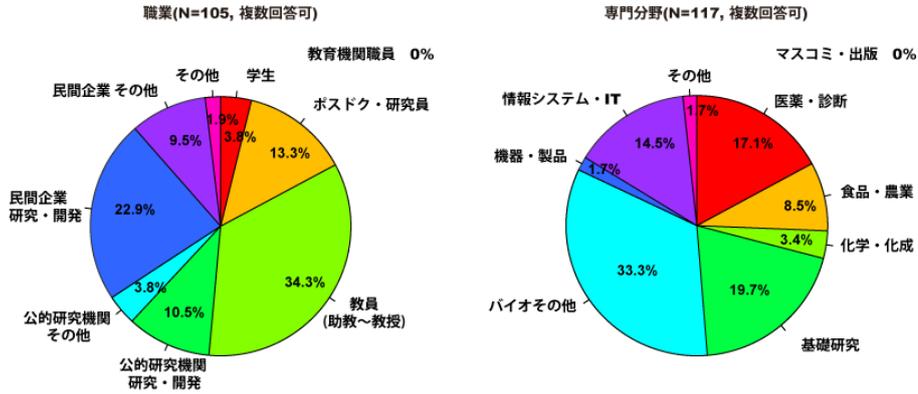


図 2.4 21 年度評価者の背景



図 2.5. 21 年度評価結果サイトの一部

表 2.1 サービスごとの 5 段階評価の結果

評価 5 : かなり役立ちそうなので、どんどん進めてほしい
評価 4 : それなりに役立ちそうなので、できる範囲で進めてほしい
評価 3 : 責任を持って評価できない
評価 2 : あまり役に立たなそうなので、進めなくてもよい
評価 1 : まったく役に立たなそうなので、やめたほうがよい

ステップの名称					
データベース、ツール、プロジェクトのカタログ化、ポータルサイトの構築（ステップ1）					
サービス名	評価				
	5	4	3	2	1
生命科学系データベースカタログ	33	28	5	0	0
生命科学系学協会カタログ	20	28	7	3	0
ゲノム・ポストゲノム主要プロジェクト一覧	24	30	5	1	1
DBCLS メタ用語集	24	20	11	2	0
生物アイコン	17	21	10	6	2
アナトモグラフィ/BodyParts3D	22	12	15	0	0
WINGpro	15	19	11	3	0
MDeR	11	13	22	3	1
データベースやツールの使い方と教材の提供（ステップ2）					
サービス名	評価				
	5	4	3	2	1
統合 TV	31	21	5	1	2
Web リソースポータルサイト	16	19	13	1	0
「バイオデータベースサーバー構築演習」の取り組み	15	14	16	1	1
「データベース高度利用者養成」の取り組み	14	18	12	1	2
「アノテータ・キュレータの教育」における取り組み	12	20	14	2	0
目的、用途ごとのデータベース統合化、解析ワークフロー（ステップ3）					
サービス名	評価				
	5	4	3	2	1
標準用 SNP データベース	19	15	15	2	0
GWAS データベース	12	17	21	1	0
日本糖鎖科学データベース 横断検索	13	18	19	1	0
日本糖鎖科学データベース 構造検索	15	9	22	0	0
日本糖鎖科学データベース 分子量検索	11	9	27	0	0
日本糖鎖科学データベース 単糖組成検索	13	9	24	0	0
ゲノムネット医薬品データベース	21	17	9	1	1
タンパク質アノテーションワークフロー	19	19	7	1	0

タンパク質比較情報ワークフロー	16	18	7	1	0
統合医科学データベース	14	10	19	1	0
かずさアノテーション	10	18	17	0	0
生体分子の熱力学データと構造データの統合	7	12	23	0	1
植物オミックス情報およびタンパク質構造情報	14	12	17	0	0
データベースやツールの統一的、シームレスな検索、利用の支援（ステップ4）					
サービス名	評価				
	5	4	3	2	1
生命科学データベース横断検索	32	16	2	1	1
生命科学系データベースアーカイブサービス	20	17	7	2	1
Allie	16	21	6	2	1
OReFiL	12	26	5	2	0
Wired-Marker	8	16	15	3	1
DNA データベース総覧と検索	10	21	8	3	0
遺伝子発現バンク(GEO)目次	14	15	9	2	2
DDBJ トレースアーカイブ/DDBJ リードアーカイブ	14	13	15	1	0
当プロジェクトが行っている情報提供について（その他）					
サービス名	評価				
	5	4	3	2	1
プロジェクトが目指すものに関する情報提供について	33	27	3	1	1
学会等でのサービスやツールに関わる研究開発内容の発表について	36	18	4	2	0

また、アクセスログの収集と集計については、平成 20 年度と同様に、集計ソフト AWStats ver.6.6 (<http://www.starplatinum.jp/awstats/awstats66/>)を用いて、統合 HP(<http://lifesciencendb.jp>)にリストされている各項目（ただし DBCLS ホームページ：<http://www.dbcls.rois.ac.jp> を除く）を 1 サービスとみなし、その単位で行った。集計期間は平成 21 年度（平成 21 年 4 月～平成 22 年 3 月）である。

アクセスログを取得している中核機関のサービス数は以下の通りである。

- 平成 20 年度末：33 サービス
- 平成 21 年度末：39 サービス

AWStats が出力する集計値としては主に以下の 3 つがある。

- 一意な訪問者数：アクセス元の IP アドレス数。何度アクセスしても同一 IP からの

アクセスは1と計数する。

- 訪問回数（セッション数）：アクセス元の IP アドレス数。1 時間以上の間隔をおいての同一 IP からのアクセスは別訪問と計数する。
- ページ数（いわゆるページビュー）：各ページへのアクセス数

ユーザ数の集計値としては、一意な訪問者数を採用した。ただし、企業、教育機関、プロバイダなどの組織からインターネットへ接続する1つの IP アドレスを経由して複数のユーザが訪問しているため過小評価となる。

また、中核機関全体の年度内の一意な訪問者数は以下の通りである。

- 平成 21 年度：331,072
- 平成 20 年度：181,932
- 前年比：182%

ただし、サービス別訪問回数の合計であるため、例えば1ユーザが2つのサービスを訪問した場合は2と計数している。サービス別上位5件は以下の通りである。

表 2.2 平成 21 年度一意な訪問者数上位 5 サービス

順位	サービス名	開始時期	一意な訪問者数	前年比
1	生命科学データベース横断検索	2007/10～	97,392	177%
2	DNA データベース総覧と検索 (DDBJ/EMBL/GenBank)	2007/10～	44,054	129%
3	統合 TV (DB やツールの動画教材)	2007/7～	30,163	153%
4	文科省「ゲノム」研究報告書 全文検索	2008/5～	18,772	279%
5	アナトモグラフィ/BodyParts3D	2007/10～	18,760	268%

アクセス数の集計値としては、訪問回数（セッション数）を採用した。ページ数はコンテンツをダウンロードした回数や検索を実行した回数を反映するが、1つの操作で複数のページを表示する場合などは過大評価となるため採用しなかった。訪問回数も一意な訪問者数と同様に IP アドレス数であるので複数のユーザが短時間に訪問した場合は過小評価になる。

また、中核機関全体の年間訪問回数は以下の通りである。

- 平成 21 年度：937,600 回
- 平成 20 年度：436,203 回
- 前年比：215%

ただし、サービス別訪問回数の合計であるため、例えば1ユーザが1時間以内に2つのサービスを訪問した場合は2回と計数している。サービス別上位5件は以下の通りである。

表 2.3 平成 21 年度訪問回数上位 5 サービス

順位	サービス名	開始時期	平成 21 年度 訪問回数	前年比
1	生命科学データベース横断検索	2007/10～	264,345	212%
2	統合 TV (DB やツールの動画教材)	2007/7～	144,740	166%
3	DNA データベース総覧と検索 (DDBJ/EMBL/GenBank)	2007/10～	135,501	245%
4	統合ぐらし	2007/9～	50,544	112%
5	文科省「ゲノム」研究報告書 全文検索	2008/5～	37,785	339%

生物科学データベース横断検索や統合 TV など生命科学分野全般に渡った幅広いユーザを対象にしたサービスは研究関係者だけでなく一般からのアクセスも含まれると考えられるため一意な訪問者数や訪問回数が多い。別の指標として 1 訪問者あたりの年間平均訪問回数 (= 訪問回数 / 一意な訪問者数) がある。これはリピータ数を示すものであり、以下が上位 5 件である。

表 2.4 平成 21 年度年間平均訪問回数上位 5 サービス

順位	サービス名	年間平均 訪問回数	一意な訪問者数	訪問回数
1	国内ウェブサービスの統合 TogoWS	5.46	2,981	16,275
2	DBCLS galaxy	4.87	740	3,607
3	統合 TV (DB やツールの動画教材)	4.80	30,163	144,740
4	inMeXes	3.98	2,516	10,026
5	MiGAP	3.95	930	3,673

また、平成 20 年度との比較については、平成 21 年度の訪問回数が前年度割れしているサービスは、26 サービス (比較可能と判断した平成 20 年 7 月以前に集計を開始したもの) 中 4 サービス (内、2 件は統合 DB 事業 (過去のコンテンツ)) であり、ほとんどのサービスで訪問回数が増加していた。

さらに、ドメイン別のアクセス集計 (訪問回数ではなくページ数で集計) は以下の通りである。

表 2.5 ページ数ドメイン別内訳

ドメイン	平成 21 年度		平成 20 年度	
	ページ数	割合	ページ数	割合
ac/edu	977, 104	10%	659, 877	18%
co	454, 635	5%	466, 048	13%
go/gov	133, 854	1%	149, 242	4%
ne	1, 319, 118	13%	701, 412	19%
others	6, 894, 492	71%	1, 638, 901	45%
合計	9, 779, 203	100%	3, 615, 480	100%

平成 20 年度と比較すると、組織を特定できないドメインからのアクセスが増加しているため、教育機関(ac/edu)、企業(co)、政府系機関(go/gov)の割合は減少している。ただ、実数にはそれほど変化がなく、教育機関では増加している。

なお、中核以外の機関についても、平成 20 年度後半に収集を依頼した機関から約 1 年分のアクセスログを提供していただき、AWStats で集計を行った。

3. 統合データベース開発

3. 1 共通基盤技術開発（情報・システム研究機構、生命情報工学研究センター、奈良先端科学技術大学院大学）

3. 1. 1 共通基盤技術開発の実施計画

（1）（情報・システム研究機構）

平成20年度に生命科学系論文を対象とした各種文献処理に必要な環境整備が整いつつあるため、平成21年度は言語横断的な文献検索システムや質問応答システム、情報抽出システムを実現するための技術開発および必要な調査を行う。また、平成20年度に引き続き、生命科学研究者の興味を反映した文献情報の検索、整理が容易にできる環境の構築を行う。さらに、これまで開発してきたTogoDB、TogoWSを利用して、研究者間で効率的にデータを共有し、活用するために必要な技術開発を行う。また、統合検索に向けた基盤技術として、セマンティックウェブ技術に関する調査を行い、必要となる技術を開発する。そして、上述の技術や開発成果を組み合わせ、容易に利用可能とするインターフェースの開発を、平成20年度のインターフェースの調査結果を生かして推し進め、サービスとして公開する。

（2）（産業技術総合研究所生命情報工学研究センター）

平成 21 年度は、平成 20 年度に構築し一般公開した「予測とデータ取得によりタンパク

質に関する網羅的な情報を得るワークフロー」と「タンパク質の比較情報を提示し、保存部位、変異部位を推定するワークフロー」の成果を基に、以下の業務を行う。なお、その際には、情報・システム研究機構が開発している DB との統合にも配慮し一般公開する。

1) タンパク質立体構造予測ワークフローの開発

アミノ酸配列を入力とし、タンパク質立体構造のテンプレートとなる構造をタンパク質立体構造データベース (PDB) から見つけ出し、それを基にタンパク質立体構造をモデリングするワークフローを開発する。

2) 利用者の要望を考慮したワークフローの開発

公開したワークフローの利用者の意見や、利用者が求めている統合 DB を調査し、その調査結果を考慮した統合 DB システムを設計・開発する。

(3) (奈良先端科学技術大学院大学)

1) 専門用語獲得技術

①専門用語辞書システムの開発

専門用語の内部構造や同義語識別子を用いた検索や専門用語の意味情報を管理することができる辞書システムの設計を行う。

②専門用語解析技術の開発

専門用語の内部構造解析のための学習データとして内部構造解析済みデータを拡大し、意味情報などこれまで用いなかった情報を利用して、内部構造解析の高性能化を行う。

2) 専門用語抽出ツールの設計と開発

①専門用語辞書拡張支援ツールの設計と開発

専門分野のテキストに現れる新規の専門用語 (新規語) の意味分類を行う手法の高性能化を行い、その結果を利用して、新規語を辞書に登録するための支援ツールを設計する。専門分野のテキストに対して、用語の出現箇所の特定制や用語の意味分類を行う自動タグ付け手法を設計する。

専門文書から専門用語を自動抽出するための手法の設計およびそのツール化を目指す。また、抽出された用語を管理するための辞書システムの設計を行う。

3. 1. 2 共通基盤技術開発の実施内容

(1) 共通基盤技術開発 (情報・システム研究機構)

言語横断的な質問応答システム技術開発として、日本語自然文を入力とし、MEDLINE の関連文献を関係のある文と共に表示するシステムのプロトタイプを作成した。このシステムは主に次の三ステップからなる：(1) 日本語自然文を解析して、主語・述語・目的語を抽出する、(2) ステップ(1)で抽出した主語・述語・目的語から、MEDIE(テキストマイニング技術を用いた知的 MEDLINE 検索システム)のクエリを生成する、(3) ステップ(2)

で作成したクエリを MEDIE 上で実行し、その結果を表示する。図 3.1 は「ミオシンと相互作用するタンパク質は何ですか？」を入力とした場合の結果画面である。質問と関連する MEDLINE の論文を表示し、そのアブストラクト内で質問と直接関連する文に下線を引いている。これにより、質問の答えとなる可能性が高い文とその文が含まれる文献が示される。

MEDLINE には 2000 万件近くに上る生命科学系の書誌情報が格納されているが、研究者がこの膨大な書誌情報から自身に興味のあるものを効率的に検索することは困難になっている。一方、多くの研究者は自身の研究分野における興味のある文献リストを自身の PC に保持している。近年のオンラインジャーナルの増加に伴い、多くの文献は PDF ファイルとして保存されていることも多い。そこで、この研究者の既に保持している文献情報を基に MEDLINE データ全体を対象として関連度の高いものを高速に検索し、提示するシステム TogoDoc を平成 20 年度より開発している。平成 21 年度は、利用者が自身の文献情報や文献 PDF ファイルを効率的に整理出来る環境を提供するために開発されている TogoDocClient (東京大学による)との連携機能の開発を中心に行い、システムを公開した。TogoDocClient は利用者の PC にインストールして動作させるソフトウェアで(図 3.2)、利用者の PC 中にある文献 PDF ファイルから自動的に書誌情報を抽出する機能を持つが、本機能を支援するシステムや、得られた文献情報、更には文献 PDF ファイルを TogoDoc に保存したり取り出したりする機能を開発した。また TogoDoc に保存されている文献情報を基にした推薦機能を TogoDocClient において利用出来る仕組みを構築した。

入力日本語文： ミオシンと相互作用するタンパク質は何ですか
MEDIE INPUT: <http://www-tsuji1.is.s.u-tokyo.ac.jp/medie/dbcls.cgi?verb=interact&object=myosin+|+protein&sm=1&count=20>

GENE DISEASE COMPOUND SYMPTOM ENZYME DRUG

PMID: 19840951
Verb: affinity
Object: myosin

TITLE: D-loop of **actin** differently regulates the motor function of **myosins II** and **V** .
ABSTRACT:
To gain more information on the manner of actin-myosin interaction, we examined how the motile properties of **myosins II** and **V** are affected by the modifications of the **DNase I** binding loop (D-loop) of **actin** , performed in two different ways, namely, the proteolytic digestion with **subtilisin** and the M47A point mutation. In an in vitro motility assay, both modifications significantly decreased the gliding velocity on **myosin II**-heavy meromyosin due to a weaker generated force but increased it on **myosin V** . On the other hand, single molecules of **myosin V** "walked" with the same velocity on both the wild-type and modified **actins** ; however, the run lengths decreased sharply, correlating with a lower affinity of **myosin** for **actin** due to the D-loop modifications. The difference between the single-molecule and the ensemble measurements with **myosin V** indicates that in an in vitro motility assay the non-coordinated multiple **myosin V** molecules impose internal friction on each other via binding to the same **actin** filament, which is reduced by the weaker binding to the modified **actins** . These results show that the D-loop strongly modulates the force generation by **myosin II** and the processivity of **myosin V** , presumably affecting actin-myosin interaction in the actomyosin-ADP. **P** (i) state of both **myosins** .

GENE DISEASE COMPOUND SYMPTOM ENZYME DRUG

PMID: 19956621
Verb: interaction
Object: DNA-binding partners, including the basic helix-loop-helix transcription factor Hes1, a prototypical anti-neurogenic WRP (W/Y) motif protein

TITLE: Cofactor-activated phosphorylation is required for inhibition of cortical neuron differentiation by Groucho/TLE1.
ABSTRACT:
BACKGROUND: Transcriptional co-repressors of the Groucho/transducin-like **Enhancer** of split (Gro/TLE) family regulate the expression of a variety of **genes** and are involved in numerous developmental processes in both invertebrate and vertebrate species. More specifically, Gro/TLE1 participates in mechanisms that inhibit/delay the differentiation of cerebral **cortex** neural progenitor cells into neurons during mammalian forebrain development. The anti-neurogenic function of Gro/TLE1 depends on the formation of protein complexes with specific DNA-binding **transcription factors** that engage Gro/TLE1 through **WRB** (W/Y) sequences. Interaction with those transcription partners results in Gro/TLE1 recruitment to selected **DNA** sites and causes increased Gro/TLE1 phosphorylation. The physiological significance of the latter event, termed " cofactor-activated phosphorylation, " had not been determined. Therefore, this study aimed at clarifying the role of cofactor-activated phosphorylation in the anti-neurogenic function of Gro/TLE1. METHODS AND PRINCIPAL FINDINGS: A combination of site-directed mutagenesis, mass spectrometry, biochemistry, primary cell culture, and

図 3.1 自然文をクエリとする MEDLINE 検索システムの出力例



図 3.2 TogoDoc と TogoDocClient の関係



図 3.3 BioHackathon 2010 集合写真。国内 36 名、海外 32 名の参加者があった。

また、統合検索に向けた基盤技術として、セマンティックウェブ技術に関する調査と開発を目的とし、国内外の研究者を招聘等した国際開発会議（BioHackathon 2010）を開催した(図 3.3、5.3(2)参照)。統合検索は、統合化されたデータの上に構築される高度なデータ検索と閲覧支援によって構成され実現するものである。内外の研究者とともに、医学生物学分野における、いま現在の最先端のデータ統合技術やその応用についての調査検討をおこない、当センターが今後の統合検索実現化に向けて必要なリソースや技術、開発する

べきことについての情報を得ることができた。セマンティックウェブ技術を応用した **LinkedData** 化が現在でのデータ統合のもっとも有望な技術であるという見地から、国内外のデータプロバイダーの実務者レベルでの利用技術の観点における合意形成がおこなわれた。それに基づいて、国際協力的に、データ統合化やそのサービスの開発をすすめる予定である。さらに、データの応用部分での議論に基づいて開発を行い、その結果、萌芽的なものではあるが、ユーザインターフェース部分やデータ統合化についてのいくつかのあたらしい技術的開発的取り組みを実施した。

また、表形式データの共有基盤として開発している **TogoDB**、データベースからのウェブサービスとデータ取得層の統合化をおこなっている **TogoWS**、**Galaxy** ベースのワークフローインターフェース **DBCLS Galaxy** の機能拡張開発を実施した。

TogoDB では、利用ユーザの利用実態の調査やフィードバックをもとに、すでにある機能をユーザにとって使い易くなるような拡張を主に行った。たとえば、カラム追加機能やデータインストールを繰り返すときの手間を省く機能、表示・編集項目の選択機能、リンク支援機能などである。新規機能拡張としては、**TogoWS REST API** 経由でのデータ取得や **DBCLS Galaxy** からのデータの入力機能があげられる。これらの連携機能は、ユーザにデータ共有の効率化のためのデータ中心のワークフローを提供する基盤となる。

TogoWS では、これまで内外の主要なデータベースプロバイダー (**NCBI**、**EBI**、**DDBJ**、**KEGG**、**PDBj**) のウェブサービスに対応して、ウェブサービスとデータ取得層の統合化をすすめてきた。ユーザは、データプロバイダーごとに違いを意識することなく、そのなかのデータを取得することができる。この枠組みにおいて、対応データベースと対応フォーマット形式の拡充が課題となっており、引き続きその開発を進めた。データ取得層の今後の方向性としては、セマンティックウェブ技術を応用した **LinkedData** 化が、今後の有望なデータ統合の姿であることが、**BioHackathon 2010** での議論により得られ、そこに向けた拡張の開発を行った。

DBCLS Galaxy は、平成 20 年度開発した **Galaxy** ベースのワークフローインターフェースのプロトタイプを改良し、平成 21 年度から公開したサービスである。平成 21 年度開発した新規機能として、**DBCLS Galaxy** での解析結果を **TogoDB** へ出力するツールが組み込まれ、上記の **TogoDB** の **DBCLS Galaxy** からのデータの入力機能と繋げて、**DBCLS Galaxy** で解析した結果をスムーズに **TogoDB** から公開できるようにした (図 3.4)。また、**BioHackathon 2010** での議論を受け、セマンティックウェブ技術を用いた知識ベース **freebase** (<http://www.freebase.com/>)からの **RDF** データ取得など、セマンティックウェブ技術関連の開発を行った。

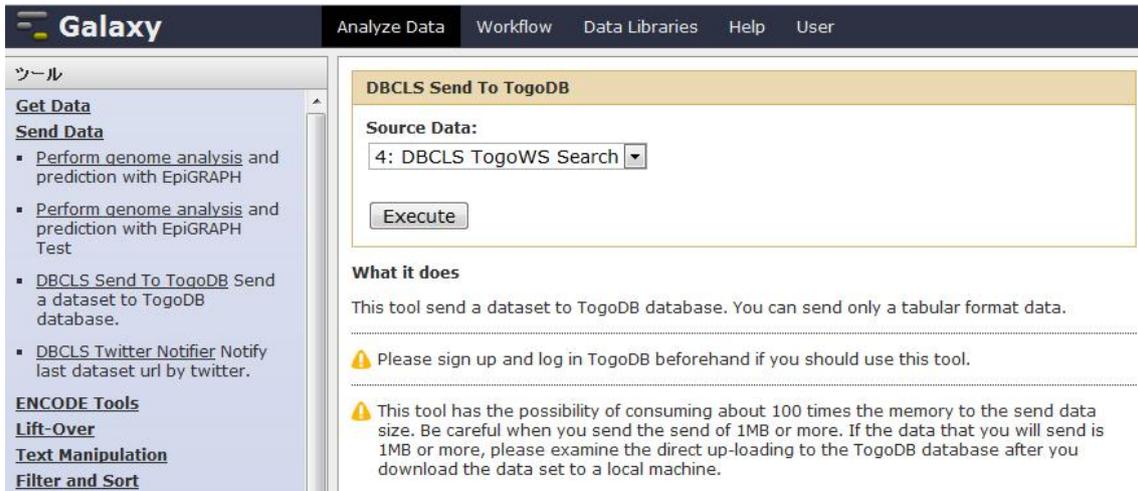


図 3.4 DBCLS Galaxy による解析結果を TogoDB へ出力するためのツール画面

(2) ワークフロー技術を用いた統合 DB 環境構築（産業技術総合研究所生命情報工学研究センター）

本年度は、タンパク質立体構造モデリングのワークフロー、ユーザが変更可能なアクティブ・ワークフロー、及び2種のSOAP対応サービスを開発した。

1) タンパク質立体構造モデリングワークフロー（Protein Structure Modelling Workflow）

本ワークフローは、立体構造未知のタンパク質の立体構造のモデル化を行う。モデル化の前処理として、BLAST/PSI-BLAST 及び fold recognition により元になる構造テンプレートを探し出し、そのテンプレートをモデル構築のプログラムに渡す。ユーザからアミノ酸配列を受取り、構造テンプレートを取得後、テンプレート毎に分散処理を行う。下記図 3.5、3.6 は本ワークフローの概念図及び結果画面である。平成 21 年 11 月末に構築が終了し一般公開した。

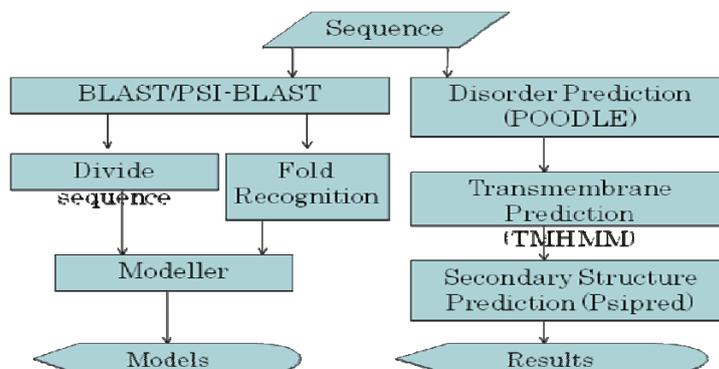


図 3.5 タンパク質立体構造モデリングワークフロー概念図



図 3.6 タンパク質立体構造モデリングワークフロー結果画面

2) ASIANアクティブ・ワークフロー (ASIAN Active Workflow)

本ワークフローは、CBRCで開発されたASIAN (Automatic System for Inferring A Network、遺伝子ネットワーク推定システム) をワークフロープラットフォーム (ワークフロー構築ソフトウェア) であるKNIMEを使用してワークフロー化したものである。ASIANは階層クラスタリングとグラフィカル・ガウシアン・モデリング (GGM) を組み合わせたネットワーク推定ツールであるが、これをいくつかのワークフローを構成する部品 (ノード) に分け、1つのワークフローとして構築した。実行形態には2種類あり、1つはThin-X、もう1つはLocal PCである。Thin-Xでは、全ての処理をCBRCのサーバ上で行い、画面表示のみユーザ自身のPCで行う。一方、Local PCは全ての処理をユーザ自身のPC上で行う。下記図3.7に実行イメージ及び結果表示画面を示す。平成22年

1 月末に構築が終了し一般公開した。

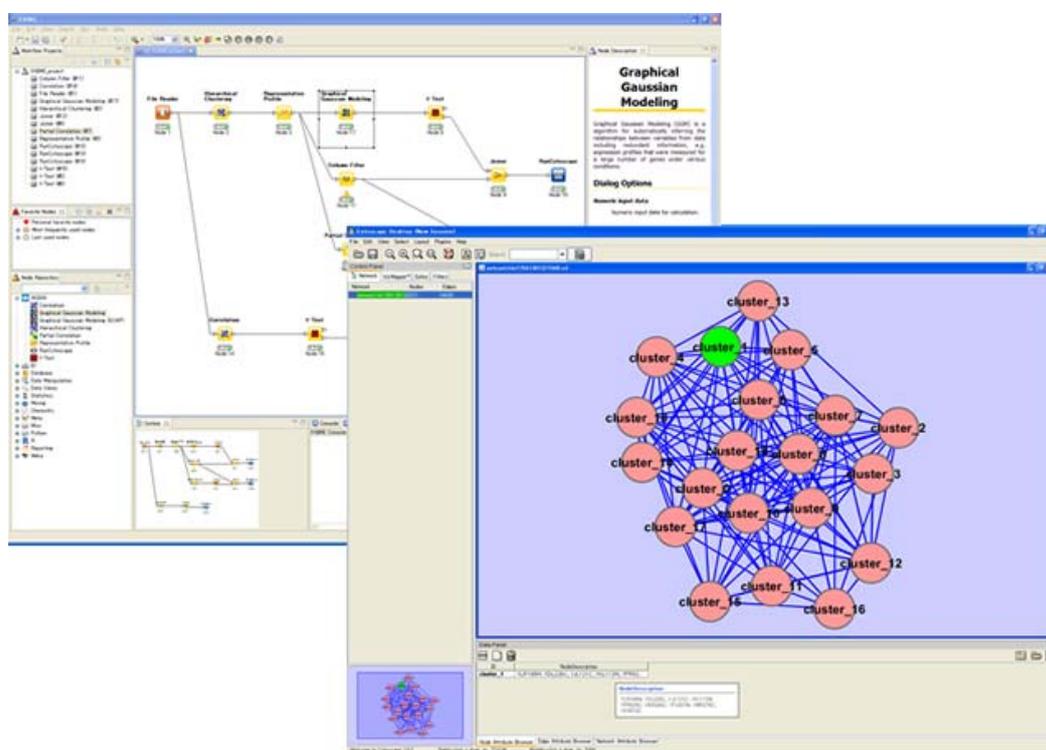


図 3.7 ASIAN ワークフロー 実行イメージ(左)、結果表示画面(右)

3) SOAP対応サービス (POODLE, WOLFPSORT)

本サービスは、CBRC で開発されたディスオーダー予測ツール POODLE (Prediction Of Order and Disorder by machine Learning) 及びタンパク質局在予測ツール WOLFPSORT (Protein Subcellular Localization Prediction) を SOAP 対応 (平成 19 年度に開発したプロトタイプを元に拡張・増強) したものである。従来ユーザは上記ツールを Web ブラウザによってのみ使用可能であったが、この SOAP 化により、プログラムによるアクセスが可能となった。平成 22 年 3 月末に構築が終了。公開は他のツールの SOAP 化と合わせ順次実施予定である。

(3) 専門用語辞書管理システムと専門用語解析技術の開発 (奈良先端科学技術大学院大学)

1) 専門用語獲得技術

① 専門用語辞書システムの開発

実施計画では、専門用語の内部構造記述、同義性、意味情報を管理する専門用語辞書システムの設計を計画していた。

医学生物学分野の専門用語の多くは複数の単語からなる複合語であり、それを構成する

単語間の関係は必ずしも隣同士の単語が直接関係するとは限らない、複雑な内部構造を持っている。複合語の内部構造は、それを構成する単語の間の係り受け関係として定義することができるが、係り受けの意味関係にいくつかの種類がある。それらを区別するために4種類の係り受け関係を定義し、内部構造のタグ付けを行うためのインターフェースをもつ辞書システムを設計した。医学生物学関係の用語では、特に、類似した2つ以上の語が並列関係として現れる場合に、一部の漢字を共有して縮退した形で表現されることがあり（例えば、動脈+静脈が「動静脈」になる）。そのような並列構造の記述法を昨年平成20年度考案した。また、平成20年度設計した辞書管理システムを用いて、約800語の専門用語（主に病名）に対して内部構造のタグ付け作業を行ったが、平成21年度は、これを約2倍の1600語にまで拡張した。

図 3.8 専門用語辞書管理システムの初期画面

図3.8に、複合専門用語の内部構造情報を管理し、表示するシステムの初期画面を示す。個々の詳細は省略するが、用語の検索はそれを構成する単語、または、単語間の関係を指定することによって検索することができる。図に示すように、単語には様々な情報を記述する項目が設定されており、どの項目を使用するかどうかは、管理者権限のある利用者が自由にカスタマイズ可能な仕様になっている。また、複数の辞書を同時に管理ことができ、複数の辞書に含まれる用語は、それらがもつ情報が矛盾しない限り、情報がマージされる。図3.9は、具体的な用語を検索した結果の用語詳細情報の表示画面である。用語が

もつ様々な情報が一覧されるとともに、内部構造をもつ用語については、図のように、その構造が木構造によって表示される。内部構造の解析が行われていない用語については、このシステムを用いて木構造のアノテーションを行う機能も実装されている。

表記が異なるが同じ概念を指す同義語について、それらを辞書システム内部で同じ識別子を与えることによって同義性を記述する機能も実装されている。この機能を用いて、指定された用語と同義の用語の一覧を表示することが可能である。

単語詳細

ID	2223089	
単語	遺伝子発現量	
読み	イデンシ/ハンゲンリョウ	
発音		
品詞	名詞一般	
活用型		
活用形		
BASE	遺伝子発現量	系列
ROOT		
辞書	WebLSD-200804*, pne_kw*, techterm*	
親概念日本語表記		
親概念英語表記		

構文詳細

状態	NEW
備考	
更新者	matsu
更新時間	2010-01-27 13:17:17

編集 削除

遺伝子発現量

構成	遺伝子, 発現量
枝の種類	D
縮退文字の位置	
省略文字の位置	none

ツリー構造

```

graph TD
    A[遺伝子発現量] --> B[遺伝子]
    A --> C[発現量]
    B --> D[遺伝]
    B --> E[子]
    C --> F[発現]
    C --> G[量]
  
```

図 3.9 用語の詳細情報の表示例

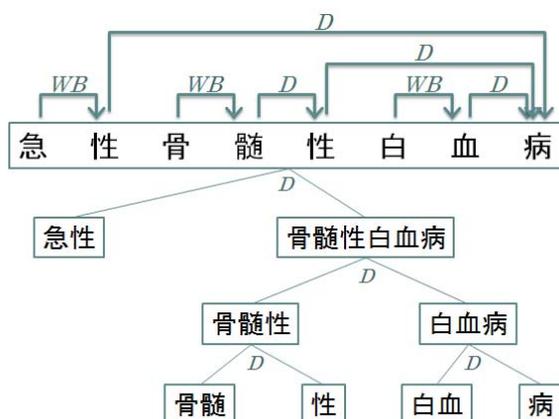


図 3.10 文字単位の係り受け解析と内部構造の対応

② 専門用語解析技術の開発

専門用語の内部構造情報のアノテーションを行ってその情報を格納し、表示する機能を上記の専門用語辞書システムに実装したが、すべての用語について手作業によるアノテーションを行うことは非効率である。平成20年度に内部構造アノテーションを行った用語を訓練データとして、機械学習による複合語内部構造自動解析プログラムを、文字間の係り受け解析として試作した。用語の内部構造解析を文字間の係り受け解析とみなすことによって、単語の縮退や単語内の並列構造など、複合語がもつ複雑な現象を考慮した内部構造解析が可能となる。図3.10は、「急性骨髄性白血病」という語の文字単位の係り受け構造を図

の上部に、単語に基づく木構造を下部に示している。上部の係り受け関係のうち、WBは単語を構成する先頭文字としての係り受け、Dは通常の依存関係を表している。他に、単語内部のつながり(WI)、並列関係(P)、逆向きの依存関係(R)などの文字係り受け関係を定義し、専門用語の内部構造の解析を、文字に基づく係り受け解析によって統一的な方法で実行できることを示した。

平成20年度は文字情報のみを用いた係り受け解析を行ったが、平成21年度は解析の過程で単語の情報（特にMeSHのシソーラスコードの情報）を用いることにより、85%以上の精度で専門用語の内部構造解析を可能にする手法を提案した。

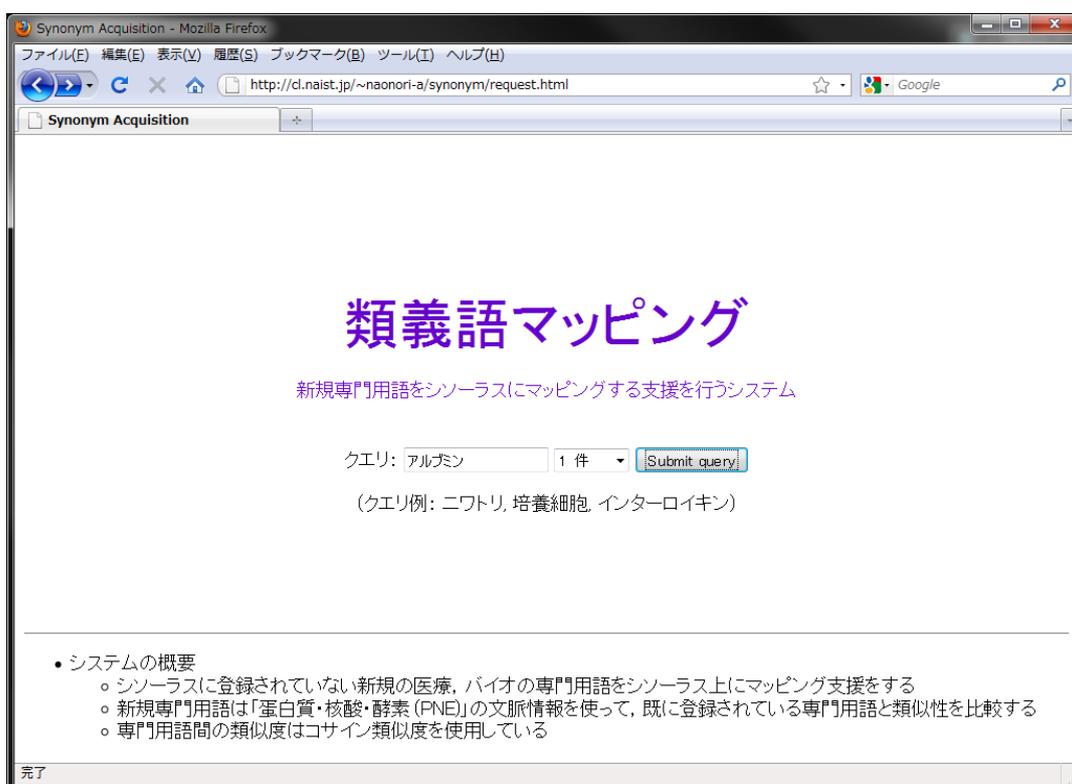


図 3.11 類義語検索ツールの初期画面

2) 専門用語抽出ツールの設計と開発

① 専門用語辞書拡張支援ツールの設計と開発

専門分野のテキストに出現する未知の専門用語を特定し、その意味クラス(MeSHのクラスラベル)の分類を行う自動分類手法を設計した。専門用語を含む文を係り受け解析し、対象とする用語の前後に出現する単語や対象用語に直接係る単語、また、それが直接係る係り先の単語、用語の末尾の文字列などを手がかりとして、意味クラスの推定を行う方法を設計し、実装した。未知の用語の意味クラス推定は、既知の専門用語との類似性に基づいて行うが、類似度尺度にリンク解析の手法を適用し、従来のベクトル空間法で用いられるコサイン尺度に比べて、コーパス中の出現頻度が低い語について提案手法の精度がよいこ

とを示した。この手法を用いて、新しく登録する専門用語に対し、既登録語の中から類似度が高い順に一定数の用語を提示し、かつ、それぞれの語のシソーラス上での位置を表示するツールを開発した。図3.11、3.12にその初期画面と検索語のシソーラス内での位置の表示画面を示す。

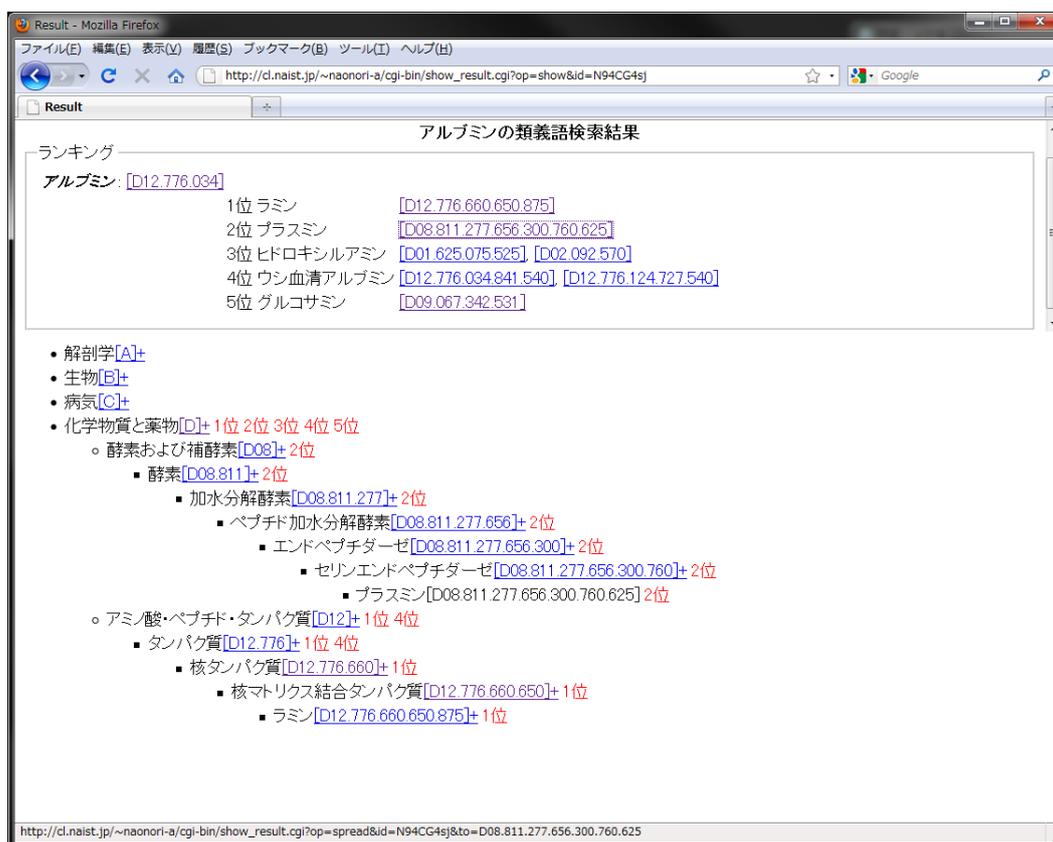


図3.12 検索語のシソーラス内の位置表示

3. 2 ヒト統合 DB の開発・運用（情報・システム研究機構、九州大学）

3. 2. 1 ヒト統合 DB の開発・運用の実施計画

(1) ヒト統合DBの開発・運用（情報・システム研究機構）

平成20年度までに、辞書シソーラス類の整備、キュレーションのための文献解読手法、ツールの開発、バンク目次システムの構築を行ってきた。特に平成20年度は、辞書のDB化とアナトモグラフィ・ボディパーツ3Dの高度化、論文正規化情報の取り出しとキャッシュ機能を持たすなどのWired-Markerの高度化、DNA DB総覧と遺伝子バンク（GEO）の融合化の検討などを行った。平成21年度は、引き続き医学研究や創薬開発研究に携わる利用者がゲノム情報、分子データベース、文献情報等の、研究開発に必要な情報を容易に統合利用できるように、背景知識の整理を行い、辞書やオントロジーの形で公開する。またこれらを用いて データベースの整理、統合を進める。具体的には、構築した辞

書やデータの管理システムへのコンテンツの充実、キュレーション情報の付加、DNA DB 総覧と遺伝子バンク (GEO) の融合化などを行う。

(2) 多型知識表現技術の開発 (九州大学)

1. JAGQC パイプラインの開発・運用

現在の大規模ジェノタイプデータを用いた疾患遺伝子のゲノムワイド関連解 (GWAS) では、データの信頼度をそのデータと共に客観的且つ検証可能な形態で明確に提示することが求められる。これまで、本研究では Affymetrix 社 SNP アレイ解析の生データ (.cel file) から出発し、実質的に国際基準である 2 種の解析ソフトウェア (GAINQC 及び PLINK) を用いて QC を行うパイプライン (JAGQC Pipeline) を作成し、且つその結果を web ページに提示するシステムを構築した。これにより GWAS 1 次スクリーニングデータの客観的/統一的手法による QC と、その結果の開示を行う一貫体制を確立した。当該年度の目標は、1. JAGQC Pipeline に EIGENSTRAT プログラム等の機能を取り入れ、その機能を強化する。2. 国内各研究グループからの GWAS 1 次スクリーニングデータを受け入れ、これを順次 JAGQC Pipeline を通すことにより QC を行い、これらの結果を WEB ページで開示する。

2. 日本人標準ハプロタイプデータベース構築作業の開始

我々はこれまで日本人胞状奇胎 (ヒトハプロイド細胞の集合体である) 由来 DNA 約 100 例を用いて日本人確定ハプロタイプ構造のゲノムワイドマップを構築し、D-HaploDB (<http://orca.gen.kyushu-u.ac.jp/>) として公開してきた。ここに「確定ハプロタイプ」とはハプロイドゲノムから直接決定された正確なハプロタイプのことである (HapMap が提供しているアジア人ハプロタイプ情報はディプロイドジェノタイプから推定されたものであり、数%の誤りがあることが知られている)。しかし D-HaploDB では、対象試料数の限界から低頻度多型に関する情報が不十分であった。そこで本研究では、最近公開されつつある標準日本人数百人のディプロイドジェノタイプデータを取り入れて、低頻度多型情報までを含むハプロタイプマップを構築してヒト統合化 DB の一部としての新たなデータベースとして公開するプロジェクトを平成 21 年度に開始する。このハプロタイプマップは、現在最高水準にあるとされる HapMap3 と比較しても 2 倍以上の試料数を用い、また胞状奇胎由来の誤りのないハプロタイプ情報をもとに決定されるものであることから、アジア人ゲノムに関する極めて精度の高い情報を提供することになる。またこのデータベースには LD-bin 情報、tagSNP 情報、コピー数多型 (CNV) 情報を含める予定であり、我が国の疾患関連解析の精度を飛躍的に高める情報基盤となると想定している。

3. 2. 2 ヒト統合 DB の開発・運用の実施内容

(1) ヒト統合 DB の開発・運用 (情報・システム研究機構)

「遺伝子発現バンク(GEO)目次システム」(以下、GEO 目次)は GEO データベースを独自の分類軸により分類整理し、全体を俯瞰可能としたデータベースを構築、Web 公開サービスとして利用者が目的とするデータに辿り着くまでのプロセスを書籍の目次からの検索に類似させたインターフェースを備えたシステムである。

なお、GEO とは NCBI が維持管理している遺伝子発現情報のデータベース、Gene expression omnibus の略称で、マイクロアレイ、タグシーケンス等の実験データを登録するデータベースである。

平成 21 年度は、公共データベースの更新に伴う保守および運用を行うとともに、GEO 目次のデータの質を向上させること、公開ウェブサイトにおけるデータ閲覧方法の改善を進めることを主眼とした開発を行った。

具体的には、データの質を向上させる作業として、次の分類ルールの見直しを行った。

- ・ 測定技術分類
- ・ 動物解剖学分類
- ・ 国別登録データ分類

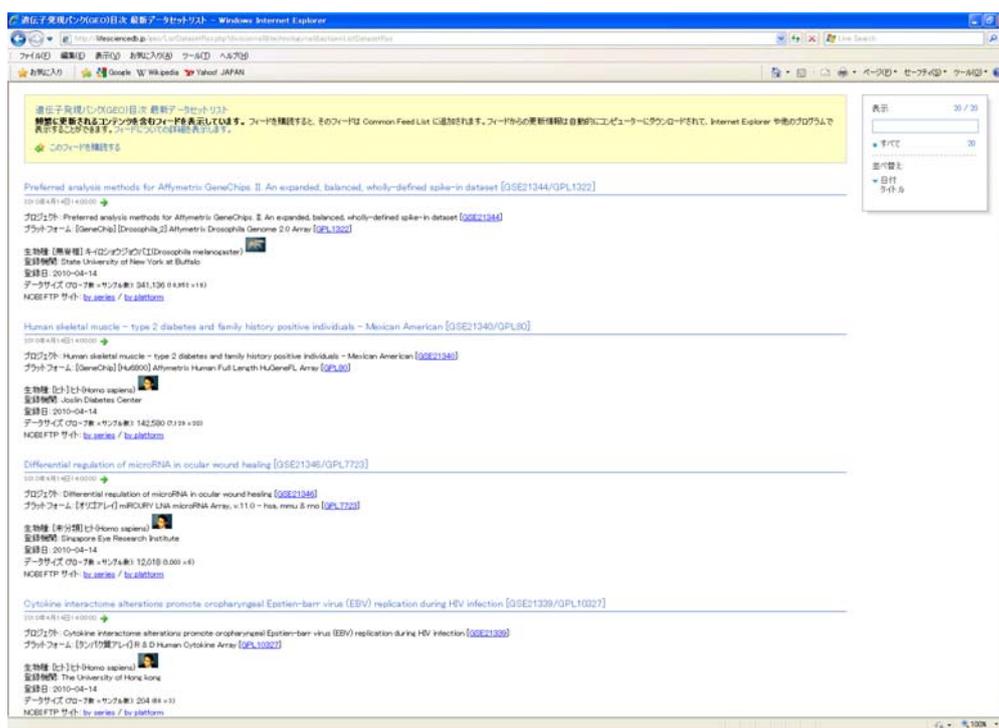


図 3.13 遺伝子発現バンク(GEO)目次最新データセットリストの例

これによって、各項目で未分類となっていたデータの数を減らした。また、臓器分類に平成 18 年度の成果である「植物解剖学分類タガー」(<http://lifesciencedb.jp/?pg=4>)を流用することにより、植物解剖学分類も表示するようにした。

次にデータ閲覧方法を改善する作業として、RSS フィードの配信サービスの実装やユーザインターフェースの改善を行った。

さらに、GEO 目次のデータ更新を週 1 回程度の間隔で行えるように、データ更新作業を自動化した。これにより、上で述べた RSS フィードの配信によって、利用者が NCBI GEO に新たに登録、更新されたデータセット、プラットフォーム、サンプルのリストを受け取ることが可能となった (図 3.13)。

次に、「DNA データベース総覧と検索システム」(以下、INSD 目次) は国際塩基配列データベース(INSD; International Nucleotide Sequence Database)を独自の分類軸により分類整理し、全体を俯瞰可能としたデータベースを構築、Web 公開サービスとして利用者が目的とするデータに辿り着くまでのプロセスを書籍の目次からの検索に類似させたインターフェースを備えたシステムである。平成 21 年度は、公共データベースの更新(DDBJ リリース 79 及び DDBJ リリース 80)に伴う保守および運用を行った。データ更新作業のうち、独自の分類軸による分類整理処理は完了している。しかし、一部の更新処理が計算機能力不足により実行できず、データ更新作業が完結できなかったため、公開には至っていない。新しい計算機の確保も諸事情により遅れたが、3月末に確保ができたため、平成 22 年度早期に公開できる見込みである。また、システムに関する文書が散在している状態であったが、これを整備することにより、運用手順書として纏めなおした。

統合データベースプロジェクトでは、日英の翻訳情報がそれぞれの出典においてどのように記載されているかを記述したメタ用語集や、基本概念での関連付けや整理を可能にするための用語シソーラ作成のおよび用語の構造化(いわゆるオントロジー)を進めている。平成 21 年度は前年度までに収集してきた辞書資源が充実してきていたので、それらをもとに「DBCLS メタ用語集」を作成し、メタ専門用語データの電子版配布

(<http://lifesciencedb.jp/lbdb.cgi?gg=dic>)を行った。その過程で、これまで曖昧なままだった辞書資源の探索・収集の対象範囲や保存形式など、一連の過程を明確にした。また、増大してゆく辞書資源を管理するデータベースを構築した。さらに、上記の説明や、

「DBCLS メタ用語集」に収集した情報源、作業を進める上での難点や苦勞を解説するサイト (<http://lifesciencedb.jp/bdls/>) を開設した (図 3.14)。平成 20 年度に引き続き、新たな情報源を選定し、辞書資源の拡充を図った。上期は動物名と植物名、下期はプランクトン、底生生物、原生動物、藻類、菌類、地衣類の生物名を中心に収集を行った。このようにして作成された「DBCLS メタ用語集」は以下のサイトで利用され、学名に和名を付与、和名による GBIF 検索などに利用されている。

- ・ Species Dictionary - 学名和名検索。

(http://bm2.genes.nig.ac.jp/Integrated_BodyMap/species.php)

- ・ 地球規模生物多様性情報機構 日本ポータルサイト 検索 (<http://gbif.ddbj.nig.ac.jp/>)。
- ・ DNA データベース総覧と検索における生物種名の和訳表示(<http://lifesciencedb.jp/ddbj/>)。

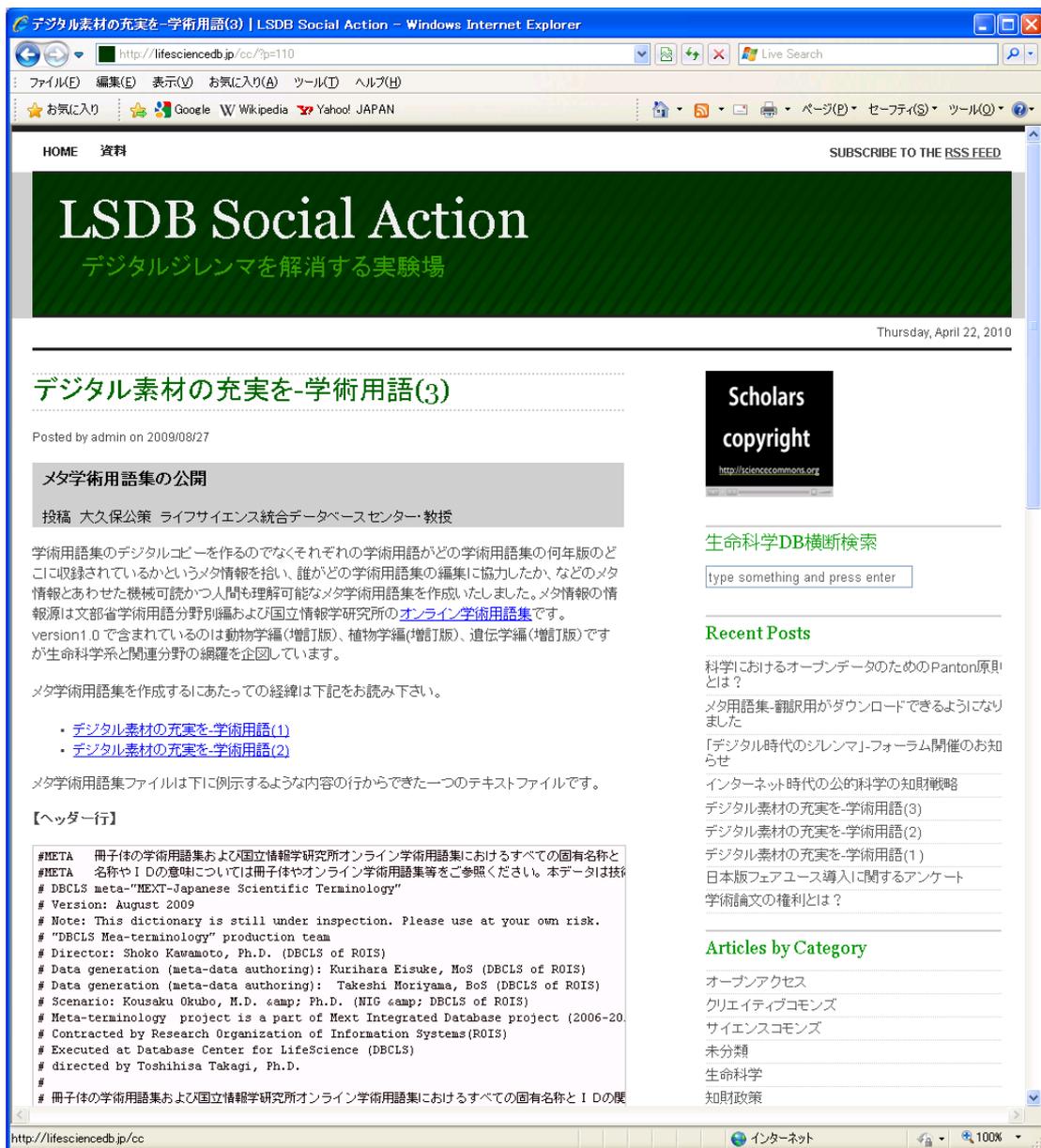


図 3.14 メタ学術用語集の公開についての表示 Web ページ

Wired-Marker はアノテーション促進の為に Fire-Fox の機能拡張ソフト(Plug-in)として開発が開始され、実用にたたる機能を持ち広く利用されるにいたったので平成 21 年度をもってその開発を終了した。

データベース統合事業では「データや情報の収集整理と共有化(再利用可能化)」を行う。これは以下の 5 過程の繰り返しである。

- ①オリジナルデータの選択と入手 (アノテーションでは関心情報を含む文献の検索と選別)。
- ②オリジナルデータの出自の記録と保全 (選択 文献のコピー、ファイリングと書誌情報の作成)。

③再利用者の為の十分なメタデータや解説の付与(文献の読解とアンダーライン・メモ付け
自分用の再利用化)。

④データやメタデータの利用制限や利用制約の解消 (③を使って書き換えやまとめを著作
する)。

⑤データ形式および記載用語などの表現の汎用化 (サマリーを出版公開する)。

アノテーションやキュレーションでは対象に関する説明や解説を専門文献中に求めて①ー
⑤の作業を行う。

近年はステップ①②は検索サイトや文献管理ソフトによってパソコン上で行えるものの、
プリントアウトを元に紙時代と同様の方法で③以下に進みサマリーが作られている。この
ままでは増加の一途であるデータにアノテーションを行うことは現実的に不可能である。

そこで、上の5つのステップの中で③の産物である「濃縮されたオリジナル情報集」を
データ形式上も制度上も共有再利用可能にする技術があればアノテーション過程に自由度
が増大し再利用性が高まり、濃縮結果に定型表現を与えれば濃縮とサマリーを分担したり
分散させたりが可能となる。現在ウェブ上のオリジナル情報を濃縮するソフトではウェブ
ページに対するマーキングソフト (scrapbook など) が代表的である。しかし、マーキン
グソフトは個人の PC 内での情報整理を目的に開発されているため、例外なくオリジナル
ページにマーク情報を付加して保存させてしまう。情報源が商業誌や教科書である場合、
この形式で作られたマーカー付きページは違法コピーであり当然公開共有することはでき
ない。そこでマーカーの位置情報をオリジナルページの情報と分離して保存し、共有再現
を可能にする仕組みを考案した (図 3.15)。

機能 1. 責任箇所 (引用 部分) への相対座標の自動付与

ウェブブラウザのページは URL で表現されるがページ内の座標を表現する方法に標
準的なものはない。Wired-Marker は X-path 形式でページ内の位置を表現し、解釈する
Hyper-Anchor 技術を利用している。アノテータも利用者も Wired-Marker を機能拡張に
加えることで Fire-Fox にカーソールで選択した箇所の情報を X-path で表現し、URL に#
区切りで付加する機能が与えられ、また同様の URL がブラウザに与えられるとそれを 2
分割してインターネット内の座標である URL とページ中の座標である X-path として解釈
し、その領域に移動し指定の修飾 (色づけする) などの機能が与えられる (図 3.15)。

機能 2. ブックマークの編集機能

ブックマークをたくさん集め、フォルダ整理を行わないとまったく再利用が不可能になる
が、フォルダ構造は収集終了時にしか最善のものが出現しない。Wired-Marker ではこれ
を助ける目的で自動と手動の 2 元的整理を採用した。平成 21 年度は追加機能として自由

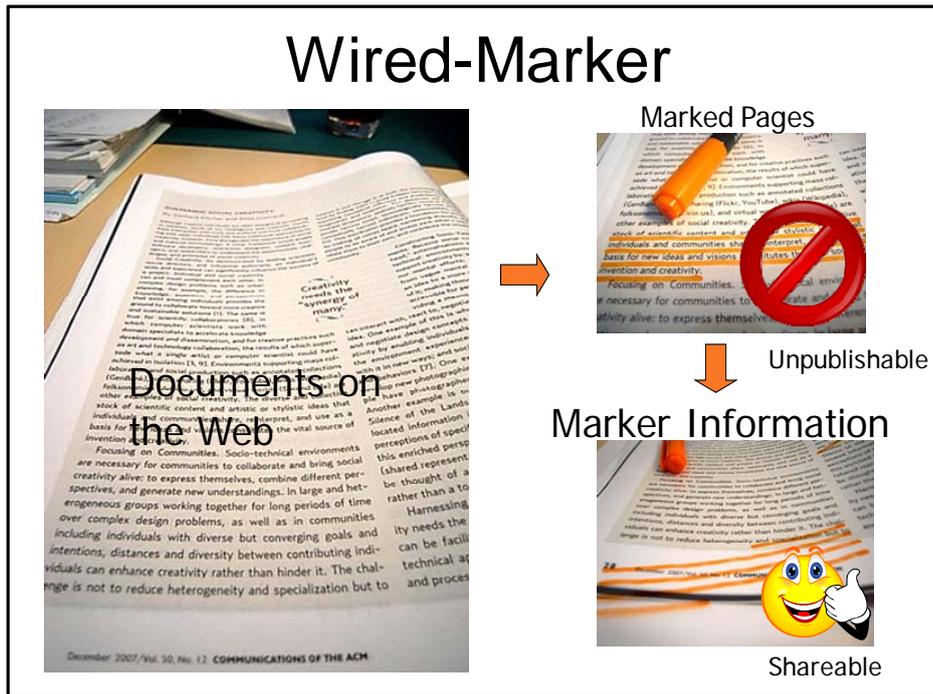


図 3.15 Wired-Marker の相対座標の自動付与機能

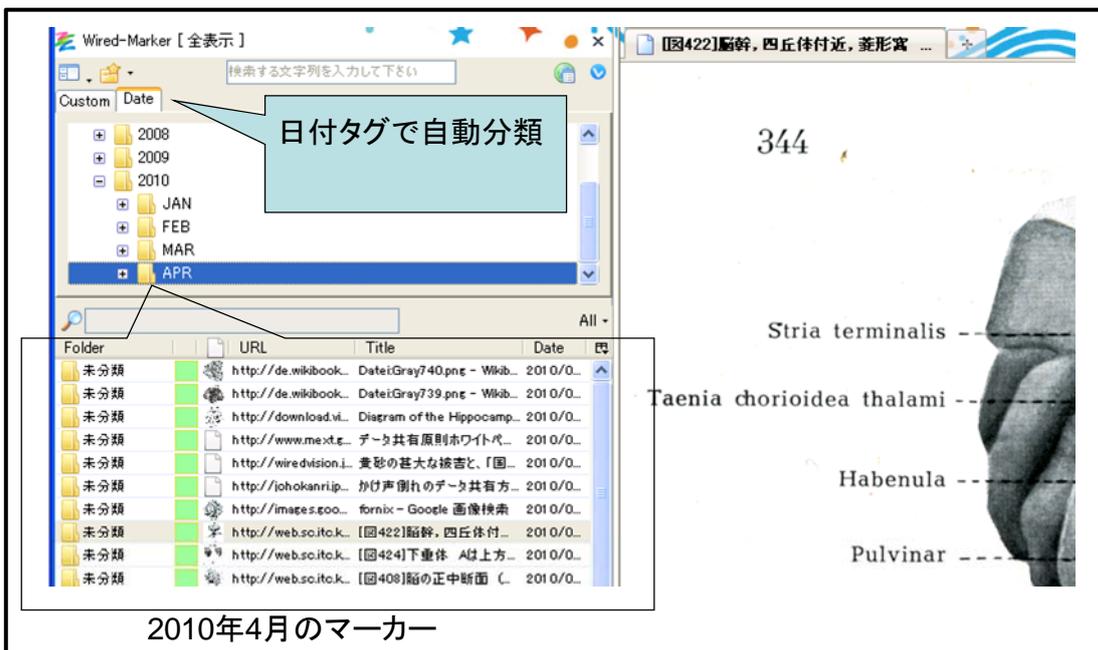


図 3.16 Wired-Marker の日付による自動分類機能

に付加階層化できる手動フォルダ以外に、自動的に年月日付の階層をもつ Date フォルダにマーカーを保存するようにした。これで未分類にマーカーを保存しても Date 分類から探し出すことができる (図 3.16)。

機能3. 共有、レポート機能

マーカー情報は通常 FF 内のサブウインドウに表示されているが、メインウインドウ内に表形式で表示することができる。また、マーカーが保持しているメタデータを自由に選んで表示することもできる。URL, Hyper-anchor adress, original-page の選択部分のキャッシュ、page title、ユーザのコメントなど再利用に必要な情報はすべて表示可能である。加えてすべての情報はこのウインドウから編集可能なので集める作業と名づける作業を分離することができる。名づけられ目的別に分類された hyper-anchor adress はそのまま Web ページから公開共有することができる。平成 21 年度は情報登録の際に多彩になったマーカー情報や Web ページの情報をどのように登録するかを設定可能とし、また、情報の入力を促すダイアログボックスを表示可能とする「登録内容編集機能」を設けた（図 3.17）。



図 3.17 Wired-Marker の登録内容編修機能

機能4. 特殊な情報源への対応

しばしば利用される情報源の中には特別なページ表現を行っておりひとつのURLとして特定のページが与えられない場合がある。代表的なものはPDF文献、オンライン教科書、PubMedCentralがある。平成21年度は(1)公共文献データベースであるPubMed Central、および、解剖学の教科書であるGray's Anatomyからの情報登録を可能とする

「ライフサイエンスデータ登録機能、(2)PDFやAjax + JavaScriptにより取得されるコンテンツを含むWebページからの情報収集を可能とする「PDFページ対応機能」を追加した。よく利用されるPubMedCentralではマーカー付与した文献についてはPubMed経由で自動的に書誌情報を収集しマーカーのメタデータに加えマーカータイトルに文献タイトルが与えられるなどの利便性が図られている。Gray's Anatomy(40版)は書籍を購入すれば出版社が提供する特殊な電子教科書サイトを利用することが可能である。このサイトに対応することでBodyParts作成グループや利用者の中で相互に関心領域を引用することを可能にした。ネット上のPDFファイルは表示法がまったく独自形式で一般公開されていない為にhtmlページのような座標分解能をだすことができない。マーカーを引くことで、URLとファイル中のページ番号を読み取り制約のない文書からは画像ファイルとして関心部分を収集可能にした。

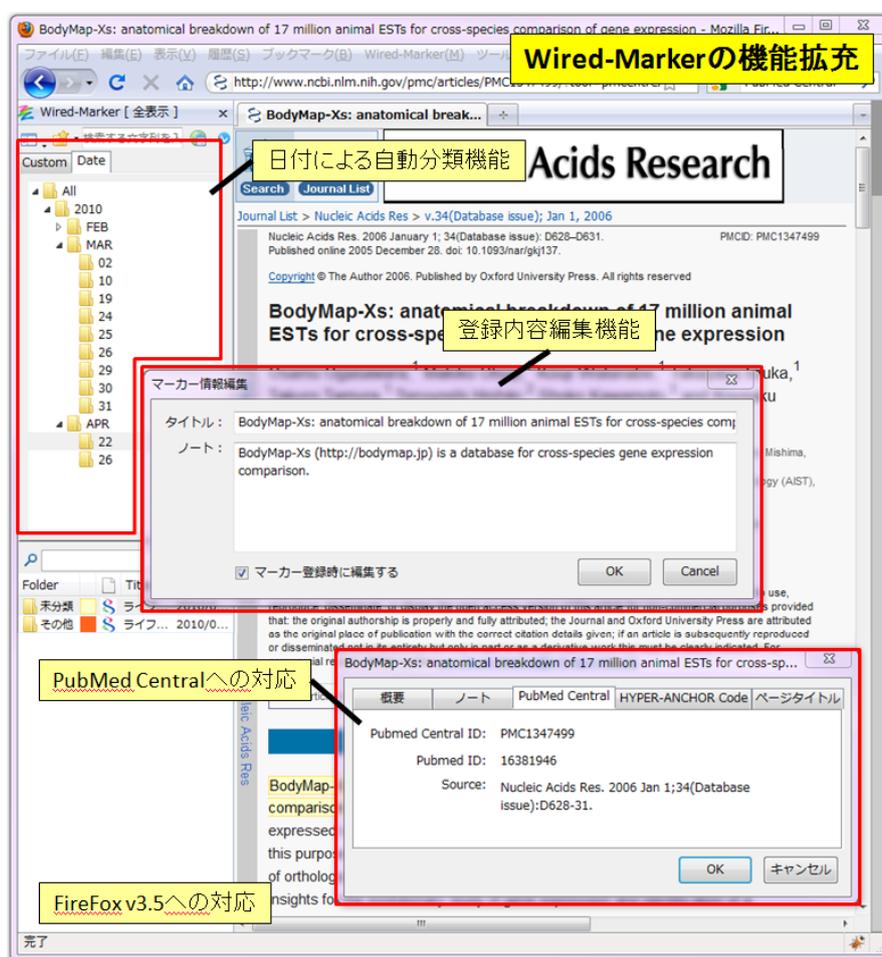


図3.18 Wired-Makerの機能拡充した主な項目

以上を纏めると、機能拡張した主な項目としては、「ライフサイエンスデータ登録機能」として、公共文献データベースであるPubMed Central、および、解剖学の教科書で

あるGray's Anatomy からの情報登録を可能とする機能を、「登録内容編集機能」として、情報登録の際にWebページの情報登録の内容を設定可能とする機能を、「登録情報一覧表示・編集機能」として、分類情報を含む登録情報を一覧表示して編集可能とする機能を、「日付による自動分類機能」として、登録情報を日付により分類・参照・利用可能とする機能を、「PDFページ対応機能」として、PDF及びAjax + JavaScript により取得されるコンテンツを含むWebページからの情報収集を可能とする機能を開発した。また、「Wired-Marker v3.5 用ユーザインターフェースの改善」として、全機能を対象にアップデートと整理を行い、ユーザインターフェースの刷新を行った（図3.18）。

BodyParts3D プロジェクトは 7000 あまりの解剖学用語で表現されている概念をすべて 3次元立体として数値で定義して公に提供するプロジェクトである。解剖学用語があらわす対象の形態を多面体データ(数千から数万の頂点座標列)として記述し BodyParts3D データベースから提供しています。すべてのデータは共通の 3次元座標系を使っており、すべての用語を集合させると一体の詳細な人体モデルを作るように設計されている。

BodyParts3D データはデータを管理提供するための①BodyParts3D サーバとブラウザ中や紙面に 3D イラストとして表現（レンダリング）するための②Anatomography サーバを通じて提供している。BodyParts3D サーバは立体構造データとして知識情報を扱う新しいカテゴリーのデータベースサーバである。

解剖学用語は医学概念中最大の概念集合であり、基礎医学や臨床医学でのコミュニケーションに欠かすことができない。NLM の索引用統制ボキャブラリである MeSH でも 5 系列中最大の系列である(医学生物概念は解剖・生物・疾患症状・化合物薬剤 D・人為に分類されている)。文献中の知識を機械的に扱う場合はもちろんのことデータベース統合で問題になる遺伝子発現や遺伝子多型などのデータの統合に際してもそのデータの由来（サンプル）を自然言語で記述したメタデータを機械的に操作可能にすることが最大の課題とされている。つまり医学用語の意味に照らして文献やデータを機械的に操作可能にすることはデジタル時代の医学研究や臨床における大きな挑戦課題である。

解剖概念はそもそも構造と形態を持った実態をクラス化した抽象概念である。したがって、形や位置関係などが主要な定義や特徴の一部となっている。オントロジーは概念関係を明示的に宣言する優れた方法であるが、十分な表現を行うためには特別な概念関係を設けなければならない。新たな関係の新設は著者に依存することや内部で矛盾することが多く、利用者にとって理解・利用困難なものになっている。すなわちオントロジーでは表現力と利用度がトレードオフの関係にある。多くの解剖オントロジーが個別の利用目的や利用ドメインで構築されて相互利用性がないのはこの為である。

一方 BodyParts では解剖概念を代表的なモデル人体（カノニカルモデル）中のセグメントとして表すものである。したがって、精緻に作れば作者によらず殆ど同じモデルになるので、より表現力のあるものはより理解しやすいものになる。ただし部分全体関係や上位下位関係などの算出も可能であるがグラフをたどるよりも計算に時間がかかる弱点がある。

以上の特徴から両者を併用することで表現力に富み、理解しやすい「形」で解剖学概念を表現することが可能になるはずである。

画像診断領域ではCTやMRIなどの断面像の濃度差で与えられる臓器輪郭の抽出課題の解決レベルが飽和し、臨床現場でも個人のCTやMRI画像の輪郭を3次元合成した3Dデータは容易に得られる。最近はその延長としてComputer Assisted Diagnosisと呼ばれる課題が注目されており、輪郭抽出 臓器同定から異常画像（器質的疾患）の検出や同定を目指した研究が盛んに行われている。診断以外の目的で解剖学的に精緻な3Dモデルの構築に関してはVisible Human Project(NIH 1990-)による特定の検体の詳細な固定断面カラー画像に由来するものが有名であるが、主にこれも輪郭抽出の課題としての研究が盛んに行われた。そのほか教育コミュニケーション用メディカルアートとしてのカノニカル画像集や3D人体モデル表示サービスもいくつか存在する。これらは教育目的や手術シミュレーションの背景として主に商業ドメインで応用が行われている。

一方、脳に限定すれば関心領域(ROI)の共通表現や操作の為に標準座標系を導入する考えは1950年代に始まる定位脳手術に源が求められる。この目的のモデルは標本脳に規準点とグリッドをはじめて持ち込んだタライラッハアトラスに始まる(J. Talairach and P. Tournoux, "Co-planar Stereotaxic Atlas of the Human Brain 1988)。この70才の女性の切片アトラスに基づく座標系は後にfMRIなどを使う機能画像診断分野でも広く使われるようになる。タライラッハアトラスは電子化共有されBrainMapなどのニューロイメージングデータの統合サイトでも座標系として利用されている。このような展開はBodyParts3Dプロジェクトの意図するところと共通である。最近では詳細なMRI画像を用いたHarvard Surgical Planning Laboratory (SPL) の分節脳MRI3Dデータとそのイメージ視覚化スイート"Slicer"(2008)などが多目的に利用されている。しかしながら、解剖学全体をカバーする精緻で分節数の大きなデータはパブリックドメインには存在せず、情報の整理統合共有目的には利用できない。

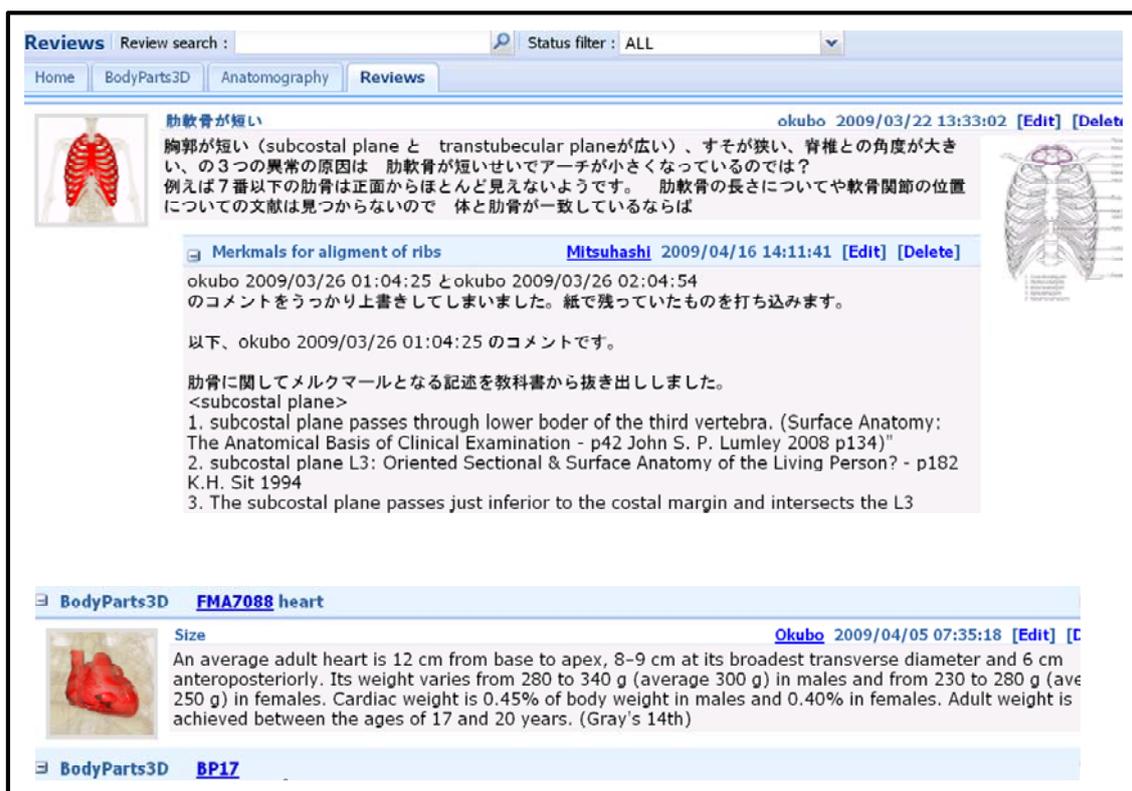
そこでこれらの前例から学び以下を行った。

- ①全身のカノニカルモデルを臓器単位で分節した簡単なデータをすぐに公開共有することで、転用を促進し広く利用されるように努める。
- ②データのカノニカルモデルとしての正当性を求め頻回のバージョンアップでデータ修正を続けて行く。
- ③解剖オントロジーや解剖学用語集の広く知られたものとのマップを常に行う。
- ④マスを持つ臓器だけでなく、点や線、面なども座標で定義することで簡単に区別できるようにする。などの特徴を持たすことで広く利用転用されることを第一目標とした。解剖学オントロジーとしては詳細でかつ共有可能なMejinoらのFoundation Model of Anatomy (FMA) を用い、名称集としては国際解剖学会が定めたNomina Anatomicaを用いている。

ユースケース：Anatomography(下)を通じての利用が主であるが、データを手術器具の開発やポリゴン編集環境の開発のテストデータとして利用されることを期待している。

平成 21 年度は主に②の方針に基づいて以下の構築と開発を行った。

①データ著作：すでに公開していた骨格系の座標修正や心臓の形態修正と詳細化を行い新規概念として四肢体幹のすべての骨格筋を追加した（図 3.19）。



The screenshot shows a web browser window with the URL 'http://www.bodyparts3d.com'. The page title is 'Reviews' and the status filter is set to 'ALL'. The main content area displays a review titled '肋軟骨が短い' (Short costal cartilage) by user 'okubo' on 2009/03/22. The review text discusses the relationship between the subcostal plane, transtubeular plane, and the angle of the ribs. Below the review, there is a reply by 'Mitsuhashi' on 2009/04/16, which provides a list of three anatomical points regarding the subcostal plane. The interface also shows other reviews, such as 'FMA7088 heart' and 'BP17', with their respective images and descriptions.

図 3.19 BodyParts3D のデータ正当化の為のレビューウインドとメルクマールの例

②BodyParts3D サーバの機能の拡張

- ・「BodyParts バージョン管理機能」

頻回におよぶデータの修正に対応する為修正前後のデータを共存させても不整合が起こらないようにデータ管理とデータと概念の対応管理を分離した（図 3.20）。

- ・「BodyParts 構造関係情報管理機能」

データと独立して FMA オントロジーの is-a 関係を保持し、一定割合異常の下位の要素が定義された場合には自動的に上位の概念に下位の和のデータを与える機能を追加した（図 3.21）。

- ・「BodyParts 構造関係検証機能」

オントロジーとデータ座標を比較して物理関係と論理関係の矛盾を自動検出する機能を開発した（図 2.21）。

- ・「切断オブジェクト管理機能」

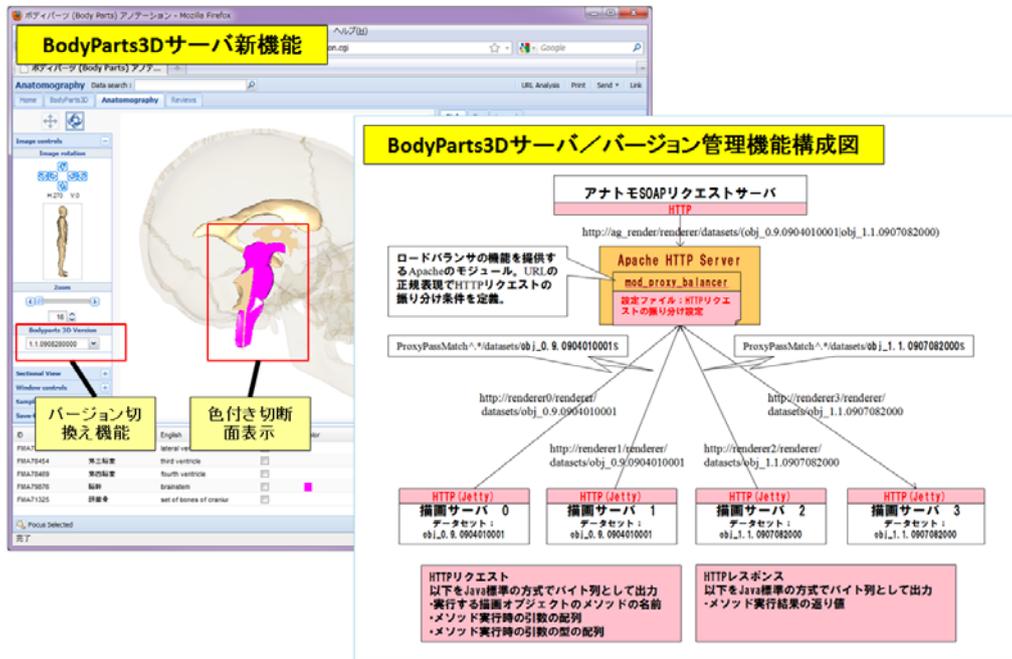


図 3.20 BodyParts3D サーバの機能の拡張

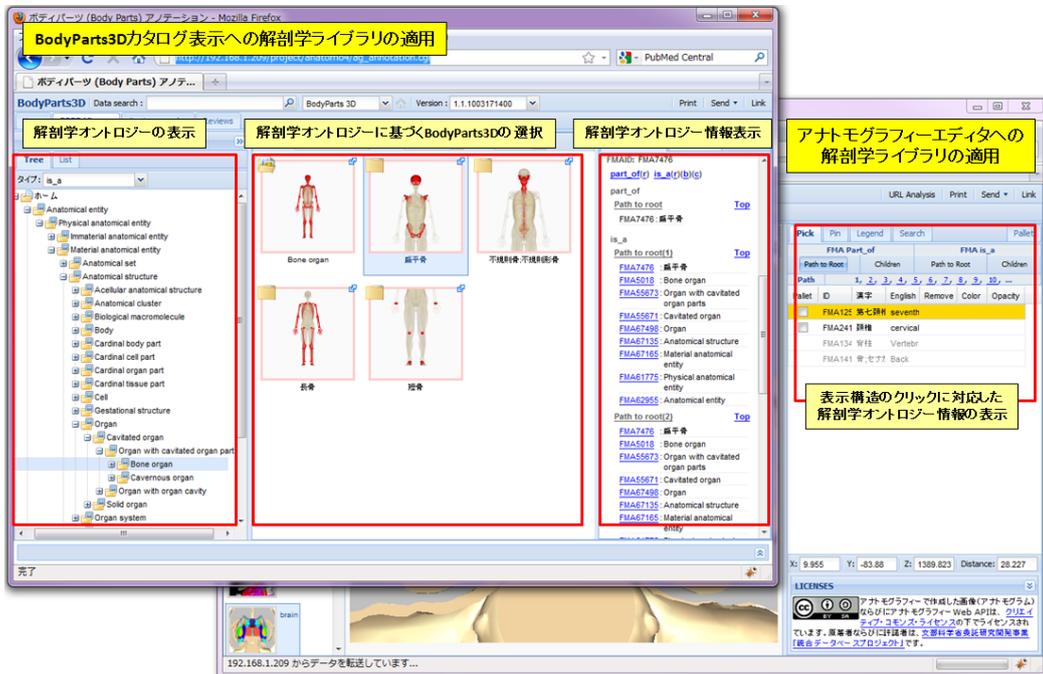


図 3.21 BodyParts 構造関係情報管理及び検証機能

Anatomography

Anatomography は BodyParts3D のアプリケーションのひとつで BodyParts3D データを呼び出して 3D イラストを描画するサーバである。医学研究者や臨床家が行うデータやアイデアの私信や出版におけるビジュアライゼーションに用いられる医学イラストレーションに提供し、すべてのイラストレーションやデータビジュアライゼーションを標準化してしまうための仕掛けである。

医学コミュニケーションでは部位名称が頻繁に利用されている。解剖学は医学の最大の基礎知識であり続けている。しかしながら、言葉を使ったコミュニケーションでは専門家にとっても俯瞰性や正確性が十分ではない。これを補う目的で医学では人体表面や人体内部の構造のイラストを多用しているが、イラストには二つの問題がある。

①著作であるために当然著作権が生まれ、転用は困難である。Venter の EST 特集はミケランジェロの BodyMap には阿形像が用いられたのは偶然ではない。

②キャプション以外に検索のための索引をつけることができない。

この為にはほとんど再利用されることなく書き捨てに等しい状態が続いている。

身体の部位に関する情報を表現するメディカルイラストレーションの流動性と再利用性、合算統合可能性を増すことができれば医学情報の整理統合への貢献は計り知れないものがある。

Anatomography は BodyParts データを呼び出してユーザの指示によって著作権のないイラスト画像(レンダリング画像) を作るためのレンダリングサーバである。この目的を果たす為には、与えるためには十分な表現力のあるイラストを誰でも簡単に書け、しかも機械可読性や集積編集可能性も持ちあわせなければならない。

ユースケース Anatomography サーバを BodyParts3D データのビューアーとして解剖知識のないエンドユーザ向けの解剖学知識の出口と同時に医学知識や解剖知識上に独自に表現したいデータやアイデアを持った臨床家や研究者の表現道具として提供し、彼らの表現に標準座標を与えることで医学データやアイデアの統合操作性を実現することを計画している。短期的には遺伝子発現データのビジュアル化、Wikipedia などのパブリック辞書でのイラスト利用、勤務医開業医の検査集計、医療統計の表現などへの利用を期待している。そこで以下を行った。

①レンダリングサーバとユーザ画面のやりとりを「臓器リストと描画パラメータでできた簡易言語 AnatomScript」で行った。

②イラストをイメージデータではなく AnatomScript として保存、共有、加算、整理、検索可能にする。

③レンダリングサーバは API を持ち URL に含まれる AnatomScript を検出実行することで常に新しい描画を返すようにする。

また、以下のことが必要であると考えて開発を進め、これまでも適宜臨床家によるレビ

ユーをいただいている。

- ①データの解剖学的正当性と詳細性。
- ②多忙でオブリゲーションのない一般医学者が直感的にストレスなく利用できる。
- ③レンダリングイメージに十分な表現力があり、その内容が機械可読で、共有や集積利用、再編集が可能である。

ボディパーツ3Dの3次元形状作成に関しては、筋肉、外皮の新規作成と胸部の詳細化を中心に行った。モデリングツールは引き続き3次元CADであるFreeFormを用いた。筋肉は主に骨格上の起始と停止の位置が構造上の目印（メルクマール）となっており、作成済みの骨格モデル上に起始、停止位置を定義し、筋肉同士の重なり具合を考慮しながら、筋肉モデルを作成した。胸部については、標準的な人体モデルを目指して、平成20年度までに作成したモデルに対して、解剖学教科書（主にグレイ解剖学やプロメテウス解剖学アトラス）の記述との差異をチェックし、それにできる限り準拠するようにモデルの精密化を行った。平成20年度までに作成したモデルは、教育用局所模型（心臓、気管、肺等）を3次元スキャナーで電子化したデータに修正を加えて作製してきたが、特に断面図上で臓器間の隙間があるなど、辻褄が合わない部分が出てきた。そのため平成21年度は局所CTデータをセグメンテーションし3次元再構成した粗データを下敷きに、もしくはスクラッチから、定性的、定量的なメルクマール記述や図譜に準拠するように部分的に再作製した。またメルクマール記述に対応するアナトモグラフィーを作成して、視覚的にモデルの確認ができるように試みた。このデータは公開データ利用時にユーザがモデルの完成度を知ることができるだけでなく、文章で表現された解剖学構造を視覚的に表現する知識ベースとなりうると思われるので、将来的には公開も視野に入れている。

モデル数については以下の表3.1の通りである。骨格系、筋肉系については、左右と両方を表す名称（例：左大胸筋、右大胸筋、大胸筋）の3つをカウントしている。循環器系や呼吸器系については、胸部再作製の際、部位の境界が確認できなかったものについて境界分けを行わなかったものがあり、モデル数が減少しているが、より大きな粒度の構造物として昨年度並みもしくはそれ以上の詳細度で作製した。新モデルのポリゴンメッシュデータを生命科学系データベースアーカイブ（<http://dbarchive.lifesciencedb.jp/bodyparts3d/dl.html>）から近日中に公開する予定である。

一方、発現データについてはヒトとマウスの様々な器官・組織の遺伝子発現データを統合して可視化し、ウェブブラウザで閲覧、一般公開し得るシステムの開発を行った（図3.22）。また、次世代シーケンサより出力された生データをEST発現頻度のデータとして可視化し、同様にウェブブラウザで閲覧、一般公開し得るシステムの開発を行った（図3.23）。

表 3.1 3次元臓器モデル数の比較

器官系名	H22年3月	H21年3月
神経系	54	55

感覚器系	4	18
循環器系	29	144
呼吸器系	11	45
消化器系	47	32
内分泌系	5	15
リンパ系	5	2
泌尿器系	9	9
生殖系	18	18
骨格系	272	272
関節系	7	0
筋肉系	749	2
外皮系	6	3
その他	3	27
合計	1,219	642

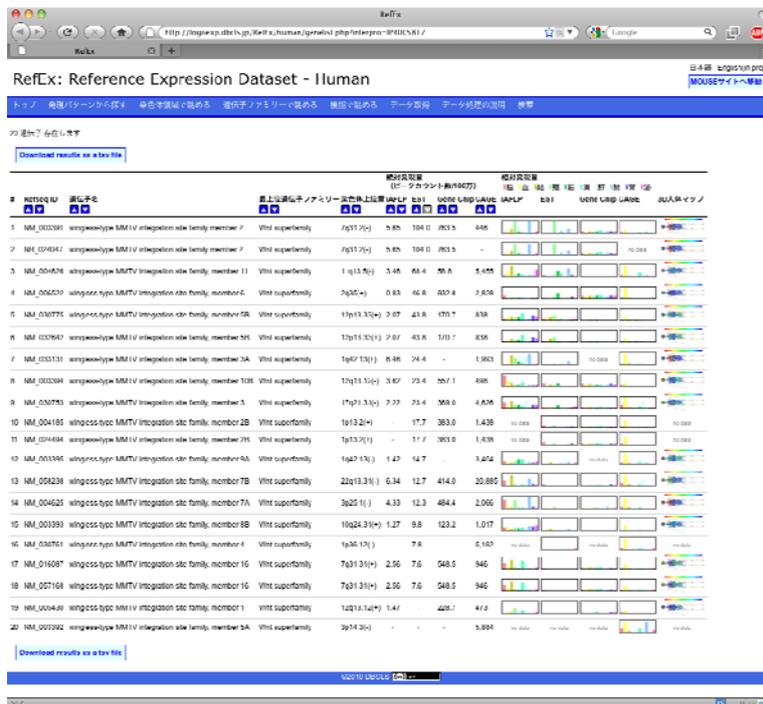


図 3.22 遺伝子発現データの統合可視化システムの Web 画面

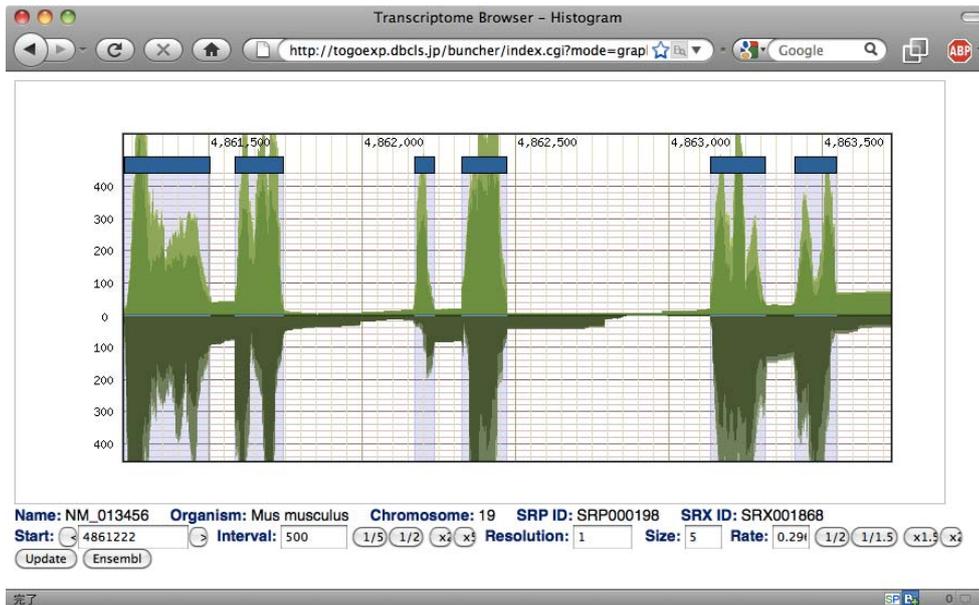


図 3.23 EST ゲノムマッピング表示アプリケーションの Web 画面

(2) 多型知識表現技術の開発 (九州大学)

1. JAGQCパイプラインの開発・運用

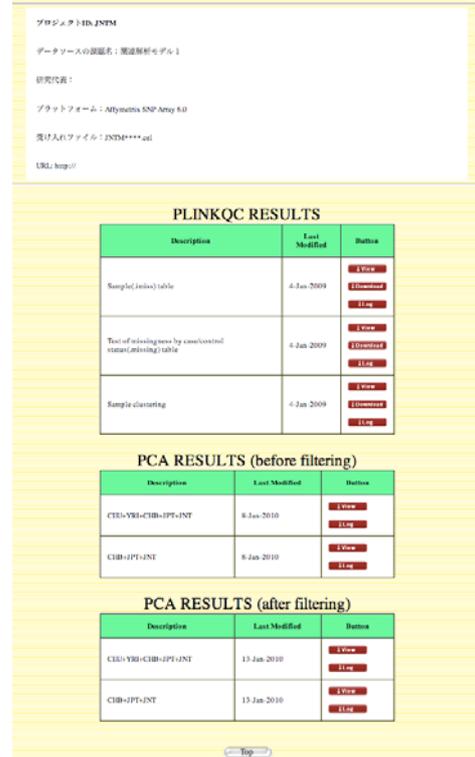
平成20年度に開発したJAGQCパイプラインは、1)ゲノムワイド疾患関連解析研究で得られたAffymetrix社DNAアレイによるジェノタイプデータのタイピングエラー率等に関するクオリティチェック (quality check、QC) をGAINQCプログラムによって行う一次QCと、2)それに続くケース集団・コントロール集団間での質の偏りの検定、両集団での遺伝的背景の均一性の検討をPLINKプログラムによって行う二次QC、の2段階からなり、これらの結果をJAGQCデータベースとして限定開示して来た。平成21年度は同パイプラインを改良して、全ての初期QCを一貫してPLINKプログラムで行うこととした。これにより、Illumina社製のDNAアレイを用いた関連解析データのQCにも対応した。さらに集団の均一性をより精細に解析するために、Eigenstratプログラムを導入して主成分解析を行うこととした。この結果、一次QCパイプラインでは、受け入れたジェノタイプデータに対してPLINKプログラムを用いて、ジェノタイプ失敗率の偏りの検定 (Test of missingness by case/control status)、集団の均一性の検討 (multidimensional scaling) を行った結果を出力する。また、二次QCパイプラインでは、受け入れたジェノタイプデータの中でcall rate 99%以上のSNPを用いて主成分分析 (principal component analysis: PCA) を行い、各試料の第一主成分および第二主成分に関する2次元展開図およびこの解析時のlog fileを出力する。

更に、解析結果を集約したJAGQCデータベースの形体を変更し、上記解析結果の表及び図による表示、解析過程のログファイルのページ表示、及び全結果のデータファイル及びログファイルのダウンロード可能な形での開示を行うこととした (図3.24)。また、日本人200

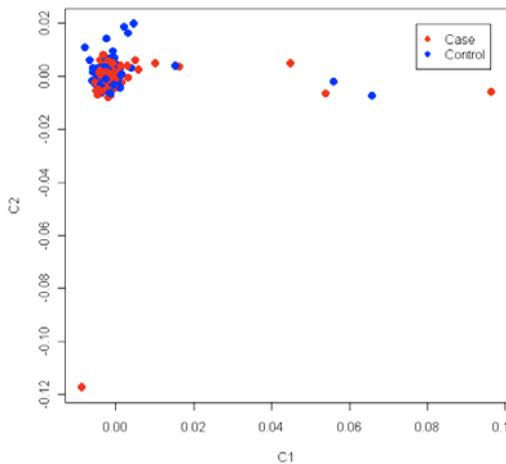
人標準ジェノタイプデータを用いて、上記全ての解析を行いそのデータを公開した。
 (http://crane.gen.kyushu-u.ac.jp/JAGQC/JAGQC_top.html)。



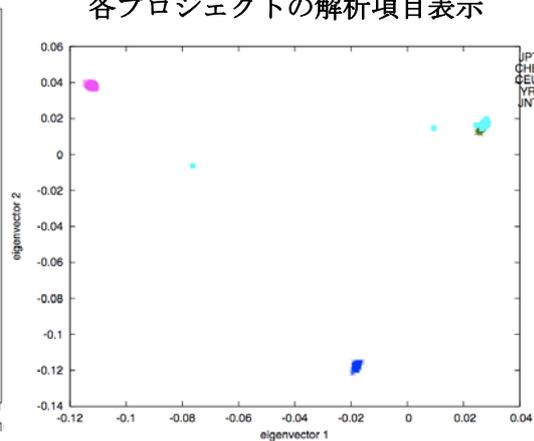
JAGQCのホームページ



各プロジェクトの解析項目表示



MDS解析結果の図示



主成分分析結果の図示

図3.24 JAGQCdbの画面表示

2. 日本人標準ハプロタイプデータベース構築作業の開始

我々が保有している日本人確定ハプロタイプデータベースにある 1.7×10^6 個の SNP マーカーに関する 85 ゲノムの確定ハプロタイプデータ、及びこれらのマーカー間での連鎖不平衡の指標である r^2 データ、の 2 個を統合 DB の他機関へ移行するためのファイル形式の検討

と整備を行った。これらのデータはAffymetrix社製のSNP Array 6.0及びIllumina社製の1M Duoアレイの両者を用いてハプロイドゲノムを持つ胞状奇胎DNAを解析して決定した確定（誤りの無い）ハプロタイプに関する情報であり、Hapmapプロジェクトが発表しているアジア人のハプロタイプに関する情報（これまで世界標準とされているが、一定頻度の誤りがあることが示されている）より高精度である。今後多数の日本人標準ジェノタイプデータと統合することにより、低頻度SNPを含む日本人標準ハプロタイプデータベースを完成させる予定である。これを日本人ハプロタイプ構造の基盤情報として用いる事により日本人での疾患関連解析の信頼度が向上する。

3. 3. モデル生物・産業応用生物統合 DB の開発・運用

（情報・システム研究機構、かずさ DNA 研究所）

3. 3. 1 モデル生物・産業応用生物統合 DB の開発・運用の実施計画

（1）（情報・システム研究機構）

主として微生物を対象とする基礎生物学研究および生物の産業応用研究に携わる利用者がゲノム情報、分子データベース、文献情報等の、研究開発に必要な情報を容易に統合利用可能とする環境を整備する。平成20年度までに、微生物情報資源のサイトの情報を収集しデータベースカタログに登録して統合データベースのポータルサイトから利用可能とした。また、平成20年度から我が国の微生物産業と新世代シーケンサの普及を想定して、これからの微生物ゲノム解析の標準を提供すべく微生物ゲノム解析パイプラインの開発に着手した。平成21年度以後、共通基盤技術の課題で開発される遺伝子名と蛋白質とPNEを紐づける技術の採用も検討しつつ、ゲノム配列に対して自動的にアノテーションを付与するパイプラインをゲノム配列アノテーションの標準足りうるように拡充する。

（2）（かずさDNA研究所）

平成20年度までに構築し改善を進めた植物および植物関連微生物のゲノムデータベースをさらに改善し、利便性の向上を目標とした業務を行う。ユーザからの情報集積とそれを基にした整理を可能とすることを特徴とするゲノムデータベース環境を活用しさらに改善を加えつつ、データの集積によるデータベース統合化をより多くの生物種について進める。

1) 高度情報集積データベースの運用と改良

平成20年度までに構築・改良した、分子データ上のポジションを統合のための基盤情報とし、アノテーション・キュレーションの統合と高度化を可能とするデータベース(KazusaAnnotation: <http://a.kazusa.or.jp/>)を植物および植物関連微生物のゲノム解析情報を対象として運用しつつ、ユーザの要望を汲みあげて改善を加え、さらなる利便性の向上を行う。同時に、集積したアノテーション情報から生物学的な意味を取り出すた

めの情報表現技術の改善を行い、それを閲覧するためのビューワ (genoDivePro, genoDiveEu)の改善と提供を行う。また、データベース中の遺伝子名や遺伝子IDの名称の食い違いを解決するサービスを維持しつつ、充実とインターフェースの向上を図る。

2) ゲノムアノテーション情報の蓄積と高度化

1) のデータベース環境を改善、運用しながら、その環境を利用したアノテーション情報の蓄積と高度化を引き続き進める。現在までの成果として、遺伝子、タンパク質に関する論文記載情報を光合成細菌であるシアノバクテリア *Synechocystis* sp. PCC 6803、*Anabaena* sp. PCC7120、*Synechococcus elongatus* PCC7942に関する合計約2100報分の論文について登録した。また、植物と共生し窒素固定を行う根粒菌、*Mesorhizobium loti* MAFF303099、*Sinorhizobium meliloti* strain 1021に関する合計約300報の論文について、登録した。さらに高等植物であるマメ科のモデル研究植物であるミヤコグサに関する論文についても同様に登録を開始した。

平成21年度は、かずさDNA研究所でゲノム塩基配列が決定されたモデル生物を中心に、*Gloeobacter violaceus* PCC 7421、*Microcystis aeruginosa* NIES-843等、より広範な生物種の遺伝子情報について、引き続き論文記載情報および実験情報の蓄積業務を実施し、ゲノムの位置情報と論文記載情報の統合を図る。登録した情報は、今後も KazusaAnnotation システムなどからすみやかに完全公開していく。あわせて、効率的なアノテーションを行うためのアノテーション方法の開発や、アノテータ・キュレータの育成、技術向上のためにIT技術を利用したミーティング手法などの開発を行う。

3. 3. 2 モデル生物・産業応用生物統合 DB の開発・運用の実施内容

(1) モデル生物・産業応用生物統合 DB の開発・運用 (情報・システム研究機構)

平成 19 年度に調査した微生物ゲノムアノテーションの国内外の状況と急速に普及し始めたいわゆる新世代シーケンサがもたらす研究動向の見通しのもとに、基礎生物学研究および生物の産業応用研究に携わる利用者がゲノム情報、分子データベース、文献情報等の研究開発に必要な情報を容易に統合利用して、データベースの整理、統合を進めることができるような環境構築の第一歩として、微生物ゲノムアノテーションパイプラインを構築して広く一般に公開した。これは、信頼されかつ簡易に利用可能なパイプラインを提供することによって、一定の質が担保されたアノテーションの普及を目指す試みである。

シーケンサの技術開発の進展とともに、大規模なシーケンシングセンターでなくても多くの研究室やグループにおける配列決定が広がっていくと見込まれる。したがって、大規模シーケンシングセンターにおいても専門知識と時間を要していたアノテーションの効率化を進めない限り、塩基の文字列がディスクに蓄積されていくに過ぎない。そこで本課題において、効率の良い微生物ゲノム配列アノテーションを実現する情報環境を目指した。

具体的には、微生物ゲノム解析において定評あるデータベースと定評あるアルゴリズムを組み合わせたパイプライン Microbial Genome Annotation Pipeline (以下、MiGAP)

を設計した。MiGAP によって、誰でも、何時でも、どこからでも、微生物ゲノム配列の生物学的意味を入手できるようになる。

MiGAP には、デフォルト解析の初級用のブロンズ (b-MiGAP)、パラメータを設定可能とする中級用シルバー (s-MiGAP)、さらにデータベースやアルゴリズムを付加することができる上級用のゴールド (g-MiGAP) の3つのレベルを設定することにした。今年度は、6月に b-MiGAP を一般公開し、s-MiGAP もゲノム解析の経験がある利用者に試験的に公開した。6月から3月までの利用状況を表 3.2 にまとめた。

MiGAP の登録利用者数は 83 名であり月平均 9 名余りに活用されている。利用の具体的内容として表 3.2 に投入件数、投入コンティグ数、投入総塩基数 (単位 Mbp)、出力 CDS 数、出力 RBS 数、出力 rRNA 数ならびに出力 tRNA 数を示す。これらの利用状況は、MiGAP の開発に着手した際の現状認識と予測が正しかったことを示唆している。シーケンサがますます高性能になり、また、低価格のモデルも出始めていることから、平成 22 年度以後 MiGAP の需要がさらに高まると思われる。12 名の利用者から得たアンケート結果でも、配列を投入するだけで一定のアノテーションを提供する MiGAP に対する評価が高く、また、より一層の高速化や、バクテリアはアーケアの配列に加えて菌類の配列への拡充が期待されている。一方で、2009 中は、DBCLS の MiGAP サーバが限界まで稼働していた。そこで、2010 年 1 月から 3 月の間に、MiGAP を国立遺伝学研究所のスーパーコンピュータに移植して、複数の配列解析を同時実行できる環境を整えた。

表 3.2 MiGAP の利用状

項目	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	Total
投入件数	24	20	6	10	13	33	61	42	29	13	251
投入コンティグ数	545	4,503	1	10	250	2,709	7,058	589	15,528	7,009	38,202
投入総塩基数 (Mbp)	28	35	3	2	12	50	56	25	39	59	250
出力 CDS 数	24,826	34,281	2,697	1,643	11,786	42,948	54,526	22,926	49,414	61,735	245,047
出力 RBS 数	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	21,028	30,222	39,887	51,250
出力 rRNA 数	86	82	6	4	70	50	384	92	98	74	946
出力 tRNA 数	476	556	44	34	204	499	942	465	431	486	4,137

(2) 植物及び植物関連微生物のゲノム情報データベース統合と高度化

(かずさ DNA 研究所)

平成 20 年度までに構築した、ユーザからの情報集積とそれを基にした整理を可能とすることを特徴とするゲノムデータベース環境を活用し改善を加え、データの集積によるデ

データベース統合化を複数の生物種について進めた。

1) 高度情報集積データベースの運用と改良

平成20年度までに構築した、分子データ上のポジションを統合基盤情報とし、アノテーション・キュレーションの統合と高度化を可能とするデータベース **KazusaAnnotation** (<http://a.kazusa.or.jp/> 図3.25参照) を、植物および植物関連微生物のゲノム解析情報を対象として運用しつつ、ユーザの要望を汲みあげて改善を加え、利便性の向上を行った。具体的には、データベースシステムの構造見直し、キャッシュ機能を追加することにより表示速度を数倍～100程度高速化した。これによりストレスの少ないデータ入力環境を実現した。また、遺伝子ごとに文献のどの部分に言及されているかを直感的に理解出来る表示などを作成した。

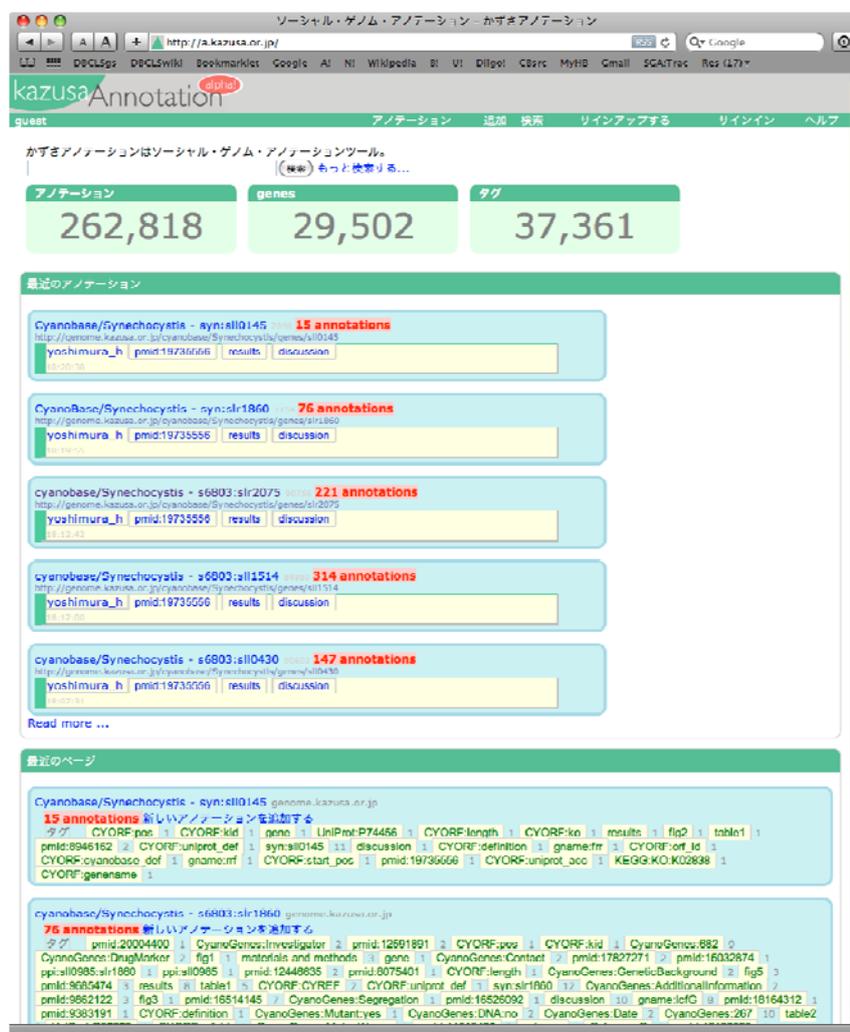


図3.25 高度情報集積データベース環境KazusaAnnotation改善版

ゲノム上の遺伝子に関わるアノテーション情報の統合的な可視化システムである **genoDive** については、遺伝子発現や書誌情報といった質の異なる情報を統合的な可視化表現をデザインし新たに加えた。さらに **BioHackathon**、**Biocuration meeting** といった国際学会や、統合データベースセンター主催の講習会で発表し、専門家の意見をききながら修正と改善を加えた。

植物ゲノムのデータベース中の遺伝子名や遺伝子IDの齟齬を解決する試みとして、個別のゲノムデータベース上で全文検索可能なシステムを開発し組み込んだ。

2) ゲノムアノテーション情報の蓄積と高度化

上記の高度情報集積データベース環境 **Kazusa Annotation** を運用しつつ、その環境を利用したアノテーション情報の蓄積と高度化を平成20年度に引き続き進めた。平成21年度は、平成20年度と同様に論文中に出現する遺伝子・タンパク質記載のインデクス情報 (**Gene Indexing**型アノテーション) を蓄積した。具体的な生物としては、*Synechocystis* sp. PCC 6803をはじめ8種類の光合成細菌に関する合計約3500報の論文について、また、*Mesorhizobium loti* MAFF303099をはじめ5種類の窒素固定根粒菌に関する合計約2,600報の論文について遺伝子の位置情報の登録を進めた。その他、高等植物としてマメ科のミヤコグサ (*Lotus japonicus*) に関する約300報の論文も登録している。合計14生物種の約6400報の文献から遺伝子とタンパク質の情報を163,173件抽出蓄積しゲノムの位置情報と論文記載情報の統合を図った。これらの情報は上記のデータベースから即時公開し、広く生物学者の利用に供している。また、新しいタイプのゲノム情報データベースであるので、その概念と活用方法を一般の生物学者に周知するため、植物研究関係のシンポジウムや講習会で紹介を行い、利用を促した。

3) 植物の統合

生物種をまたいだ横断的なゲノム情報の取得や、理化学研究所など他機関の植物オーミクス情報との連携のため、エントりに付与した種々の属性により高度なデータの絞り込みを可能とするデータマネジメントシステムである **BioMart** (<http://mart.kazusa.or.jp>) を導入した。このシステムに酸素発生型光合成細菌の15種類のゲノム情報を蓄積し公開した。

4. 統合データベース支援 (情報・システム研究機構)

4. 1 ポータル整備・運用、広報、普及啓発

4. 1. 1 ポータル整備・運用、広報、普及啓発の実施計画

平成19年10月に10件のサービスからスタートしたポータルサイト「統合ホームページ」は、平成20年6月には横断検索をはじめとする32件のコンテンツを公開し本格公開をスタートさせた。その後、平成20年度末現在で関連機関の開発したサービスを含む43件のコンテンツが公開されている。平成21年度は主にその運用を継続し、各コンテンツの更

新や改良を続けるとともに、開発されたサービスや技術を連携させ機能向上に努める。国内DB中心に開発を進めてきた横断検索は平成20年度末までに国内のDBやウェブサイトから200DBを対象に検索インデックスの作成を実施した。平成21年度はそのメンテナンスを続けるとともに、海外DBにも対象を広げ検索の充実化を図る。検索においては、学名や解剖名、疾患名などに対するキーワードの日英翻訳機能を強化するとともに、統合検索に向けてメタデータやキーワードを利用した検索手法を検討する。また、統合プロジェクト内の各機関や各省の統合プロジェクトとの検索機能の連携を進める。

広報・普及啓発では、平成20年度、大学院生学生を対象に8回の講習会を開催し統合プロジェクトの普及を図ったが、平成21年度も各地での講習会の開催を予定している。動画教材である統合TVは平成20年度150本にのぼる動画を製作し重要なサービスへと成長したが、今年度も引き続き製作を続け、DBやツールの更新に対応するように過去の動画の更新も行っていく。広報活動は平成20年度BMB2008でのブース出展などを行い利用者への普及を図ったが、平成21年度も国内の主要学会への出展を計画している。またインターネットやパンフレット、各種広告など様々な媒体形態を利用して統合DB活動の普及を行う。またそれと同時に、サービスに関するアンケート等を定期的実施し利用者の意見をサービスに反映させて行くこととする。

4. 1. 2 ポータル整備・運用、広報、普及啓発の実施内容

(1) 統合DBプロジェクトポータルサイトの整備・運用

統合データベースプロジェクトの成果全体の公開サイト「統合ホームページ」は平成19年10月に32のサービス及びコンテンツからスタートし、平成20年度は46件となり、平成21年度は新規サービス8件を加え、平成22年3月末現在53件を掲載している。平成21年度はサイトの外観や操作機能に大きな変更を加えず、各コンテンツの更新や改良を続けるとともに、開発されたサービスや技術を連携させ機能向上を進め、後述する利用者への普及活動に努めた。平成21年度新規に追加されたサービス及びコンテンツは、MiGAP(平成21年6月)、inMeXes(平成21年7月)、DDBJリードアーカイブ(平成21年8月)、DBCLS Galaxy(平成21年10月)、シンポジウム・講演会のページ(平成21年10月)、LSDB Lab(平成21年12月)、辞書の構築と公開(平成22年1月)、KazusaMart(平成22年3月)である(図4.1)。

このうちLSDB Lab.ページは正式公開前のサービスの公開サイトとして開設されたもので、平成22年3月末現在で、Gendoo(Gene, Disease Features Ontology-based Overview System)、togoexp (Website for constructing/maintaining reference expression dataset)、TogoProt(蛋白質関連データベース統合検索)、TogoDoc(文献情報管理・推薦システム)の4サービスを公開中である。一部の物は今後正式サービスへと展開する予定である。

統合ホームページにリンクされた統合データベースプロジェクト発のデータベースや公開資料は、従来サービス別の検索は可能であったが、平成21年度新たに生命科学データ

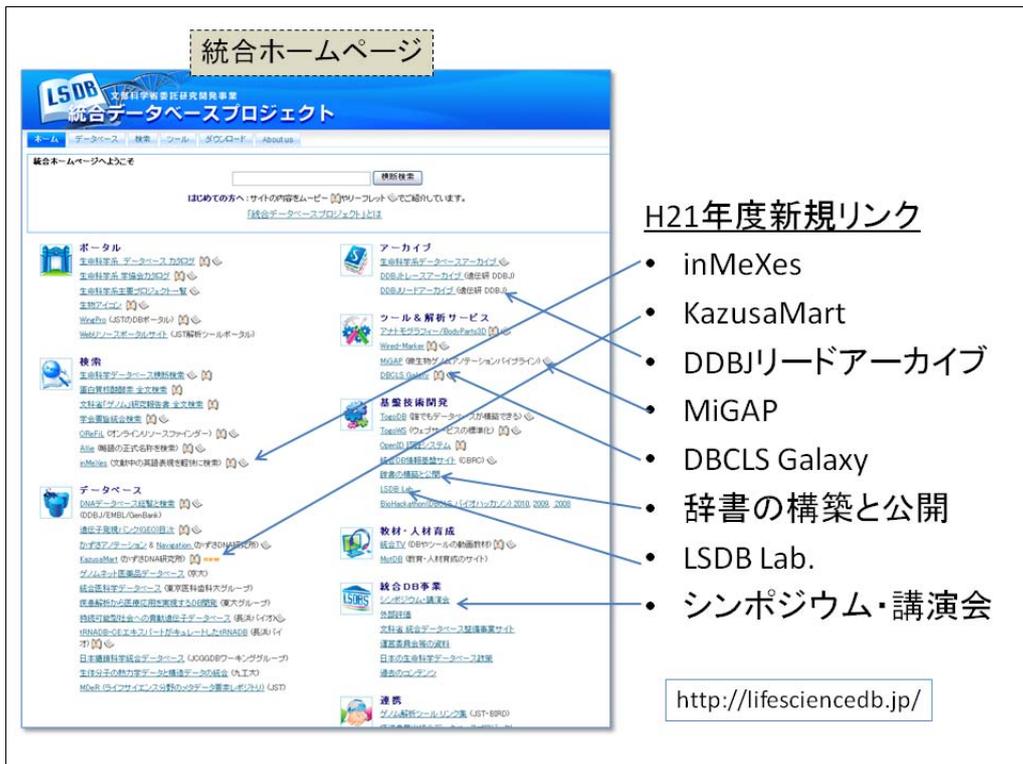


図 4.1 プロジェクト成果公開サイト「統合ホームページ」



図 4.2 統合 DB プロジェクト内コンテンツの横断検索

ベース横断検索(以下、横断検索)にインデックスを追加し、一括して検索が可能となった(図 4.2)。これにより、プロジェクトで開催された研究運営委員会や作業部会、日本の生命科学データベース政策に掲載された、関連資料の検索閲覧が可能となった。プロジェクトには分子 DB から動画、会議資料まで種々のものがあるため、一括したサイト内キーワード検索は実効的では無いが、横断検索の結果をふまえて、平成 22 年度(最終年度)には統合 DB プロジェクトの成果検索サービスを設置する予定である。

横断検索サービスは国内のデータベースを中心に約 200 のデータベースを一括して検索することが可能なサービスであり、検索エンジンにフリーソフトウェアの HyperEstrailer(<http://hyperestraier.sourceforge.net/>)を利用している。HyperEstrailer の特徴の一つはクライアント/サーバ方式のプログラムが提供されている点である。個々のデータベースに対するインデックスをネットワーク経由で操作できるため、検索サーバを分散管理することが可能である。平成 20 年度は補完課題の日本糖鎖科学統合データベース

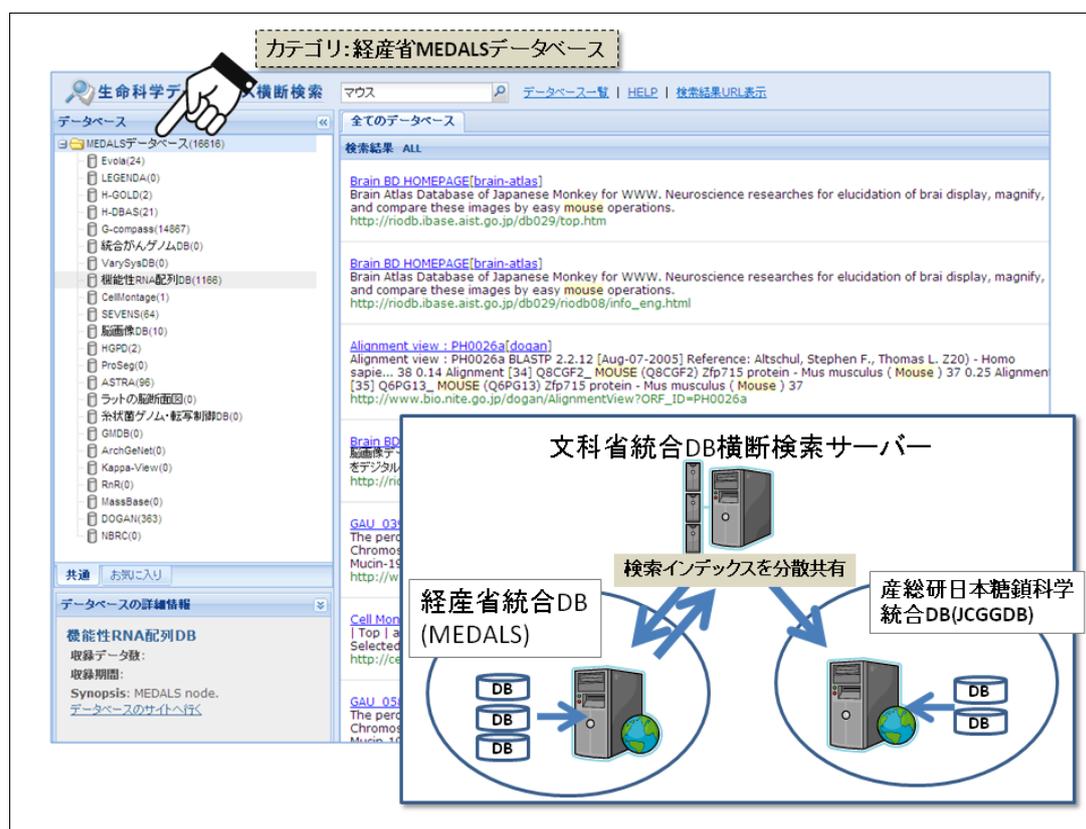


図 4.3 HyperEstrailer を利用したサーバ分散型の検索サービスの構築

(JCGGDB)と連携し相互検索を実現したが、平成 21 年度は他省庁との連携を進める中で特に、経産省統合 DB プロジェクト・MEDALS に中核機関で開発した検索プログラムを提供し、これにより両プロジェクトの検索連携が可能となった。図 4.3 に経産省の計算サ

一バにクエリを送信し、検索結果を取得し表示させた例を示す。利用者側からは、検索サーバが物理的にどこに存在するかは関係無く、各機関で保有するデータベースの最新の検索結果を得られるなどの利点があり、国内データベースの検索の統合化に向けた第一歩となるものである。一方、さらに利用者の立場に立てば、文科省、経産省、各機関といった運営者の違いは検索時には不要であることから、検索は1機関が担うので十分であるという意見が経産省側からも提出されており、今後の検討課題である。

横断検索は、日本語キーワード検索機能やカテゴリツリーの操作による結果の絞り込み機能など2、3の機能をあわせ持っているが、平成20年度までの設計は搭載するデータベース数のスケーラビリティを重視したものになっている。データベース別のインデックスの更新管理や、データベースの追加、削除は容易に行える一方で、関連検索、配列検索など、統合検索に必要な機能は未開発であった。そこで平成21年度は、海外データベースの索引作成並びにキーワードからキーワードへ、キーワードから配列への連携を高めるために、関連キーワード表示機能並びに Blast 検索連携機能のプロトタイプを開発した(図4.4)。関連キーワード表示機能では、プロジェクトで構築した辞書を追加した他、検索においては、これまで検索に組込まれていた学名や解剖名、疾患名などに対するキーワードの日英翻訳機能を強化するとともに、関連キーワード並びに、Wikipedia のリダイレ



図 4.4 横断検索連携機能の開発

クト情報を利用した同義語を表示するようにした。それぞれのキーワードから再度検索をかけることが可能である。一方の Blast 検索連携機能は、検索結果が核酸配列の場合 Genbank に対して Blast ホモロジー検索を行う機能である。検索結果のエントリーから Blast のクエリとなる核酸配列を取得する際に一部 TogoWS を利用した。また Blast 検索は DDBJ の WABI (http://xml.nig.ac.jp/index_jp.html) で提供されている Web サービスを利用している。この成果をふまえ、平成 22 (最終) 年度には、キーワードの整備とバイオインフォマティクスツール解析との連携を進めるとともに、セマンティックウェブの技術を用いた統合検索を実現する予定である。

以上の他、ポータルメニューであるデータベースカタログ、生物アイコン、学協会カタログは大きな変更点は無く主としてデータの追加もしくは更新を行った。データベースやツールの動画教材である統合 TV は、統合ホームページ英語版用の英語動画作成と各種有用データベースの使い方のアップデートを中心に新規に 120 件の動画を作成し計 274 件となった。

(2) 広報、普及啓発 (情報・システム研究機構)

統合 DB プロジェクトの普及のために、実際にインターネットにつながった端末を前にして主として初心者向けに行う統合データベース講習会を 21 年度は計 7 回 (平成 21 年 4 月 17 日於東京大学本郷キャンパス浅野地区(DBCLS)、平成 21 年 5 月 15 日於日本大学生物資源科学部、平成 21 年 7 月 10 日於自然科学研究機構基礎生物学研究所、平成 21 年 9 月 7、8 日於北海道大学、平成 21 年 10 月 16 日於長浜バイオ大学、平成 21 年 11 月 6 日於大阪府立大学りんくうキャンパス、平成 22 年 1 月 18、19 日於久留米ビジネスプラザ) 開催した。このうち、2 回の講習会については 2 日にわたる構成とし、初日は通常の講習会、2 日目は中上級者向けのリクエストに基づく内容とした。このほか、中級者以上向けに Nucleic Acids Research 誌 DB 特集号およびウェブサーバ特集号を俯瞰する統合データベース勉強会を 1 回 (平成 21 年 11 月 25 日於東京大学本郷キャンパス浅野地区(DBCLS))、および単独講師による短いバージョンあるいは実習なしの講演会を計 5 回 (平成 21 年 10 月 16 日於長浜バイオ大学、平成 21 年 11 月 9 日於東京農業大学、平成 21 年 12 月 4 日於静岡県立大学、平成 22 年 3 月 10 日於酒類総合研究所、平成 22 年 3 月 19 日於東京大学本郷キャンパス浅野地区(DBCLS)) 開催した (図 4.5)。

統合データベース講習会/勉強会で使用したテキストやその様子を録画した動画は、ウェブサイト MotDB から利用可能となるよう過去の分も合わせて維持管理し、常に最新の情報が提供されるようにした。一部は ustream などで開催と同時配信も行った。平成 20 年度に引き続き、プロジェクトで開発したデータベースやサービスの紹介、また、その他データベース統合化に必要な活動を紹介するために、学会年会におけるワークショップやフォーラムの企画・運営 (日本分子生物学会 [12 月])、ブースの出展 (BioExpo [7 月])、セルロース学会 [7 月]、脂質生化学会 [7 月]、日本癌学会 [9 月]、日本生化学会 [10 月]、

日時	場所	参加者
〈講習会〉		
2009/4/17	DBCLS	30名
2009/5/15	日本大学生物資源科学部	24名
2009/7/10	基生研	17名
2009/9/7-8	北海道大学	20名+11名
2009/10/16	長浜バイオ大学	43名
2009/11/6	大阪府立大学りんくうキャンパス	33名
2010/1/18-19	久留米ビジネスプラザ	20名+12名
〈講演会・勉強会〉		
2009/10/16	長浜バイオ大学	NA
2009/11/9	東京農大	NA
2009/11/25	DBCLS	11名
2009/12/4	静岡県立大学	NA
2010/3/10	酒類総合研究所	27名
2010/3/19	DBCLS	30名+80名



図4.5 講習会の開催

日本生物物理学会〔10月〕、日本分子生物学会〔12月〕、日本農芸化学会〔3月〕)やランチョンセミナーの開催(日本生物物理学会〔10月〕、日本農芸化学会〔3月〕)を実施した。また、出展した学会等の展示会においても引き続きアンケートを実施し、サービス改善のための情報を得るとともに、ユーザ評価者の募集を行い、合計132名の新たな評価者を得た。統合DBプロジェクトのサービスを紹介した総合パンフレットについては内容を改訂し、合計4300部以上配布、個別サービス紹介のためのリーフレット(現在35種〔うち英語版9種〕)についても内容を随時改訂して展示会等で配布した。ウェブサイトについては、昨年平成20年度公開したプロジェクトサイト英語版の充実のために、説明資料(リーフレットや統合TV)の英語化を行った。また、センターサイトのリニューアルも行い、平成22年度早々に公開する予定で作業を進めている。情報発信のために、プロジェクトウェブサイトおよびJSTの「文科省 統合データベース整備事業サイト」を通じたニュース配信を84回行い、そのうちの一部(9回)については日経のBiotechnology Japanのサイトを利用したPRも実施した。さらにDBCLS用のTwitterアカウントを取得し、ウェブサイトに掲載したニュースをそのまま流すことができる外部サービスを利用して、情報発信を開始した。Twitterアカウントでは、このほか、DBCLSメンバーが発表する講演会等の情報も配信している。また、製薬企業を中心としたメーリングリスト、「ゲノム」特定領域研究の班員のメ

ーリングリスト等に対し、プロジェクトの案内や人材育成プログラムへの参加者募集を行った。

4. 2 データベース受入

4. 2. 1 データベース受入の実施計画

我が国のプロジェクトデータベースについて、データベースのタイプに応じた方法で、データベース受け入れを実施する。平成 20 年度は、メタデータ仕様や、利用許諾体系、覚書等の受け入れの基盤整備を行い、それを用いて主にプロジェクト参加機関のデータベース受け入れを実施した。平成 21 年度は、プロジェクト参加機関のデータベース受け入れと連携を進めると同時に、より広い範囲の機関からのデータベース受け入れを実施する。また、平成 20 年度開発した受け入れデータベースの公開システムに、平成 21 年度は、検索機能・横断検索との連携機能等の拡張を行い、運用を行う。さらに、我が国で実施された主要な生命科学系プロジェクトの調査を実施し、その結果を公開する。

4. 2. 2 データベース受入の実施内容

1) 生命科学系データベースアーカイブの機能拡張

①TogoDB の機能の拡張と応用

DB テンプレート TogoDB に関して、まずユーザ側の利便性向上のために、テーブル間の 1 対多リンク、リスト表示での件数の設定、詳細表示でのセルの高さの切替などを設けた。図 4.6 に、Dicty_cDB (粘菌の cDNA データベース) の例を示す。Contig テーブル (図 4.6-(a)) から Clone テーブル (図 4.6-(b)) へと 1 対多リンクを張ることにより、例えば cDNA ライブラリプロファイルによりコンティグを検索し、その中に含まれる各クローンの配列等の情報へと辿ることが可能になっている。また Contig 詳細情報ウィンドウ (図 4.6-(c)) では、情報量の多い項目のセルの高さを切替可能にして見やすくしている。このようにして、複雑で多様なコンテンツを有する DB についても、利便性をあまり落とすことなく TogoDB の枠組みに収めることができるようになった。

また DB 管理者側の利便性向上のために、テーブルや詳細ウィンドウでの非表示やダウンロード非対象を項目毎に設定できるようにした。また、テーブルに新たに項目を追加しその項目のデータのみをアップロードできるようにした。これらの拡張によって寄託ごとに異なる様々な状況に効率よく柔軟に対処できるようになった。

②アーカイブサイトの改定

データベースアーカイブサイトのユーザ向けの改定としては、サイトの Top Page にアーカイブ一覧を設置し、関連項目を隣に配置し、オリジナル DB の Top ページの画像を左

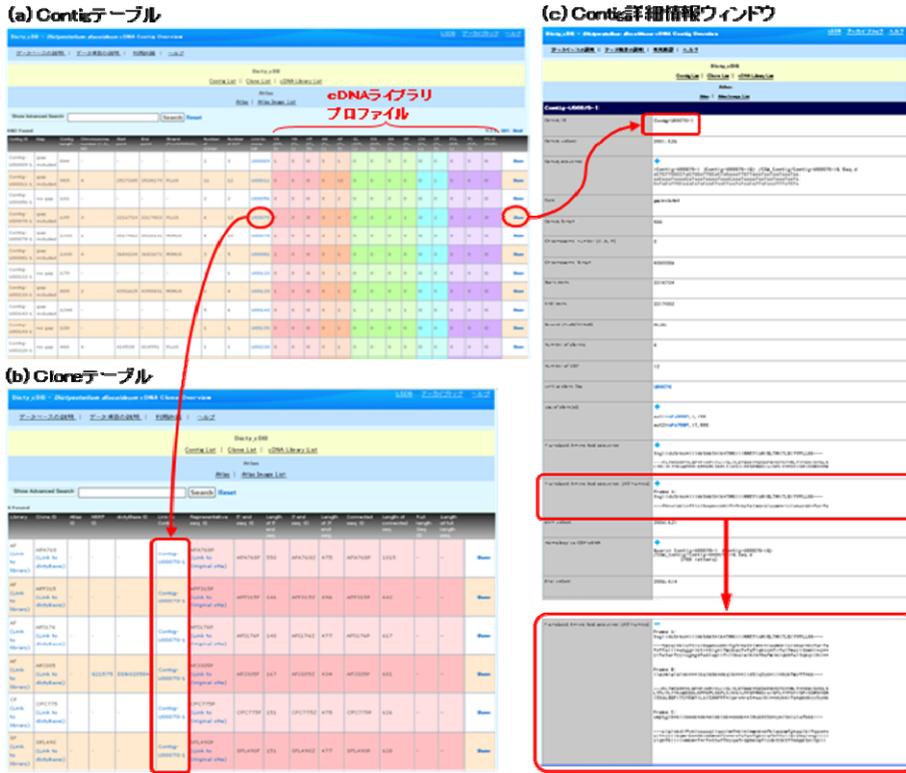


図 4.6 TogoDB のインターフェースの機能拡張 (Dicty_cDB の例)



図 4.7 生命科学系データベースアーカイブのトップページ

端に置くなどの工夫により、見やすく使いやすいページになるよう改定した（図 4.7）。利用許諾について、寄託者の手間を低減するために、一般的に採用されやすいと思われる標準利用許諾と追加利用許諾の組み合わせを利用許諾テンプレートとして作成し公開した（図 4.8）。また寄託者との間の交渉をスムーズに進めるために、覚書の文言を簡素化した寄託規約を作成し（図 4.9）、寄託者の理解を取り付ける形の契約締結を行うこととした。これら利用許諾テンプレートと寄託規約の採用により、寄託者との間の寄託契約の締結を、迅速にかつ一定の様式で進めることができるようになった。さらに、DB コンテンツ中の個別の写真の利用許諾について、個別の写真の権利を保護するために、利用許諾表示を画像に付加する方式を考案し、データベースの事例において採用した（図 4.10）。

アーカイブサイトについても、その構築と運用を管理する機能の構築を行った。これまで、各 DB のメタデータや利用許諾等のページ情報を個別に管理していたため、枠組みや共通の文言の修正を行うときにも個別に修正を行う必要があった。そのため、次第に寄託 DB の数が多くなってきて管理運用に非常に手間がかかるようになってきた。そこで、各 DB のメタデータやその他の情報を一元的に管理する仕組みを構築した。これらにより、今後ますます増加するアーカイブ化作業において、重複する手間を減らすことによる効率化と間違いの低減を実現することができるようになった。

③横断検索への導入

データベースアーカイブのメタデータをインデックス化することにより、横断検索への導入を行った。これによりキーワード検索でアーカイブを探すことが可能になり、またキーワード検索を主な DB 利用法とするユーザにもアーカイブ利用を促せるようになった。

④アーカイブ以外の受け入れ方法の検討

ほとんどのデータベースはこれまでに構築したアーカイブの形で受け入れて公開することができると考えられるが、今後データがさらに多様化するとアーカイブ以外の受け入れ方法を検討することが必要であると考え、検討を行った。

その結果、現在主に用いている簡易検索機能を併用した形に比べると、自前サーバもしくはクラウドを用いたミラーサーバの構築は、個別 DB 独自の機能を移管でき DB 寄託者側からの支持は得易いが、DB 間で利用法に統一感がないためユーザからの支持は得難く、またコストが高いことがわかった。しかし今後受け入れる DB の内容によっては、クラウドを用いたミラーサーバ構築も検討の余地があると考えられる。

2) データベースの受け入れと公開

①各種機関からの受け入れ

i) 参加機関からの受け入れと公開

平成 20 年度に受け入れていたデータベースのうち、理研 1 件（理研播磨 重原子データベ

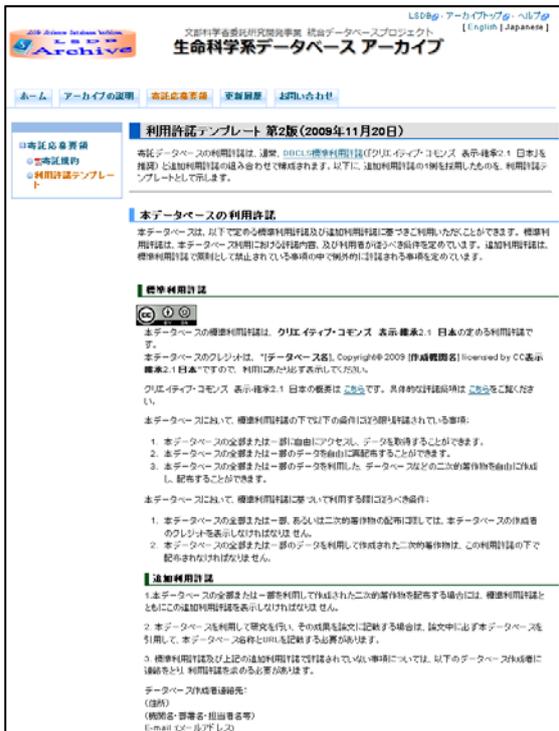


図 4.8 利用許諾テンプレート

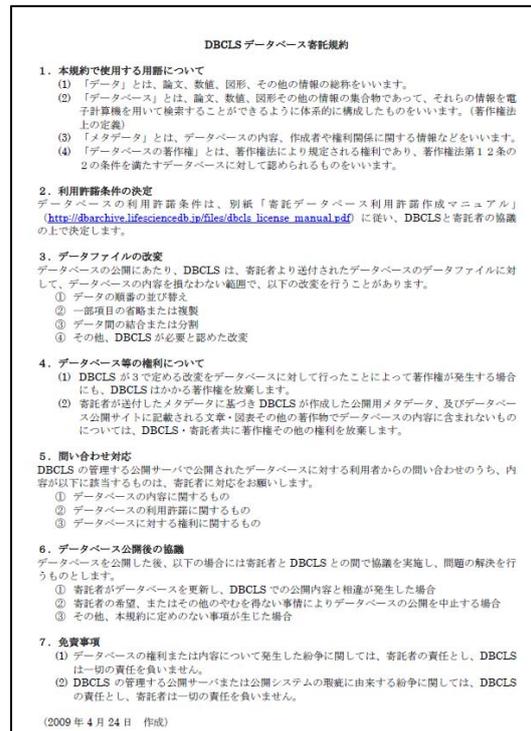


図 4.9 データベース寄託規約



図 4.10 写真の利用許諾表現

ース) についてアーカイブ公開を行った。また新たに長浜バイオ大学より 1 件のデータベース (tRNADB-CE) の受け入れとアーカイブ公開を行った。

ii) データベース構築者へのアンケート回答者からの受け入れと公開

平成 19 年度に実施したデータベース構築者へのアンケートで協力的な返答を寄せていた機関に対して、寄託への賛同を呼びかけた。その結果 4 件の DB (Fluorome (東京大学、現早稲田大学)、Gclust Server (東京大学)、Dicty_cDB - Dictyostelium cDNA Database (筑波大学)、FAMSBASE(北里大学))を受け入れ、3 件の DB(Fluorome、Gclust Server、Dicty_cDB - Dictyostelium cDNA Database) のアーカイブ公開を行った。

iii) 文科省ゲノム特定研究の課題機関からの受け入れと公開

ゲノム特定研究でデータベースを構築している課題 135 件を対象にして、データの種類によって分類を行った結果、公共バンクが存在する 1 次データ、公共バンクが存在しない 1 次データ、2 次データがそれぞれ、51 件、22 件、62 件あるという調査結果を得た。これらのうち公共バンクが存在しない 1 次データがアーカイブ化の意義がより高いものと考え、それらのデータとして変異体と表現型、及び cDNA 等の発現画像データベースをアーカイブ化の優先的なターゲットに選び、寄託への賛同を呼びかけた。その結果、以下に示す 8 件の DB (下記は運用場所) について受け入れを実施しアーカイブ公開を行った。

・ Atlas (ISH Data Base)	筑波大学
・ EGTC - Database for the Exchangeable Gene Trap Clones	熊本大学
・ KTmAbDB - C. elegans Monoclonal Antibody Collection	理化学研究所
・ Transposon insertion site database - germline	大阪大学
・ WorTS - Worm TS mutant Database	大阪大学
・ GETDB - Gal4 Enhancer Trap Insertion Database	京都工芸繊維大学
・ The NAISTrap database	奈良先端科学技術大学院大学
・ DGSP Database - Drosophila Gene Search Project Database	首都大学東京

iv) その他の受け入れと公開

文科省ゲノムネットワークプロジェクトの成果である DMPD (Dynamic Macrophage Pathway CSML Database)、Yeast Interacting Proteins Database、及び Budding yeast cDNA sequencing project を東京大学より、AcEST を首都大学東京より、DIAM-安全に関する文献情報をバイオインダストリー協会 (JBA) より受け入れた。これらのうち、JBA からのデータベースを除いてアーカイブ公開を行った。

以上から、これまでの受け入れと公開状況をまとめると、平成 20 年度受け入れた 12 件のうち 6 件はアーカイブ公開済み (5 件は 20 年度、1 件は 21 年度公開)、2 件はアーカイブ公開作業中、4 件はアーカイブ公開検討中である。平成 21 年度は、プロジェクト関係機関だけでなくより広い範囲の機関から合計 18 件を受け入れて、そのうち 16 件はアーカイブ公開済み、2 件はアーカイブ公開作業中である。受け入れの累積総数は 30 件である。この内訳は、アーカイブ公開済みが 22 件、アーカイブ公開作業中が 4 件、アーカイブ公開検討中が 4 件である。また、19 年度はアーカイブではないが、2 件ミラー公開の形で受け

入れている。これらのうち、すでに公開中のデータベースについて受け入れ年度や公開年度、寄託ソース毎にまとめた表を、以下の表 4.1 に示す。データベース毎にアーカイブファイルの数と内容についても記載している。

表 4.1 公開中のデータベースアーカイブの内容

データベース	データベース運用場所	代表者	データベースカテゴリ	生物種	要約	ファイル数	ファイル内容	寄託元	受け入れ年度	公開年度
Arabidopsis thaliana EST index	かずさ DNA 研究所 植物ゲノム研究部	中村 保一	DNA 配列	シロイヌナズナ	シロイヌナズナの EST 配列と、それらの配列と NCBI nr データベースとの間で行った blastx 検索結果。	4	・EST 配列一覧 ・5' EST 配列 ・3' EST 配列 ・blastx 検索結果	1.参加機関	20年度	20年度
Chlamydomonas reinhardtii EST index	かずさ DNA 研究所 植物ゲノム研究部	中村 保一	DNA 配列	クラミドモナス	クラミドモナスの EST 配列と、それらの配列と NCBI の nr データベースとの間で行った blastx 検索結果。	4	・EST 配列一覧 ・5' EST 配列 ・3' EST 配列 ・blastx 検索結果	1.参加機関	20年度	20年度
Lotus japonicus EST index	かずさ DNA 研究所 植物ゲノム研究部	中村 保一	DNA 配列	ミヤコグサ	ミヤコグサの EST 配列と、それらの配列と NCBI の nr データベースとの間で行った blastx 検索結果。	4	・EST 配列一覧 ・5' EST 配列 ・3' EST 配列 ・blastx 検索結果	1.参加機関	20年度	20年度
Trifolium pratense EST index	かずさ DNA 研究所 植物ゲノム研究部	中村 保一	DNA 配列	ムラサキツメクサ	ムラサキツメクサの EST 配列と、それらの配列と NCBI nr データベースとの間で行った blastx 検索結果。	4	・EST 配列一覧 ・5' EST 配列 ・3' EST 配列 ・blastx 検索結果	1.参加機関	20年度	20年度
GlycoProtDB	独立行政法人 産業技術総合研究所 糖鎖医学研究センター グライコプロテオーム解析チーム	梶 裕之	構造・糖	生物全般	質量分析法を主体としたプロテオミクスの手法で実験的に同定された N 結合型糖タンパク質の糖結合位置のデータベース。	1	・3N 結合型糖鎖の修飾位置	1.参加機関	20年度	20年度
HATODAS - 重原子データベース	独立行政法人 理化学研究所 播磨研究所 放射光科学総合研究センター	国島 直樹	立体構造	生物全般	タンパク質の X線結晶解析で行われる重原子化の事例を収集したデータベース	2	・重原子レコード ・モチーフ	1.参加機関	20年度	21年度
tRNADB-CE	長浜バイオ大学 生体分子情報学研究室	阿部 貴志	RNA 配列	原核生物	世界的に最も精度が高い、エキスパートが精査を加えた tRNA 遺伝子データベース	1	・tRNA 配列、アノテーション及びキュレーションのデータ	1.参加機関	21年度	21年度
Dicty_cDB - Dictyostelium cDNA Database	筑波大学 大学院生命環境科学研究所 ゲノム情報	漆原 秀子	cDNA	粘菌	Dictyostelium discoideum cDNA の塩基配列と機能アノテーション結果 (相同性検索・モチーフ検索・産物の局在性予測・発現特性・ゲノムへのアライメント等) の情報を収録したもの	3	・Contig List ・Clone List ・cDNA Library List	2.アンケート回答者	21年度	21年度
Gclust Server	東京大学大学院 総合文化研究科 広域科学専攻 生命環境科学系	佐藤 直樹	蛋白質配列	95 生物種	総当たり Blast 比較の結果得られた 95 生物種のタンパク質配列のクラスタのデータベース	3	・Cluster List ・Sequence List ・Similar Sequence List	2.アンケート回答者	21年度	21年度
Fluorome	早稲田大学 教育学部 理学科	園池 公毅	変異体・表現型	シアノバクテリア	原核光合成生物であるシアノバクテリアのクロロフィル蛍光強度の時間変化を変異株ごとに取得し、変異株間で定量比較したデータベース	4	・Fluorescence Data ・Similar ORF list ・Fluorescence Numerical Data ・Fluorescence Image Data	2.アンケート回答者	21年度	21年度
GETDB - Gal4 Enhancer Trap Insertion Database	京都工芸繊維大学	林 茂生	変異体・発現	ショウジョウバエ	ショウジョウバエにおいて、Gal4 エンハンサートランプ系統の挿入位置とエンハンサー活性のパターンをまとめたデータベース	4	・Strain Table ・Clone Table ・Cluster Table ・Expression Image Data	3.ゲノム特定の課題機関	21年度	21年度
WorTS - Worm TS mutant Database	大阪大学 微生物研究所 細胞機能分野	中村 邦明	変異体・表現型	線虫	線虫の胚発生致死温度感受性変異株のデータベース	1	・線虫の変異株データ	3.ゲノム特定の課題機関	21年度	21年度
The NAISTrap database	奈良先端科学技術大学院大学 バイオサイエンス研究科 動物遺伝子機能学講座	石田 靖雅	変異体・発現	マウス	遺伝子トランプで得たマウス ES 細胞のデータベース	1	・トランプされた遺伝子に関する情報	3.ゲノム特定の課題機関	21年度	21年度
DGSP Database - Drosophila Gene Search Project Database	首都大学東京 大学院理工学研究科 生命科学専攻 細胞遺伝学研究室	相垣 敏郎	変異体	ショウジョウバエ	ショウジョウバエの GS ベクター挿入系統のデータベース	3	・GS 系統リスト ・アクセッション番号リスト ・GS ベクター情報	3.ゲノム特定の課題機関	21年度	21年度
EGTC - Database for the Exchangeable Gene Trap Clones	熊本大学 生命資源研究・支援センター	荒木 正健	変異体	マウス	マウスの 可変型遺伝子トランプクローンのデータベース	4	・クローン ・トランプベクター ・セルライン ・メソッド	3.ゲノム特定の課題機関	21年度	21年度
Transposon insertion site database - germline	大阪大学大学院 医学系研究科 環境・生体機能学	堀江 恭二	変異体	マウス	Sleeping Beauty トランスポゾンを用いて作製した変異マウスに関するデータベース	1	・SB トランスポゾン挿入部位データ	3.ゲノム特定の課題機関	21年度	21年度
KTmAbDB - C.elegans Monoclonal Antibody Collection	理化学研究所 発生・再生科学総合研究センター 発生ゲノミクス研究チーム	杉本 亜砂子	局在	線虫	線虫 C. elegans 胚の特異的な細胞種・細胞内構造を認識するモノクローナル抗体コレクションのデータベース	2	・モノクローナル抗体コレクションと免疫染色顕微鏡画像 ・免疫染色顕微鏡画像ファイル	3.ゲノム特定の課題機関	21年度	21年度

Atlas (ISH Data Base)	筑波大学 大学院生命環境科学研究科 ゲノム情報	漆原 秀子	発現	粘菌	Dictyostelium discoideum cDNA クローンについて、Whole-mount in situ hybridization に基づく空間的遺伝子発現情報を収録したもの	3	・cDNA クローンリスト(発生時期別の局在画像付き) ・局在画像リスト ・局在画像ファイル	3.ゲノム特定の課題機関	21年度	21年度
AcEST	首都大学東京 大学院理工学研究科生命科学専攻植物環境応答研究室	鐘ヶ江 健	cDNA	シダ	シダの EST 配列と機能アノテーション結果のデータベース	5	・AcEST の DNA 配列とそのアノテーション ・BlastX Result : Swiss-Prot ・BlastX Result : TrEMBL ・AcEST Contig ・About Libraries	4.その他	21年度	21年度
Budding yeast cDNA sequencing project	東京大学 大学院理学系研究科 生物化学専攻	伊藤 隆司	ゲノム(脊椎動物以外)-真菌ゲノム	酵母	出芽酵母の完全長クローンの5'cDNA 配列データベース	4	・クローンリスト ・5'末端配列 ・クオリティデータ ・ベクター配列	4.その他	21年度	21年度
Yeast Interacting Proteins Database	東京大学 大学院理学系研究科 生物化学専攻	伊藤 隆司	代謝系/シグナル伝達系パスウェイタンパク質相互作用	酵母	Y2H による出芽酵母のタンパク間相互作用のデータベース	4	・Core Data(更新版) ・Full Data(更新版) ・Core Data(Original) ・Full Data(Original)	4.その他	21年度	21年度
DMPD (Dynamic Macrophage Pathway CSML Database)	東京大学 医科学研究所	宮野 悟	代謝系/シグナル伝達系パスウェイ	ヒト、哺乳類	マクロファージの分化、活性化に関するパスウェイデータベース	2	・論文リスト ・パスウェイデータ	4.その他	21年度	21年度
BodyParts3D	ライフサイエンス統合データベースセンター	大久保 公策	器官	ヒト	解剖学用語が示す人体の部品(臓器、器官)の位置と形状を3次元人体モデルで記述したデータベース	7	・3次元臓器モデルの座標系の説明図 ・IDと臓器名称等との対応表 ・臓器、器官間の包含関係記述データ ・臓器 ID 毎のポリゴンデータ1 ・臓器 ID 毎のポリゴンデータ2 ・臓器 ID 毎のポリゴンデータ2	DBCLS	20年度	20年度
生命科学系データベースカタログ	ライフサイエンス統合データベースセンター	川本 祥子	カタログ	—	生命科学系データベースの情報(URL・稼働状況・提供機関・生物種・データ種)を対象や生物種などの項目別に関覧検索できるカタログ型のデータベース。	1	・データベースに関する情報	DBCLS	20年度	20年度
生命科学系協会カタログ	ライフサイエンス統合データベースセンター	川本 祥子	カタログ	—	生命科学系学会の情報(URL・名称・連絡先・会員数・学会誌)を分類別に閲覧検索できるカタログ型のデータベース。	1	・生命科学系の学会に関する情報	DBCLS	20年度	20年度
生命科学系主要プロジェクト一覧	ライフサイエンス統合データベースセンター	西川 哲夫	カタログ	—	日本国内の主要プロジェクトに関する情報(体制、予算、概要、データベースサイト、ダウンロードサイト、報告書や文献へのリンク情報等)を収集したカタログ型のデータベース。	1	・生命科学系主要プロジェクト一覧データ	DBCLS	20年度	20年度
施設名辞書	ライフサイエンス統合データベースセンター	大久保 公策	辞書	—	施設名称や研究室名称と、それらの各表記の対応辞書	1	・施設名辞書	DBCLS	20年度	20年度
遺伝子名称シソーラス	ライフサイエンス統合データベースセンター	大久保 公策	辞書	9 生物種	9 生物種(ヒト、マウス、ラット、ゼブラフィッシュ、ショウジョウバエ、線虫、出芽酵母、分裂酵母、枯草菌)を対象に遺伝子名、遺伝子ファミリー名を集め並列関係【同義】および上下関係【ファミリー】で関係付けました。頭文字表記や代表的な遺伝子・ゲノムデータベース ID 情報も名称と扱います。	1	・遺伝子名称シソーラス	DBCLS	20年度	20年度

表からわかるように、ダウンロードできるデータのファイル数としては 3~4 の場合が多く、その多くは TogoDB のテーブルとして表現しておりテーブル間のリンクによって効率的に情報の閲覧をできるようにしている (図 4.6 も参照)。

②蛋白質関連 DB 統合検索システムの開発

受け入れ DB の整理の一環として、蛋白質関連 DB 統合検索システム TogoProt (図 4.11) を開発した。このシステムにより、ユーザは望みの蛋白質に関する多様なデータを一覧することができる。ユーザがまず日本語または英語のキーワードを入力すると、検索の結果、蛋白質のリストが PIRSF、Pfam、UniRef50 などのファミリーに分類された形で表示され

る。ここで1つの蛋白質を選択すると、その蛋白質に関連づけられる各種データが一覧表示される。ここで蛋白質の代わりにファミリーを選択すると、ファミリーのいずれかのメンバーに関連づけられる各種データが一覧表示される。今年度は12月にTogoProtのプロトタイプを公開し、年度末までにインターフェースの開発を行うとともに、配列・分類・構造・機能・発現・相互作用・辞典など多種多様なDBから40種のDBを選んで、UniProt accessionやPDBIDなどの蛋白IDの付加を行いTogoProtに収録した。

③他省庁との連携

DB受け入れに関連して、経産省統合DBプロジェクトとの連携のため9回の会合を開催し、メタデータや横断検索、アーカイブの共有化について討議した。その結果、メタデータ共有化案の作成と横断検索の仕様統一化を実現した。さらに、DBの法的な問題への当センターの取り組みを経産省統合DB側関係者に説明し討議する会を、コンサルテーションをいただいている弁護士を含めて実施した。これらの活動により、経産省統合DBプロジェクトとの間でメタデータや横断検索は共通仕様に、アーカイブについては文科省統合DBプロジェクトのアーカイブサイトに集約する方向でコンセンサスを得ることができた。

一方、農水省統合DBプロジェクトとの間では、今後の連携のために第一回の会合を開催し、カタログや横断検索、アーカイブなどの連携のあり方や問題点について討議を行い、今後も協議を続けることとした。

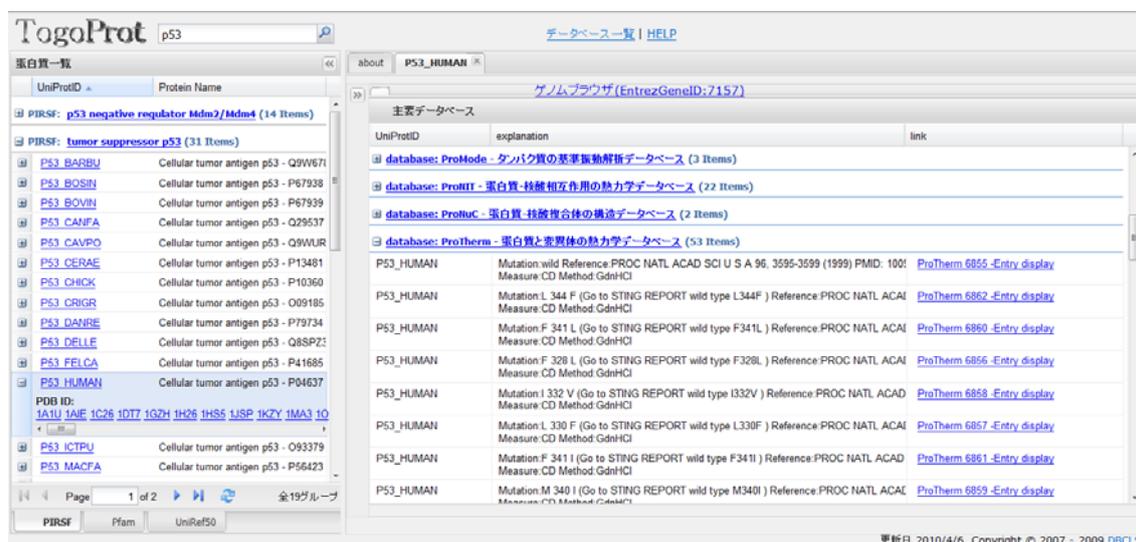


図 4.11 TogoProt の検索画面

3) 生命科学系主要プロジェクトの調査

生命科学系主要プロジェクトの調査については、主要プロジェクト一覧のデータ内容の更新と22件の追加（文科省18件、経産省4件）、及びそれらの英語化を実施した。

4. 3 人材育成（東京大学、お茶の水女子大学、長浜バイオ大学）

4. 3. 1 人材育成の実施計画

（東京大学）

DB 構築者を養成するために以下の 2 つの演習を実施する。

1) バイオ DB サーバ構築演習

データベースサーバのミラーサイトを構築する。OS, apache, MySQL 等の主要ソフトウェアのインストールおよびネットワークセキュリティに習熟することが目標である。参加者はプログラミング経験があることを条件とする。参加者には各自にサーバ構築用ワークステーションを配布する。演習は 1 年間コースと 1 ヶ月コースを設け、後者は 1 年間の継続的受講が困難な企業等の受講希望者に対して 1 年間コースを要約した演習を提供する。各コースとも受け入れ人数は 2-3 名程度、全体で 5 名の DB 構築者を養成する。

2) 独創的サーバー構築演習

平成 20 年度にバイオ DB サーバ構築演習を修了した受講者に対して開講する。大規模計算のためのクラスタ利用技術を習得させ、他に類の無いバイオ DB サーバを設計、実装、公開することを目標とする。平成 21 年度の受け入れ人数は 10 名とする。

演習は英語で書かれたプログラミングの教科書および DB サーバシステムの管理マニュアルの読破も要求し、自ら立ち上げたサーバの英語マニュアルが書けるように指導する。したがって受講者にはある程度の英語能力を想定する。平成 20 年度で作成し公開中の演習ノートも改善する。

（お茶の水女子大学）

統合 DB は日々更新される膨大なデータベースである。このデータベースから最新のデータを取得し、生命理解へと繋げるには、生命科学の知識のみならず DB から必要なデータを抽出、編集、蓄積し、解析や知識発見ができる DB 高度利用者が必要である。平成 21 年度は、平成 20 年度に改良した DB 高度利用者養成カリキュラム案を基に、授業、演習を引き続き実施し、社会人・主婦 2 名以上を含む DB 高度利用者の養成を行う。また、DB 高度利用者養成のための自習用教材の作成とそのインターネット上での内容を充実させる。

（長浜バイオ大学）

1) アノテータ・キュレータの教育

長浜バイオ大学は、これまでの実績を基礎に、情報・システム研究機構と連絡を密に取り、平成 21 年度も引き続き実データに基づくゲノムアノテーション実習を行う。対象とするメタゲノム配列が新規に大量に公的データベースに登録されているので、その新規配列を対象に有用遺伝子のアノテーションを行い、本統合 DB からの公開データとして提供する。また、国立遺伝学研究所・極地研究所・理化学研究所等の公的機関を中心に、決定

された塩基配列へも新規アノテーション情報を付加し、公的 DNA Databank への登録用基礎データとする。公開を前提としたアノテーションを実践することは、技術の習得に加えて、データベース作成の意義や責任を体験することになり、ミスマッチの少ない人材育成となる。なお、これらの人材育成プログラムには、情報・システム研究機構でアノテータやキュレータとして雇用する人材の育成も含む。アノテータ育成のための自習用教材を作成・拡充し、その成果教材を昨年同様インターネット上で公開を行い、他大学、社会人、高校生や高専生を対象とした実習中心の講習会を開く。

2) 自己組織化マップによる大量ゲノム配列データへのアノテーション手法を用いた人材育成

バイオデータの特徴である複雑で大量な情報のデータベース化と、その効率的な活用においては、可視化技術の利用が重要となる。大量複雑情報の可視化技術として確立している自己組織化マップのような既存手法を用いて、混合微生物やウイルス由来の大量遺伝子のアノテーション結果の可視化を行い、得られたアノテーションの成果と可視化データを本統合 DB からの公開データとして更新を行い、利用しやすい形式で提供する。また、国立遺伝学研究所・極地研究所・理化学研究所等の我が国の公的機関を中心にして決定された大量塩基配列データの対象についても、自己組織化マップを用いたアノテーション結果の可視化を行う。バイオデータベースの作成と活用に必須な要素技術となる、可視化技術を備えた4回生を、バイオデータの解析を実例に育成する。また、解析で得られた成果をインターネット上で公開を行う。

3) シニア研究者のチームによる分散型の高度なキュレーションシステムの改良・運用

上記2業務で得られたアノテーション成果の質の向上を目的に、平成21年度も引き続き専門知識を有するシニア世代研究者を中心に、若手人材も含むチームによるネットワークを介した分散型の高度なキュレーションシステムの開発・改良と運用を行う。平成20年度は、tRNA 遺伝子のキュレーションに関するシステムを中心に運用してきた。平成21年度も引き続きシニア世代の知識を組み入れた機能部位の検索システムの改良と、新規に登録された大量塩基配列データを対象にして、tRNA 遺伝子のデータベース追加更新を行う。多くの大学研究者や企業が関心の高いタンパク質遺伝子に関しても、シニア専門家人材の発掘を含めて、高度キュレーションシステムを立ち上げて運用を開始する。又、他大学や一般企業、研究機関の人材を対象に講習会を開催し、大学や研究機関へのキュレーションモデルの普及を図る。

4. 3. 2 人材育成の実施内容

(1) DB 構築者の養成 (東京大学)

公開成果の利用状況は Webalizer または google analytics を使ったアクセス解析により調査した。(Sites: いくつの IP アドレスからリクエストを受けたかを示す。KBytes: 送信したデータ量。Visits: サイトを訪問したユーザー数。Pages: 送信したページ数。画像などのファイルは含まない。Files: 送信したファイル数。セッション: サイトの訪問数。ページビュー: 送信したページ数。)

1) バイオ DB サーバ構築演習

データベースサーバのミラーサイトを構築する演習を実施した。OS, apache, MySQL 等の主要ソフトウェアのインストールおよびネットワークセキュリティに習熟することが目標とし、参加者には各自にサーバ構築用ワークステーションを配布した。1年間コース受講者は4名で、演習を完了するまでには計画通り1年間の時間を要した。平成19年度より作成している演習ノートの web ページ (下記) を充実させた。

<http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php?Introduction>

アクセス数

	Sites	KBytes	Visits	Pages	Files
1月	795	314117	6397	11144	16014
2月	956	286172	5440	10212	15742
3月	999	350515	6953	14348	19413

ユーザーからの評価

詳しい説明があり勉強になった。トラブルへの対処法の記述が充実していた。

2) 独創的サーバ構築演習

バイオ DB サーバ構築演習およびプログラミング演習を修了した受講者に対して独創的サーバ構築演習を実施し、ゲノムブラウザで表示される新規トラックの開発・新規データの表示・塩基配列をゲノムにマッピングし結果をウェブ上で表示するプログラムの実装などを行った。

- ・塩基配列をメダカゲノムにマッピングし結果をウェブ上で表示するプログラムの開発。

<http://medaka.utgenome.org/>

- ・cDNA アセンブルプログラムの開発。

<http://musica.gi.k.u-tokyo.ac.jp/>

- ・ゲノム中の短い配列の出現頻度データベース RepeatScape (下記) を公開した。

<http://repeatscape.utgenome.org>

アクセス数

	セッション	ページビュー
1月	32	68
2月	36	70
3月	31	47

ユーザーからの評価

ゲノム中のリピート配列情報は次世代シーケンサを用いたゲノム解析に有用。

(2) DB 高度利用者の養成（お茶の水女子大学）

平成 21 年度の業務においては、平成 20 年度までに開発した DB 高度利用者の養成に必要なカリキュラム案を基に、授業・演習を行った。平成 20 年度策定のカリキュラムが今後の生物学データの変化に対応できるか評価のため、世界の第一線で活躍する進化生物学者および生態学者が集まる国際会議 ESEB2009 に出張し、研究者達とディスカッションを行った。この会議には、平成 19 年度に調査した SMBE2008 に参加している研究者とは研究者層が異なり、研究者の扱う生物種が多様なため平成 19 年度とは異なる角度で本カリキュラムの評価ができた。その結果、次世代シーケンサを利用した塩基配列や遺伝子発現量データを用いた研究の急速な広がりと共に、そのデータを解析できる人材の需要が急速に拡大していることが確認できた。本カリキュラムでは策定時点から、これらの兆候を捕らえ大規模データ解析に必須となるプログラミングや基礎統計を重点的に設定していたため、海外の研究者からも本カリキュラムが妥当であると意見を頂いた。

平成 21 年度に行った DB 高度利用者養成のための授業・演習受講生のうち、27 名の修士課程、博士課程の学生と 3 名の主婦を含む 5 名の学外からの社会人受講生が所定のコースを修了した。また、別の社会人 5 名が本コースに部分参加した。平成 21 年度の受講生の内、3 名は本自習用教材改訂において中心的な役割を果たし、1 名は統合データベースセンターにおいて文書検索システム開発に携わる事で、データの抽出、編集、蓄積に関する人材として活躍している。また、平成 20 年度の受講生の内 2 名は本自習用教材公開用サーバ管理において中心的な役割を果たし、別の 2 名は本コースで学んだ内容を基に、統合 TV の教材作成を行っている。

内容に関し「教材が丁寧に作られているので、これまで知らなかったプログラミング言語を自分で使う際にとっても助けになる」や「家で復習しているとエラーが出ることもあったが、メールや次の回に質問することで実行できた」と評価の声がある一方で 2 割程度の受講生には「ついていくのが大変」との感想もあった。しかし、統合 DB で活躍できる人材育成のためには、講義のレベルを下げるのではなく、テキストやメール等で質問できるサポート態勢の充実により、各自の進度に応じて教育できる態勢強化を行う。

平成 21 年度も平成 20 年度に引き続き、授業・演習と平行して自習用教材の作成とイン

ターネットでの公開を行った。カリキュラム、自習教材は、<http://togodb.sel.is.ocha.ac.jp>で公開中である。平成 21 年 4 月以降、3 ヶ月毎の Google analytics を利用して調査したアクセス数は下記の通りである。

	セッション	ページビュー
平成 20 年 4 月～6 月	429	2601
平成 20 年 7 月～9 月	1232	3017
平成 20 年 10 月～12 月	1326	3057
平成 21 年 1 月～3 月	1434	5397

(3) アノテータ・キュレータの教育 (長浜バイオ大学)

1) アノテータ・キュレータの教育

初級アノテーション教育 (3 回生 250 名) として、大量に公的データベースに登録されている環境由来メタゲノム配列からの「健康に貢献する遺伝子発掘」を平成 21 年 7 月に完了し、シニア研究者が精査完了したデータを統合 DB から公開した。又、実習授業「バイオ分野のデータベースの統合的な利用」(2 回生 270 名) を平成 21 年 10～11 月に完了した。他大学等で利用可能な自習用テキスト最新版も WEB 公開している。中級アノテーション教育 (50 名) は、国立大学法人の研究グループが次世代シーケンサで決定した 1000 万件に近い微生物ゲノム由来の大量断片配列データを対象に、アノテーション実習を平成 21 年 11 月～同 22 年 1 月に実施し、結果を当該研究グループへ報告した。この実習用に次世代シーケンサ由来配列解析用の学内版テキストを作成したが、その学内版テキストを元に、学外でも使用可能な自習用公開教材の作成を開始した。近隣の 3 高等学校で総計 180 名の高校生を対象に土曜講座や出前実習を開き、バイオ分野の様々なデータベースの利用実習等を実施した。

2) 自己組織化マップによる大量ゲノム配列データへのアノテーション手法を用いた人材育成

公的データベースに登録されている環境由来のメタゲノム配列を対象に、自己組織化マップを用いて、2名の4回生が約2000件の遺伝子の生物系統を推定し、この配列相同性によらない生物系統のアノテーション結果を情報・システム研究機構統合DBセンターから公開した。混合微生物やウイルス由来の大量遺伝子を対象に、自己組織化マップを用いたアノテーション結果の可視化を行い、論文発表ならびにその詳細結果のインターネット上での公開を行った。

3) シニア研究者のチームによる分散型の高度なキュレーションシステムの改良・運用

シニア研究者チームと学部生の共同作業による大量塩基配列データを対象とした

tRNADB-CE に、今年度は新たにウイルスゲノムの tRNA 遺伝子の追加更新を行い、新規に追加した検索機能と共に公開した。更に、農業生物資源研究所から依頼を受けてイネゲノムの tRNA データの精査も完了し、更新バージョンを公開した。多くの大学研究者や企業が関心の高いタンパク質遺伝子に関しても、シニア専門家 1 名が抗生物質生産遺伝子を中心に、公的データベースに登録されている環境由来の大量メタゲノム配列からの探索と、発掘された遺伝子のアノテーション、専門的知識に基づく高度キュレーションを開始した。シニア研究者チームによる高度キュレーションシステムについて、関連学会でのミニシンポジウムの企画を含む、複数学会での発表を行った。

5. プロジェクトの総合的推進（情報・システム研究機構）

5. 1 プロジェクトの総合的推進の実施計画

中核機関である情報・システム研究機構は、分担機関である京都大学、東京医科歯科大学、東京大学大学院医学系研究科および補完課題の実施機関である理化学研究所、産業技術総合研究所糖鎖医工学研究センター、情報・システム研究機構国立遺伝学研究所、九州工業大学に全体戦略を提示し連携して本事業を推進する。平成 20 年度はこのため、分担機関および補完課題実施機関を含めた研究運営委員会作業部会を組織することにより、プロジェクト遂行に関わる実体的な議論ができる場を整備した。また、必要に応じて、個別の参加機関との打合せの実施や書面を用いて、プロジェクト参加機関の連携・調整、及び進捗の確認を行った。平成 21 年度は、これまでの経緯を踏まえ、共通の課題やターゲットを抱えた機関を集めた分科会の開催を増やすなど、引き続き、研究運営委員会作業部会その他の会議の開催等を通してプロジェクト参加機関の連携・調整、及び進捗の確認を行う。また、必要に応じて外部有識者を招聘して意見を聞くなど、プロジェクトの推進に資する。

5. 2 プロジェクトの総合的推進の実施内容

研究運営委員会及び研究運営委員会作業部会

中核機関である情報・システム研究機構は、分担機関である京都大学、東京医科歯科大学、東京大学大学院医学系研究科および補完課題の実施機関である理化学研究所、産業技術総合研究所糖鎖医工学研究センター、情報・システム研究機構国立遺伝学研究所、九州工業大学に全体戦略を提示し連携して本事業を推進するために、平成21年度は、昨年同様、共通の課題やターゲットを抱えた機関を集めた分科会等の開催数を増やすなど、引き続き、研究運営委員会作業部会その他の会議（下記）の開催等を通してプロジェクト参加機関の連携・調整、及び進捗の確認を行った。また、平成22年度は最終年度に当たるため、当初立てた計画通りに進んでいるのか、あるいは修正が必要なのか等の議論も積極的に行った。尚、研究運営委員会及び作業部会分科会のメンバーは下記の通りである。

(研究運営委員会委員 敬称略、順不同)

大倉 克美	(独) 科学技術振興機構研究基盤情報部長
勝木 元也	自然科学研究機構理事
金岡 昌治	(株) 大日本住友製薬研究本部副本部長
金久 實	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
榊 佳之	豊橋技術科学大学長
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長
長洲 毅志	エーザイ (株) 理事・研究開発担当付部長
中村 桂子	JT生命誌研究館館長
中村 春木	大阪大学蛋白質研究所蛋白質情報科学研究系教授
長村 吉晃	(独) 農業生物資源研究所基盤研究領域ゲノムリソースセンター長
松尾 泰樹	(独) 理化学研究所横浜研究所研究推進部長
松原 謙一(委員長)	(株) DNAチップ研究所代表取締役社長
湯元 昇	(独) 産業技術総合研究所理事
吉田 輝彦	国立がんセンター研究所腫瘍ゲノム解析・情報研究部長
吉田 光昭	東京大学大学院新領域創成科学研究科客員教授
大久保公策	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ研究センター教授
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
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長
堀田 凱樹(副委員長)	情報・システム研究機構長

(作業部会分科会委員 敬称略、順不同)

浅井 潔	(独) 産業技術総合研究所生命情報工学研究センター長
黒田 雅子	(独) 科学技術振興機構研究基盤情報部バイオインフォマティクス課長
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
田中 博	東京医科歯科大学情報医科学センター長
田畑 哲之	(財) かずさ DNA 研究所副所長
徳永 勝士	東京大学大学院医学系研究科教授
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長
成松 久	(独) 産業技術総合研究所糖鎖医工学研究センター長
松原 謙一	(株) DNA チップ研究所代表取締役社長
森下 真一	東京大学新領域創成科学研究科教授
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
菅原 秀明	情報・システム研究機構国立遺伝学研究所

生命情報・DDBJ 研究センター特任教授

高木 利久(主査) 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長

上記の体制で、開催した研究運営委員会及び研究運営委員会作業部会は次の通りである。

研究運営委員会拡大懇談会 平成 21 年 4 月 13 日
研究運営委員会 (第 7 回) 平成 21 年 6 月 25 日
研究運営委員会 (第 8 回) 平成 21 年 10 月 23 日
研究運営委員会 (第 9 回) 平成 22 年 1 月 25 日

作業部会分科会 (第 5 回) 平成 21 年 5 月 20 日
作業部会分科会 (第 6 回) 平成 21 年 5 月 25 日
作業部会分科会 (第 7 回) 平成 21 年 5 月 27 日
作業部会分科会 (第 8 回) 平成 21 年 5 月 29 日
作業部会分科会 (第 9 回) 平成 22 年 1 月 22 日
作業部会分科会 (第 10 回) 平成 22 年 1 月 29 日
作業部会分科会 (第 11 回) 平成 22 年 2 月 2 日

5. 3 その他シンポジウム、セミナー等

(1) 統合データベース講習会

(情報・システム研究機構、かずさ DNA 研究所、長浜バイオ大学)

4. 1. 1. 項で記載した統合データベース講習会の詳細は以下のとおりである。

①統合データベース講習会: AJACS 本郷 4

日時: 平成 21 年 4 月 17 日 (金) 13:00-17:30

場所: DBCLS 教室

参加者: 30 名

対象: ライフサイエンス分野のデータベースを効率よく利用することに興味のある方。

内容:

13:00-13:15 坊農秀雅「はじめに: 統合データベースプロジェクトとは？」

13:15-13:45 坊農秀雅「データベースを使いこなすための前準備」

13:45-14:30 川本祥子「生命科学横断検索の利用法」講習で利用する URL のリスト

14:30-14:50 休憩

14:50-15:20 坊農秀雅「DNA データベース総覧と検索、遺伝子発現バンク(GEO)目次の使い方」

15:20-16:05 大久保公策「アナトモグラフィ/BodyParts3D の利用法」
<http://lifesciencedb.jp/ag/>
16:05-16:35 仲里猛留「文献検索にまつわる各種ツールの使いこなし術」
16:35-16:50 休憩
16:50-17:30 大喜利「明日から使えるデータベース・ウェブツール」(河野信、小野浩雅、他)

②統合データベース講習会：AJACS 湘南 2

日時：平成 21 年 5 月 15 日 (金) 10:00-17:30

場所：日本大学生物資源科学部

参加者：24 名

対象：日本大学生物資源科学部の学部 3、4 年生や大学院生、教員を始めとしたライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

10:00-10:15 「はじめに：統合データベースプロジェクトとは？」(坊農秀雅)

10:15-10:45 「データベースを使いこなすための前準備」(坊農秀雅)

10:45-11:30 「生命科学横断検索の利用法」(川本祥子)

11:30-13:00 休憩

13:00-13:30 「DBCLS 目次サービス (DNA データベース総覧と検索、遺伝子発現バンク(GEO)目次) の利用法」(坊農秀雅)

13:30-14:00 「アナトモグラフィ/BodyParts3D の利用法」(三橋信孝・藤枝香)

14:00-14:30 「文献検索にまつわる各種ツールの使いこなし術」(仲里猛留)

14:30-14:45 休憩

14:45-15:45 「統合 TV を駆使した遺伝子発現データの使い倒し術」(小野浩雅)

15:45-16:15 「Genome Annotation と Kazusa Annotation」(岡本忍)

16:15-16:30 休憩

16:30-17:30 大喜利「明日から使えるデータベース・ウェブツール」

③統合データベース講習会：AJACS 三河

日時：平成 21 年 7 月 10 日 (金) 13:00-18:00

場所：自然科学研究機構 基礎生物学研究所

参加者：17 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのような

に作成、維持管理されているかに興味のある方

内容：

13:00-13:15 「はじめに：統合データベースプロジェクトとは？」

13:15-14:15 「生命科学横断検索の利用法」

14:15-14:45 「DNA データベース総覧と検索の使い方」

14:45-14:55 休憩

14:55-15:35 招待講演

o 内山郁夫「微生物ゲノム比較のためのデータベース MBGD」

o 成瀬清「メダカ完全長 cDNA リソースの整備とデータベース構築」

15:35-16:20 「遺伝子発現データベースを使い倒す」

16:20-16:30 休憩

16:30-17:00 「アナトモグラフィ/BodyParts3D の利用法」

17:00-17:45 「Genome Annotation と KazusaAnnotation」

17:45-18:00 「おわりに」

④統合データベース講習会：AJACS 蝦夷 2

日時：平成 21 年 9 月 7 日（月）09:45-17:00、8 日（火）10:00-15:00

場所：北海道大学札幌キャンパス

参加者：1 日目 20 名、2 日目 11 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

1 日目：ライフサイエンス統合データベースセンターが提供するサービスの使い方

9:45-10:05 「はじめに：統合データベースプロジェクトとは？」

10:05-10:15 「統合 TV」

10:15-11:15 「生命科学横断検索の利用法」

11:15-12:00 「DNA データベースの使い方」

12:00-13:30 昼休憩

13:30-14:00 招待講演：瀬々潤（お茶の水女子大学）「データベースからの生命科学：タンパク質相互作用と遺伝子発現を例に」

14:00-14:45 「遺伝子発現データベースを使い倒す」

14:45-15:15 休憩

15:15-16:00 「アナトモグラフィ/BodyParts3D の利用法」

16:00-16:45 「Genome Annotation と KazusaAnnotation」

16:45-17:00 「おわりに」

2 日目：一歩進んだバイオデータベース・ツールの使い方（中上級者向け）

10:00-12:00

続・遺伝子発現データベースを使い倒す

次世代シーケンサーから得られる遺伝子配列データの解析手法：配列データとしての発現データを扱う解析パイプラインの開発から

12:00-13:00 昼休憩

13:00-15:00

TogoWS の使い方

Galaxy を使った解析ワークフローの作り方

他、質問など

⑤統合データベース講習会：AJACS 近江 2

日時：平成 21 年 10 月 16 日（金）13:30-16:50

場所：長浜バイオ大学

参加者：43 名

対象：長浜バイオ大学生命情報コース配属の 3 回生及び一般参加者

内容：

13:30-13:45 「はじめに」

13:45-14:00 「統合 TV」

14:00-14:45 「生命科学横断検索の利用法」

14:45-15:15 「遺伝子配列・発現データベースの利用法」

15:15-15:25 休憩

15:25-15:45 招待講演

15:45-16:15 「アナトモグラフィ/BodyParts3D の利用法」

16:15-16:45 「Genome Annotation と KazusaAnnotation」

16:45-16:50 「おわりに」

⑥統合データベース講習会：AJACS りんくう

日時：平成 21 年 11 月 6 日（金）09:45-17:00

場所：大阪府立大学りんくうキャンパス

参加者：33 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

9:45-10:05 「はじめに：統合データベースプロジェクトとは？」

10:05-10:15 「統合 TV」
10:15-11:15 「生命科学横断検索の利用法」、番外編/生物アイコン
11:15-11:45 特別講演「科学データは誰のものか」国立遺伝学研究所 大久保公策 先生
11:45-13:00 昼休憩
13:00-13:30 招待講演「FANTOM4 プロジェクトに見る次世代シーケンサの可能性とデータベース」理化学研究所 オミックス基盤領域 川路英哉 先生
13:30-14:00 「DNA データベースの使い方」
14:00-14:45 「遺伝子発現データベースを使い倒す」
14:45-15:15 休憩
15:15-16:00 「アナトモグラフィー/BodyParts3D の利用法」
16:00-16:45 「Genome Annotation と KazusaAnnotation 」
16:45-17:00 「おわりに」

⑦統合データベース講習会：AJACS 筑後

日時：平成 22 年 1 月 18 日（月）09:30-17:30、19 日（火）10:00-15:00

場所：久留米ビジネスプラザ

参加者：1 日目 20 名、2 日目 12 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

1 日目：ライフサイエンス統合データベースセンターが提供するサービスの使い方

9:30-10:00 「はじめに：統合データベースプロジェクトとは？」

10:00-10:30 「統合 TV」

10:30-12:00 「生命科学横断検索、生命科学系 DB カタログの利用法」

「概要」(30mins)

「実習」(1hr)

12:00-13:00 昼休憩

13:00-15:00 「遺伝子発現データベースを使い倒す」

15:00-15:30 休憩

15:30-17:00 「DNA データベース総覧と検索、MiGAP、アナトモグラフィー／BodyParts3D の使い方」

2 日目：一歩進んだバイオデータベース・ツールの使い方（中上級者向け）

10:00-12:00 「遺伝子発現データを使い倒す～R/BioConductor の導入と解析例～」

12:00-13:00 昼休憩

13:00-15:00 午後の部

Reactome を使い倒す http://lifesciencedb.jp/image/small_video_icon.png
Biomart を使ってマイクロアレイ probeID の対応表を作る
Biomart を使って遺伝子上流配列を取得する
Galaxy を使って転写因子予測結合領域と遺伝子上流領域の「交差点」をリストアップする
その他

(2) BioHackathon2010 (情報・システム研究機構)

統合検索に向けた基盤技術の調査と開発を目的とし、国内外の研究者を招聘等した国際開発会議 (BioHackathon 2010) を開催した。統合検索は、統合化されたデータの上に構築される高度なデータ検索と閲覧支援によって構成され実現するものである。内外の研究者とともに、医学生物学分野における、いま現在の最先端のデータ統合技術やその応用についての調査検討をおこない、当センターが今後の統合検索実現化に向けて必要なリソースや技術、開発するべきことについての情報を得ることができた。初日の 2 月 8 日は独立行政法人産業技術総合研究所生命情報工学研究センターでシンポジウムを開催した。プログラム等スケジュールは下記の通りである。

2/8 (Mon)

Symposium 10:00-18:00

Program

Open 9:00

Talks 10:00-12:00

10:00-10:05 Toshihisa Takagi (DBCLS/DDBJ)

Opening remarks [video](#)

10:05-10:10 Toshiaki Katayama (HGC, University of Tokyo)

Introduction to the BioHackathon 2010 for Semantic Web

10:10-10:40 Erick Antezana

[Towards a Semantic Systems Biology: Biological Knowledge Management Using Semantic Web Technologies](#)

10:40-11:00 Matthias Samwald

[High-level knowledge representation on the Semantic Web: the Concept Web Alliance and related efforts](#)

11:00-11:20 Thomas Kappler and Jerven Bolleman (Swiss Institute of Bioinformatics)

[UniProt in RDF](#)

11:20-11:40 François Belleau (Bio2RDF)

[Bio2RDF cognoscope : A killer app for the life science.](#)

11:40-12:00 Heiko Horn (Novo Nordisk Foundation Centre for Protein Research)

Reflect - text mining in Semantic Web

Lunch break

Talks 13:30-14:50

13:30-13:50 Tetsuro Toyoda (RIKEN)

RIKEN SciNeS

13:50-14:10 Mark Wilkinson (Genome Canada Bioinformatics Platform)

SADI - Semantic Web Services

14:10-14:30 Andrea Splendiani (Rothamsted Research)

Visualization and analysis of biological networks on the Semantic Web

14:30-14:50 Mitsuteru Nakao (DBCLS) and Toshiaki Katayama (HGC, University of Tokyo)

Extending TogoWS and Open Bio* libraries for Linked Data

Break

Poster session 15:00-16:30

All invited participants

Open space discussion 16:30-18:00

Reception 19:00-21:00/2/9 - 2/11 9:00-18:00

Day 1 (Tuesday)

• The first activity of the group was to watch Tim Berner Lee presentation about Linked Data at TED The whole group also shouted RAW DATA NOW for the the life science.

• RDF primer lecture by François Belleau

- SadiDayOne
- TextMiningDayOne
- OpenBio
- (ongoing: FAQ)

Day 2 (Wednesday)

• SPARQL primer lecture by François Belleau

- TextMiningDayTwo
- 16:30 Data provider session
- NextGenerationSequencing
- (ongoing: FAQ)

Day 3 (Thursday)

- Predicate
- DDBJ-KEGG-PDBi

- [TextMiningDayThree](#)

- (ongoing: [FAQ](#))

Day 4 (Friday)

- [TextMiningDayFour](#)

- (ongoing: [FAQ](#))

Hacks

- [G-languageProject](#)

- [SystemsBiologySBML](#)

- [VirutuosoQuickPrimer](#)

- [VisualizationAndInteraction](#)

- [GFF3toRDF](#)

Hack, hack, hack

2/12 9:00-18:00

Hack, hack, hack 9:00-16:00

SummarySession 16:00-18:00

Farewell party 18:00-

(3) バイオインフォマティクスへの招待(お茶の水女子大学)

お茶の水女子大学で DB 高度利用者養成の一環として行ったセミナー「バイオインフォマティクスへの招待」の一覧を以下に示す。

第 19 回 バイオインフォマティクスへの招待

日時：平成 21 年 6 月 25 日

場所：お茶の水女子大学

参加人数：47 人

タイトル：分子シミュレーションで探るバイオの世界：蛋白質から生体超分子
まで

発表者：北尾 彰朗

第 20 回 バイオインフォマティクスへの招待

日時：平成 21 年 7 月 16 日

場所：お茶の水女子大学

参加人数：25 人

タイトル：世界初のデータベース総合インキュベーションセンターと女性人材の
活躍

発表者：豊田 哲郎

第 21 回 バイオインフォマティクスへの招待

日時：平成 21 年 10 月 15 日

場所：お茶の水女子大学

参加人数：40 人

タイトル：ゲノム解析から見えてきたヒトの進化

発表者：颯田 葉子

第 22 回 バイオインフォマティクスへの招待

日時：平成 21 年 11 月 19 日

場所：お茶の水女子大学

参加人数：28 人

タイトル：Functional Movements in Proteins –from Loops to Domains

発表者：Steven Hayward

第 23 回 バイオインフォマティクスへの招待

日時：平成 22 年 1 月 21 日

場所：お茶の水女子大学

参加人数：24 人

タイトル：生命科学・生化学に活用できる機械学習技術

発表者：加藤 毅

(4) アノテータ、キュレータ養成のための運営委員会と公開イベント (長浜バイオ大学)

長浜バイオ大では、学部レベルでのアノテータ・キュレータの人材養成を効果的に行うために、外部の専門家を含んだ運営委員会を独自に構成している。この運営委員会は、シニア世代の研究者を中心メンバーとしており、学部学生の作成したデータベースの精査をも実施し、品質の高いデータベース構築のための実践的な活動を、学部生との共同作業として実施している。人材育成の実施に積極的に関与しながら、問題的の具体的な洗い出しや評価、取り組むべき課題の方向性を検討した。

本運営委員会は、本年度は以下の日程で開催した。なお、学部学生との共同作業によるデータベース構築の実践例として、「エキスパートのキュレートした tRNA 遺伝子データベース」の精査と構築をおこなった。シニア世代によるデータベースの精査とキュレーションは、日常的にはインターネットを活用して分散型で行い、問題点や取り組むべき課題の洗い出しの検討や評価は長浜バイオ大で開催する会議で行った。シニア世代専門家の精査・キュレーション後のデータベースは、長浜バイオ大学、及び情報・システム研究機構統合 DB センターより公開している。

① 2009.5.28-30

参加者：池村淑道、阿部貴志、上原啓史、井口八郎、武藤あきら、山田優子、金谷重彦、学部学生 (1 名)

内容：tRNA 遺伝子精査結果判定会議

② 2009.11.4-6

参加者：池村淑道、阿部貴志、上原啓史、井口八郎、武藤あきら、山田優子、金谷重彦、渡辺公綱、学部学生（1名）

内容：tRNA 遺伝子精査結果判定会議、RNA に関する勉強会（講師：渡辺公綱）

③ 2010.3.22-24

参加者：池村淑道、阿部貴志、上原啓史、井口八郎、武藤あきら、山田優子、学部学生（1名）

内容：2009 年度更新用 WGS データ、酵母 tRNA 遺伝子の精査内容確認

また、シニア世代研究者を中心とした上記運営委員会に加えて、バイオ分野データベースに関する産業界からの視点を獲得の目的で、バイオ産業分野専門家との委員会を下記の日程にて開催した。

① 2009.11.10

参加者：池村淑道、阿部貴志、上原啓史、大橋武久、伊藤菁菫

内容：産業利用の高度な知識を付加したデータベースのキュレータ教育に係わる調整

② 2010.3.19

参加者：池村淑道、阿部貴志、上原啓史、大橋武久、伊藤菁菫

内容：産業利用の高度な知識を付加したデータベースのキュレータ教育に係わる調整

アノテータ・キュレータの教育実施体制

① アノテータ・キュレータの教育

(a) 担当教員と研究者：池村淑道、阿部貴志、大島一彦、齋藤美保子、上原啓史

内容：ゲノム情報解析法のバイオ実課題への応用「環境由来メタゲノム配列からの健康に貢献する遺伝子発掘」ならびにデータベース構築

2009.5.28 - 7.9 学部 3 回生(250 名)

(b) 担当教員と研究者：池村淑道、白井剛、大島一彦、高橋健一、依田隆夫、塩生真史、阿部貴志、齋藤美保子、上原啓史

内容：実習事業「バイオ分野のデータベースの統合的な利用」

2009.10.5 - 11.30 学部 2 回生(270 名)

(c) 担当教員と研究者：池村淑道、白井剛、大島一彦、高橋健一、依田隆夫、塩生真史、阿部貴志、齋藤美保子、上原啓史

内容：微生物ゲノムアノテーション実習

2009.11.12 - 1.21 学部 3 回生(50 名)

② 自己組織化マップによる大量ゲノム配列データへのアノテーション手法を用いた人材養成

担当教員と研究者：池村淑道、阿部貴志、上原啓史

2009.4.15-2010.3.19 学部 4 回生(1 名)

- ③ シニア研究者のチームによる分散型の高度なキュレーションシステムの改良・運用
担当教員と研究者：池村淑道、阿部貴志、上原啓史、井口八郎、武藤あきら、山田優子、
金谷重彦、木ノ内誠

2009.4.1-2010.3.31 学部 4 回生(1 名)

6. 成果の外部への発表

論文、学会発表等の発表内容については、添付の様式第 21 (78 ページ以降) を参照されたい。また、新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
文科省、バイオライフ DB を JST 事業と一本化	化学工業日報	H22.1.28
記者発表、情報・システム研究機構、 H20 年度「統合データベースプロジェクト」事業成果に対する ユーザー評価の結果	Biotechnology Japan (日経 BP)	H22.2.3

7. 特許出願状況

該当なし。

8. 実施体制

別表 1、2、3 (98 ページ以降) を参照されたい。

9. 整備実績一覧

別紙 1 (101 ページ以降) を参照されたい。

様式第 2 1

学 会 等 発 表 実 績

委託業務題目: 「ライフサイエンス統合データベース開発運用」(戦略立案・実行評価/統合データベース開発/統合データベース支援)

機 関 名: 大学共同利用機関法人 情報・システム研究機構、独立行政法人 産業技術総合研究所生命情報工学研究センター
財団法人 かずさディー・エヌ・エー研究所、国立大学法人 九州大学、国立大学法人 奈良先端科学技術大学院大学
国立大学法人 東京大学、国立大学法人 お茶の水女子大学、学校法人 関西文理総合学園長浜バイオ大学

1. 学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
大学共同利用機関法人情報・システム研究機構				
生命科学系データベースアーカイブ[口頭]	畠中 秀樹	日本農芸化学会 2010 年度大会 ランチ オンセミナー「情報整理術 2010 ～文 献とデータを賢く管理する～」	2010 年 3 月 27-30 日	国内（東京）
内閣府統合データベース事業と理研への期待[口頭]	高木 利久、大石道夫	第 18 回理事長ファンドワークショップ 「理研総合データベースを核にした研 究基盤連携網の構築」	2010 年 3 月 12-13 日	国内（葉山）
内閣府統合データベース事業と理研への期待、文科 省ライフサイエンス統合データベースプロジェクト の目的、現状、課題[口頭]	高木 利久	第 18 回理事長ファンドワークショップ 「理研総合データベースを核にした研 究基盤連携網の構築」	2010 年 3 月 12-13 日	国内（葉山）
国際開発会議DBCLSバイオハッカソン2010レポート [口頭]	中尾 光輝	人工知能学会第二種研究会 分子生物情 報研究会（SIG-MBI）	2010 年 3 月 5-6 日	国内（能美）
BIOLOGICAL DOCUMENT CLUSTERING USING A GENE-GENE NETWORK[ポスター]	Hong-Woo Chun, Shinobu Okamoto, Yasunori Yamamoto, Atsuko Yamaguchi, Toshihisa Takagi	PSB(Pacific Symposium on Biocomputing) 2010	2010 年 1 月 4-8 日	国外（Hawaii）

DDBJ Read Archive and DDBJ Read Annotation Pipeline: An Archive Database and an Analytical Tool for Next-Generation Sequence Data T003/P120[ポスター／口頭]	Eli Kaminuma, Yuichi Kodama, Satoshi Saruhashi, Takeshi Konno, Takako Mochizuki, Hidemasa Bono, Hideaki Sugawara, Kousaku Okubo, Toshihisa Takagi, Yasukazu Nakamura	The 20th International Conference on Genome Informatics (GIW2009)	2009年12月14-16日	国内（横浜）
なぜ今共有なのか？個人プレーから社会プレーへ[口頭]	大久保 公策	第32回日本分子生物学会年会、フォーラム「デジタル時代のジレンマ」”共有と秘匿のバランス”	2009年12月10日	国内（横浜）
データベース統合センターの役割[口頭]	高木 利久	第32回日本分子生物学会年会、フォーラム「デジタル時代のジレンマ」”共有と秘匿のバランス”	2009年12月10日	国内（横浜）
実験生物学者にとって有益な情報環境を提案する機会：「統合データベースプロジェクト」のユーザ評価[口頭]	高祖 歩美、箕輪 真理	第32回日本分子生物学会年会、ワークショップ	2009年12月10日	国内（横浜）
ブタ成熟脂肪細胞および顆粒層細胞における脱分化ならびに多能性獲得機構の統合トランスクリプトミクス[ポスター]、[口頭]	小野 浩雅、坊農 秀雅、加野 浩一郎	第32回日本分子生物学会年会、ワークショップ	2009年12月10日	国内（横浜）
かずさアノテーションを用いた分散型ゲノムアノテーションの実証実験[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、佐藤 修正、金子 貴一、中村 保一	第32回日本分子生物学会年会、ワークショップ	2009年12月10日	国内（横浜）
統合データベースプロジェクト：オープンで知のめぐりの良い分子生物学を目指して[口頭]	坊農 秀雅	第32回日本分子生物学会年会、ワークショップ	2009年12月10日	国内（横浜）
DBCLS で提供中の文献関係サービス[ポスター]	山本 泰智、山口敦子、坊農 秀雅、高木利久	第32回日本分子生物学会年会	2009年12月9-12日	国内（横浜）
KazusaMart: 植物および関連微生物のゲノムデータ統合[ポスター]	藤澤 貴智、岡本 忍、中尾 光輝、中村 保一	第32回日本分子生物学会年会	2009年12月9-12日	国内（横浜）
「DDBJ Read Archive」：次世代シーケンサからの出力データのためのアーカイブ[ポスター]	児玉 悠一、猿橋 智、五條 堀 孝、館野 義男、神沼 英里、中村 保一、菅原	第32回日本分子生物学会年会	2009年12月9-12日	国内（横浜）

次世代シーケンサーのハイスループット・データ解析システム「DDBJ Short Read Annotation Pipeline」の構築[ポスター]	神沼 英里、児玉 悠一、大久保 公策、高木 利久、中村 保一	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
微生物ゲノムアノテーションパイプライン MiGAP[ポスター]	大山 彰、森 宙史、黒川 顕、菅原 秀明	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
DBCLS Galaxy: 生物データ処理の対話的ツール組み合わせインターフェース[ポスター]	山口 敦子、全 弘宇、中尾 光輝、大内田 賢太、山本 泰智、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
生命科学研究者のための統合文献情報管理システム [ポスター]	岩崎 渉、山本 泰智、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
Representation of the multiple roles of proteins from anatomical perspective by data analysis[ポスター]	坂東 明日佳、中村 桂子、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
キーワードと蛋白 ID を用いた横断検索システム [ポスター]	畠中 秀樹、平井 信一、大野 忠、高橋 順子、牧口 大旭、奥村 利幸、熊谷 禎洋、八塚 茂、川本 祥子、西川 哲夫、中村 春木、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
生命情報データベースの横断検索システム: Web-API サービスの開発[ポスター]	牧口 大旭、高橋 順子、川崎 恭実子、川本 祥子、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
データバンク (INSD, GEO) 目次システムの開発 [ポスター]	小笠原 理、渡邊 康司、栗原 英輔、森山 丈士、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
BodyParts3D と Anatomography: 3次元人体構造データベースと人体解剖図作成サービス[ポスター]	三橋 信孝、藤枝 香、田村 卓郎、川本 祥子、高木 利久、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）

生命科学系データベースアーカイブサービス[ポスター]	西川 哲夫、八塚 茂、熊谷 禎洋、畠中 秀樹、片山 俊明、川本 祥子、大久保 公策、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内 (横浜)
ライフサイエンス統合データベース : アップデート 2009[ポスター]	河野 信、Lsdb 開発スタッフ、大久保 公策、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内 (横浜)
Functional analysis pipeline of transcript sequences as gene expression data[口頭]	Hidemasa Bono	第 7 回がんとハイポキシア研究会	2009 年 12 月 5-6 日	国内 (京都)
GuideLink: A Corpus Annotation System that Integrates the Management of Annotation Guidelines[ポスター]	Kenta Oouchida, Jin-Dong Kim, Toshihisa Takagi and Jun' ichi Tsujii	The 23rd Pacific Asia Conference on Language, Information and Computation	2009 年 12 月 3-5 日	国内 (香港)
デジタル科学における独占と共有の賢いバランス [口頭]	大久保 公策	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内 (お台場)
統合データベースプロジェクトのサービス紹介~複数のデータベースを繋ぐツールたち - [口頭]	山口敦子、中尾 光輝	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内 (お台場)
ライフサイエンス統合データベースの課題 : 権利と法律、技術[口頭]	中尾 光輝	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内 (お台場)
「生命科学情報の共有を目指す” ライフサイエンス統合データベース” 」 [ポスター]	箕輪 真理、高祖 歩美、坂東 明日佳、川本 祥子	情報・システム研究機構シンポジウム 情報とシステム 2009	2009 年 11 月 30 日	国内 (品川)
Integrated Database Project for Life Science in Japan[口頭]	Toshihisa Takagi	BI01NF02009, CBI-KSBSB Joint International Conference	2009 年 11 月 4-6 日	国外 (釜山)
統合 DB プロジェクトのサービス紹介~データベースを役立てよう~[口頭]	川本 祥子	第 47 回日本生物物理学会年会 ランチオンセミナー「データベース統合の多面的展開」	2009 年 10 月 30 日-11 月 1 日	国内 (徳島)
多様なデータが広く活用されるために[口頭]	畠中 秀樹	第 47 回日本生物物理学会年会 ランチオンセミナー「データベース統合の多面的展開」	2009 年 10 月 30 日-11 月 1 日	国内 (徳島)

Systematic organization of gene expression data in Japan[口頭]	Hidemasa Bono, Eli Kaminuma, Yuichi Kodama, Yasukazu Nakamura, Kousaku Okubo, and Toshihisa Takagi	12th International MGED Meeting, focusing on translational genomics and high throughput sequencing	2009年10月5-8日	国外(Phoenix, Arizona)
「デジタル科学への最後のステップ」[口頭]	大久保公策	シンポジウム「科学技術における情報の上手な権利化と共有化」	2009年10月5日	国内(東大弥生)
ライフサイエンス統合データベースプロジェクト[口頭]	高木 利久	日本人類遺伝学会 第54回大会 ワークショップ「ゲノムインフマティクス」	2009年9月26日	国内(品川)
ライフサイエンス分野の統合データベースの学部や高校教育における活用[口頭]	坊農 秀雅	日本遺伝学会第81回大会 ミニ・シンポジウム「自分が興味を持つ遺伝子がどの範囲の環境微生物に存在するのかを探索する方法:環境微生物ゲノム配列からのお宝遺伝子発掘の学部や高校教育における活用」	2009年9月16日	国内(長野)
統合データベースプロジェクト:オープンで知のめぐりの良いライフサイエンス研究を目指して[口頭]	坊農秀雅	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東大本郷)
BodyParts3D/Anatomography: 3次元人体構造DBと人体モデル図作成ツール[ポスター]	三橋 信孝	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東大本郷)
Gene Indexing:植物と関連微生物におけるオープンゲノム注釈実証実験[ポスター]	岡本 忍	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東大本郷)
生物種ごとの日本の分子生物学研究一覧の構築に向けて ~「生物アイコン」の現状と今後[ポスター]	坂東 明日佳	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東大本郷)

生命科学系データベースアーカイブサービス[ポスター]	西川 哲夫	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
DBCLS Galaxy: 生物データ処理の対話的ツール組み合わせインターフェース[ポスター]	山口 敦子	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
発現統合：配列データとしての発現データを扱う解析パイプラインの開発[ポスター]	坊農 秀雅	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
BioHackathon 2009 レポート[ポスター]	中尾 光輝	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
MeSH キーワード集による疾患の特徴プロファイリング[ポスター]	仲里 猛留	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
テキスト処理して情報整理[ポスター]	山本 泰智	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
統合データベース講習会 AJACS[ポスター]	河野 信	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）

統合 TV-動画によるバイオデータベースとウェブツールの活用法解説サイト[ポスター]	小野 浩雅	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
MeSH キーワード集による疾患の特徴プロファイリング[ポスター]	仲里猛留、坊農秀雅、松田秀雄、高木利久	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
統合データベースポータル「統合ホームページ」 lifesciencedb.jp[ポスター]	仲里 猛留、坂東 明日佳、坊農 秀雅、全 弘宇、藤枝香、畠中 秀樹、平井 信一、川本 祥子、河野 信、高祖歩美、熊谷 禎洋、箕輪 真理、三橋 信孝、永井 啓一、中尾 光輝、西川 哲夫、岡本 忍、小野 浩雅、大内田賢太、谷口 仁、山口 敦子、山本 泰智、八塚 茂、大久保 公策、高木 利久	第9回日本分子生物学会春季シンポジウム	2009年5月12-13日	国内（宮崎）
独立行政法人産業技術総合研究所生命情報工学研究センター				
KNIME ガイド：CBRC KINIME 活用	野口 保	共同研究会「生物情報資源の相互運用性」	2010年3月24日	国内（東京）
情報統合基盤技術によるアクティブ・ワークフロー解析[ポスター]	福井 一彦、田代 俊行、矢葺 幸光、野口 保、浅井 潔	平成21年度 産総研・産技連 LS-BT 合同研究発表会	2010年2月4日	国内（つくば）
情報統合基盤技術によるワークフロー解析[口頭]	福井 一彦、田代 俊行、矢葺 幸光、堀本 勝久、野口保、浅井 潔	産総研 生命情報工学研究センター2009年度研究報告シンポジウム	2009年12月3日	国内（東京）
ライフサイエンス統合 DB プロジェクトワークフロー[ポスター（デモンストレーション）]	田代 俊行、矢葺 幸光	産総研 生命情報工学研究センター2009年度研究報告シンポジウム	2009年12月3日	国内（東京）

ライフサイエンス統合 DB プロジェクト ワークフロー[ポスター]	田代 俊行、矢葺 幸光、福井 一彦、野口 保、浅井 潔	産総研 生命情報工学研究センター2009年度研究報告シンポジウム	2009年12月3日-4日	国内（東京）
財団法人かずさディー・エヌ・エー研究所				
ゲノム情報のアノテーションと可視化[口頭]	岡本 忍	第2回若手研究者による知の創出のための異分野共同セミナー	2010年3月28-29日	国内（札幌）
微生物ゲノム基盤情報の高度化と統合のためのシステム[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、佐藤 修正、金子 貴一、中村 保一	第4回日本ゲノム微生物学会年会	2010年3月7-9日	国内（福岡）
ゲノムアノテーション情報の高度化と統合化への取り組み[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、佐藤 修正、金子 貴一、中村 保一	かずさ研究交流セミナー	2010年1月27日	国内（千葉）
KazusaMart:the integration of genomic data in plants and the related microorganisms	藤澤貴智、岡本忍、中尾光輝、中村保一	第32回日本分子生物学会	2009年12月10日	国内（横浜）
シアノバクテリアのゲノム情報基盤[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、金子 貴一、中村 保一	ラン藻ワークショップ2009	2009年12月4-5日	国内（千葉）
RhizoBase の新機能[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、佐藤 修正、金子 貴一、中村 保一	第19回植物微生物研究会	2009年9月8-10日	国内（長野）
KazusaMart: RhizoBase のデータ統合と管理	藤澤貴智、岡本忍、中尾光輝、金子貴一、佐藤修正、中村保一	第19回植物微生物研究会	2009年9月8-10日	国内（長野）
新しいCyanoBase の使い方[口頭]	岡本 忍	ラン藻ゲノム研究交流会	2009年7月4日	国内（東京）
ラン藻のゲノム解析とその支援ツール KazusaMart	藤澤貴智	ラン藻ゲノム研究交流会	2009年7月4日	国内（東京）

Gene Indexing: 植物と関連微生物におけるオープンゲノム注釈実証実験[ポスター]	岡本 忍、 中尾 光輝、 藤金 沢 貴智、 佐藤 修正、 子 貴一、 中村 保一	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東京)
KazusaMart: 植物関連生物ゲノムデータの統合と管理	藤澤貴智、中尾光輝、岡本忍、中村保一	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東京)
国立大学法人九州大学				
Identification of loci for systemic lupus erythematosus by pooling-based genome-wide association study. [ポスター]	Tahira T, Masumoto K, Kukita Y, Okazaki Y, Yoshinaga A, Higasa K, Horiuchi T, Hayashi K	58th Annual Meeting of the American Society of Human Genetics, Honolulu, U. S. A.	2009年10月20日-25日	国外(Hawaii)
Definitive SNP/CNV haplotyping of Asian genomes using DNAs derived from complete hydatidiform moles. [ポスター]	Kenshi Hayashi, Yoji Kukita, Koichiro Higasa, Tomoko Tahira, Koji Yahara, Kiyoko Kato, Norio Wake	58th Annual Meeting of American Society of Human Genetics. Honolulu, U. S. A.	2009年10月20日-25日	国外(Hawaii)
SNP/CNV haplotype map determined using complete hydatidiform moles carrying haploid genomes. [ポスター]	Yoji Kukita, Koichiro Higasa, Tomoko Tahira, Kiyoko Kato, Norio Wake, Kenshi Hayashi	Personal Genome, Cold Spring Harbor Laboratory, New York, U. S. A.	2009年9月14日-17日	国外(NY)
国立大学法人奈良先端科学技術大学院大学				
述語語義と意味役割の結合学習のための構造予測モデル[口頭]	渡邊陽太郎, 浅原正幸, 松本裕治	言語処理学会第16回年次大会	2010年3月	国内(東京)
A Graph-based Approach for Biomedical Thesaurus Expansion[口頭]	Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, Yuji Matsumoto	ACM Third International Workshop on Data and Text Mining in Bioinformatics	2009年11月	国外(香港)

Coordinate Structure Analysis with Global Structural Constraints and Alignment-Based Local Features[口頭]	Kazuo Hara, Masashi Shimbo, Hideharu Okuma, Yuji Matsumoto	Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing	2009年8月	国外 (Singapore)
Bypassed Alignment Graph for Learning Coordination in Japanese Sentences[口頭]	Hideharu Okuma, Kazuo Hara, Masashi Shimbo, Yuji Matsumoto	Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing	2009年8月	国外 (Singapore)
オンライン最大マージン学習アルゴリズムに基づく多言語依存構造-述語項構造解析[口頭]	渡邊陽太郎, 浅原正幸, 松本裕治	情報処理学会自然言語処理研究会	2009年7月	国内 (北海道)
Multilingual Syntactic-Semantic Dependency Parsing with Three-Stage Approximate Max-Margin Linear Models[口頭]	Yotaro Watanabe, Masayuki Asahara, Yuji Matsumoto	Thirteenth Conference on Computational Natural Language Learning	2009年6月	国外 (Boulder, USA)
国立大学法人東京大学				
バイオデータベース構築者養成コースの紹介[ポスター]	中谷洋一郎	文部科学省「統合データベースプロジェクト」シンポジウム「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内(東大本郷)
国立大学法人お茶の水女子大学				
A course for biological knowledge discovery by handling genome-wide data(ポスター)	金子 聡子	第32回日本分子生物学会年会	2009年12月9-12日	国内(横浜)
A Course for Biological Knowledge Discovery by Handling Genome-Wide Data[ポスター]	金子 聡子	Bioinformatics beyond Omics Data Analyses	2009年12月3-4日	国内(東京)
情報量が増え続けるデータベースを使いこなすために: データベースからの知識発見[ポスター]	金子 聡子	第82回日本生化学会大会	2009年10月21-24日	国内(神戸)

Makorin1 遺伝子の mRNA 量制御機構についての予測 [口頭]	金子 聡子	日本遺伝学会第 81 回大会	2009 年 9 月 16-18 日	国内 (長野)
データベースからの生命科学：タンパク質相互作用 と遺伝子発現を例に[口頭]	瀬々 潤	統合データベース講習会：AJACS 蝦夷 2	2009 年 9 月 7-8 日	国内 (札幌)
ヒト特異的に発現量が低下している Carboxylesterase1 遺伝子群の進化[ポスター]	金子 聡子	第 11 回日本進化学会大会	2009 年 9 月 2-4 日	国内 (札幌)
Evolutionary dynamics of “functional” processed pseudogenes that stabilize Makorin1 mRNAs in mammals (ポスター)	金子 聡子	ESEB 12th Congress	2009 年 8 月 24-29 日	国外 (Italy)
DB 高度利用者の養成：情報量が増え続けるデータベ ースを使いこなすために[ポスター]	金子 聡子	文部科学省委託研究開発事業 統合デー タベースプロジェクト シンポジウム 2009 「データベースが拓 くこれからのライフサイエンス」	2009 年 6 月 12 日	国内 (東京)
学校法人関西文理総合学園長浜バイオ大学				
学部教育としての健康への貢献遺伝子データベース ～抗生物質生産遺伝子を中心とした遺伝子探索～ (URL: http://dbcls.nagahama-i-bio.ac.jp/) [ポ スター]	上原啓史、阿部貴志、中泉 友紀、和田千恵子、井口八 郎、池村淑道	第 4 回日本ゲノム微生物学会年会	2010 年 3 月 7 日-9 日	国内 (福岡)
tRNADB-CE を用いたバクテリアの tRNA 遺伝子構造の 解析[ポスター]	三崎優香、阿部貴志、上原 啓史、池村淑道、井口八郎	第 4 回日本ゲノム微生物学会年会	2010 年 3 月 7-9 日	国内 (福岡)
オリゴペプチド組成に基づく一括学習型自己組織化 マップ (BLSOM) 法を用いた機能未知のタンパク質類 の機能推定法の確率[ポスター]	阿部貴志、上原啓史、金谷 重彦、池村淑道	第 4 回日本ゲノム微生物学会年会	2010 年 3 月 7-9 日	国内 (福岡)
学部教育としての持続可能型社会への貢献遺伝子デ ータベース構築[ポスター]	上原啓史、阿部貴志、池村 淑道	第 10 回グリーン・サステイナブル・ケ ミストリー (GSC) シンポジウム	2010 年 3 月 4-5 日	国内 (東京)

tRNADB-CE: tRNA gene database curated manually by experts (http://trna.nagahama-i-bio.ac.jp/cgi-bin/trnadb/index.cgi) [口頭]	T. Abe, T. Ikemura, Y. Ohara, H. Uehara, M. Kinouchi, J. Sugahara, A. Kanai, S. Kanaya, Y. Yamada, A. Muto, H. Inokuchi	23rd tRNA workshop (Aveiro, Portugal)	2010年1月28日-2月2日	国外 (Portugal)
Unveiling microbial diversity of uncultured environmental microbe mixtures and protein functions of metagenome sequences using Batch Learning Self-Organizing Map (BLSOM) [口頭]	T. Abe, S. Kanaya, T. Ikemura	International Symposium on Marine Genomics 2009	2009年12月15-18日	国内 (沖縄)
A型インフルエンザウイルスの全ゲノム配列を対象とした一括学習型自己組織化マップ法 (BL-SOM) による効率的な知識発見とその応用 [ポスター]	岩崎裕貴、阿部貴志、伊藤正恵、和田健之介、池村淑道	第32回日本分子生物学会年会	2009年12月11日	国内 (横浜)
学部生とシニア世代の共同作業としての“健康への貢献遺伝子データベース”構築 (URL: http://dbcls.nagahama-i-bio.ac.jp/) [ポスター]	上原啓史、阿部貴志、中泉友紀、和田千恵子、井口八郎、池村淑道	第32回日本分子生物学会年会	2009年12月11日	国内 (横浜)
tRNADB-CE: エキスパートがキュレートした tRNA 遺伝子データベース [ポスター]	阿部貴志、池村淑道、小原康雄、上原啓史、平野晋也、木ノ内誠、金谷重彦、菅原潤一、金井明夫、山田優子、武藤あきら、井口八郎	第32回日本分子生物学会年会	2009年12月11日	国内 (横浜)
データベースに蓄積の著しい機能未知のタンパク質類の機能推定のための一括学習型自己組織化マップ法 [ポスター]	阿部貴志、金谷重彦、池村淑道	第32回日本分子生物学会年会	2009年12月10日	国内 (横浜)
「生命の謎にコンピュータで挑戦」 [口頭]	上原啓史 (池村淑道)	ひらめき☆ときめきサイエンス	2009年10月31日	国内 (長浜)
学部教育としての“健康”ならびに“持続可能型社会”への貢献遺伝子データベース構築 (http://dbcls.nagahama-i-bio.ac.jp/) [ポスター]	上原啓史、阿部貴志、中泉友紀、和田千恵子、井口八郎、池村淑道	第82回日本生化学会大会	2009年10月23日	国内 (神戸)

「多様な環境微生物からの“健康ならびに持続可能型社会への貢献遺伝子”の発掘とデータベース構築」 「高性能スパコンによる先端的大規模ゲノム解析」 [展示]	池村淑道	びわ湖ビジネス環境メッセ	2009年10月21日-23日	国内（長浜）
全インフルエンザAウイルスを対象にした新規情報学的手法による俯瞰的可視化とそこからの知識発見 [口頭]	岩崎裕貴、阿部貴志、伊藤正恵、池村淑道	日本遺伝学会 第81回大会	2009年9月18日	国内（長野）
オリゴヌクレオチド組成を用いた一括学習型自己組織化地図法によるA型インフルエンザウイルスの俯瞰的な特徴解明[口頭]	岩崎裕貴、池村淑道、伊藤正恵、阿部貴志	第18回バイオ情報学研究会	2009年9月17日	国内（札幌）
健康や持続可能型社会への貢献遺伝子の発掘とその学部や高校教育における活用[口頭]	上原啓史、阿部貴志、中泉友紀、和田千恵子、井口八郎、池村淑道	日本遺伝学会 第81回大会	2009年9月16日	国内（長野）
学部生シニア世代の共同作業としての「高精度・高性能な遺伝子データベース」の構築とその利用[口頭]	阿部貴志、小原康雄、上原啓史、平野晋也、木ノ内誠、金谷重彦、菅原潤一、金井明夫、山田優子、武藤昱、井口八郎、池村淑道	日本遺伝学会 第81回大会	2009年9月16日	国内（長野）
学部生とシニア世代の共同作業としての知識発見とそのデータベース化[口頭]	池村淑道	日本遺伝学会 第81回大会	2009年9月16日	国内（長野）
全地球レベルの環境微生物多様性を把握・俯瞰するための新規情報学手法の開発[口頭]	阿部貴志、金谷重彦、池村淑道	第11回日本進化学会大会	2009年9月2-4日	国内（札幌）
「生命の謎にコンピュータで挑戦」[口頭]	上原啓史 (池村淑道)	ひらめき☆ときめきサイエンス	2009年8月30日	国内（長浜）
tRNADB-CE: エキスパートがキュレートした tRNA 遺伝子データベース[口頭]	阿部貴志、池村淑道、小原康雄、上原啓史、平野晋也、木ノ内誠、金谷重彦、菅原潤一、金井明夫、山田優子、武藤昱、井口八郎	RNA学会第11回大会	2009年7月27-29日	国内（新潟）

Unveiling microbial diversity of uncultured environmental microbe mixtures on the basis of Batch Learning Self-Organizing Map (BLSOM) [ポスター]	T. Abe, S. Kanaya, T. Ikemura	Xth SCAR International Biology Symposium	2009年7月26-31日	国内（札幌）
エキスパートがキュレートした tRNA 遺伝子データベース [ポスター]	阿部貴志、小原康雄、上原啓史、平野晋也、木ノ内誠、金谷重彦、菅原潤一、金井明夫、山田優子、武藤昱、井口八郎、池村淑道	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）
学部教育としての持続可能型社会ならびに健康への貢献遺伝子データベース構築 [ポスター]	上原啓史、阿部貴志、中泉友紀、和田千恵子、井口八郎、池村淑道	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009年6月12日	国内（東大本郷）

2. 学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期	国内・外の別
大学共同利用機関法人情報・システム研究機構				
GuideLink: A Corpus Annotation System that Integrates the Management of Annotation Guidelines	Kenta Oouchida, Jin-Dong Kim, Toshihisa Takagi and Jun'ichi Tsujii	Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation	2009年12月3-5日	国外
CyanoBase: the cyanobacteria genome database update 2010	Mitsuteru Nakao, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko and Yasukazu Nakamura	Nucleic Acids Research, 2010, Vol. 38, Database issue D379-D381	published online on 2009年10月30日	国外 (Oxford University Press)

生命科学分野におけるテキストマイニング	山本 泰智	情報処理学会誌 2009, Vol. 50, No. 9, 853-858	2009年9月	国内
生命情報学が直面する大規模ゲノムデータ時代の課題 : 1. 分散データの統合とセマンティック Web	中尾 光輝、片山 俊明	情報処理学会誌 2009, Vol. 50, No. 9, 836-844	2009年9月	国内
生命情報学が直面する大規模ゲノムデータ時代の課題 : 0. 編集にあたって	片山 俊明	情報処理学会誌 2009, Vol. 50, No. 9, 833-835	2009年9月	国内
Gendoo: Functional profiling of gene and disease features using MeSH vocabulary	Takeru Nakazato, Hidemasa Bono, Hideo Matsuda and Toshihisa Takagi	Nucleic Acids Research, 2009, Vol. 37, Web Server issue W166-W169	2009年6月	国外 (Oxford University Press)
ライフサイエンス統合データベースセンターと統合データベースプロジェクト	坊農 秀雅	情報の科学と技術, 59巻4号, 165-170	2009年4月	国内
国立大学法人奈良先端科学技術大学院大学				
述語語義と意味役割の結合学習のための構造予測モデル	渡邊陽太郎, 浅原正幸, 松本裕治	人工知能学会論文誌, Vol. 25, No. 2	2010年1月	国内
学校法人関西文理総合学園長浜バイオ大学				
Novel bioinformatics for inter- and intraspecies comparison of genome signatures in plant genomes	T. Abe, K. Wada, Y. Iwasaki, T. Ikemura	Plant Biotechnology, 26, 469-477	2009年	国外
A novel bioinformatics strategy for function prediction of poorly-characterized protein genes obtained from metagenome analyses	T. Abe, S. Kanaya, H. Uehara, T. Ikemura	DNA Research, 16, 287-298	2009年	国外
tRNADB-CE: tRNA gene database curated manually by experts	T. Abe, T. Ikemura, Y. Ohara, H. Uehara, M. Kinouchi, S. Kanaya, Y. Yamada, A. Muto and H. Inokuchi	Nucleic Acids Research, 37, D163-D168	2009年	国外

Batch-Learning Self-Organizing Map for Predicting Functions of Poorly-Characterized Proteins Massively Accumulated	T. Abe S. Kanaya, T. Ikemura	J.C. Príncipe and R. Miikkulainen (Eds.): WSOM 2009, LNCS 5629, pp. 1-9	2009 年	国外
A Large-Scale Genomics Studies Conducted with Batch-Learning SOM Utilizing High-Performance Supercomputers	T. Abe, Y. Hamano, S. Kanaya, K. Wada and T. Ikemura	J. Cabestany et al. (Eds.): IWANN 2009, Part I, LNCS 5517, pp. 829-836	2009 年	国外
環境由来大量 DNA 配列を利用した難培養性生物群の系統推定のための新規情報学手法	阿部貴志、上原啓史、金谷重彦、池村淑道	マリンメタゲノムの有効利用(松永是、竹山春子 監修), シーエムシー出版, p228-239	2009 年	国内
オリゴヌクレオチド組成を用いた一括型自己組織化地図法による A 型インフルエンザウイルスの俯瞰的な特徴解明	岩崎裕貴、池村淑道、伊藤正恵、阿部貴志	第 18 回バイオ情報学研究会、Vol. 2009-BIO-18 No5, 1-6.	2009 年	国内
国立大学法人九州大学				
Mutations in TSPAN12 cause autosomal-dominant familial exudative vitreoretinopathy.	Poulter JA, Ali M, Gilmour DF, Rice A, Kondo H, Hayashi K, Mackey DA, Kearns LS, Ruddle JB, Craig JE, Pierce EA, Downey LM, Mohamed MD, Markham AF, Inglehearn CF, Toomes C	Am J Hum Genet. 86(2):248-253.	2010 年	国外
Association of killer cell immunoglobulin-like receptor 2DL5 with systemic lupus erythematosus and accompanying infections.	Kimoto Y, Horiuchi T, Tsukamoto H, Kiyohara C, Mitoma H, Uchino A, Furugo I, Yoshizawa S, Ueda A, Harashima S, Sawabe T, Tahira T, Hayashi K, Yoshizawa S, Shimoda T, Akashi K, Harada M.	Rheumatology	in press	国外

Impact of group IVA cytosolic phospholipase A (2) gene polymorphisms on phenotypic features of patients with familial adenomatous polyposis.	Umeno J, Matsumoto T, Esaki M, Kukita Y, Tahira T, Yanaru-Fujisawa R, Nakamura S, Arima H, Hirahashi M, Hayashi K, Iida M	Int J Colorectal Dis.	in press	国外
Estimation of SNP allele frequencies by SSCP analysis of pooled DNA.	Tahira T, Kukita Y, Higasa K, Okazaki Y, Yoshinaga A, Hayashi K	Methods Mol Biol. 578:193-207.	2009 年	国外
Novel mutation in PAX3 gene in Waardenburg syndrome accompanied by unilateral macular degeneration.	Kozawa M, Kondo H, Tahira T, Hayashi K, Uchio E	Eye 23(7):1619-1621.	2009 年	国外
Evaluation of haplotype inference using definitive haplotype data obtained from complete hydatidiform moles, and its significance for the analyses of positively selected regions.	Higasa K, Kukita Y, Kato K, Wake N, Tahira T, Hayashi K	PLoS Genetics 5:e1000468.	2009 年	国外

別表1 平成21年度に於ける実施体制

研究項目	担当機関等	研究担当者
1. 戦略立案・実行評価	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 情報・システム研究機構 国立遺伝学研究所 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 同上 同上	◎○高木 利久 大久保 公策 永井 啓一 箕輪 真理 西川 哲夫 吉羽 洋周
2. 統合データベース開発 (1) 共通基盤技術開発 ① 共通基盤技術開発 ② ワークフロー技術を用いた統合 DB 環境構築 ③ 専門用語辞書管理システムと専門用語解析技術の開発	情報・システム研究機構 国立遺伝学研究所 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 同上 同上 同上 同上 同上 東京大学医科学研究所ヒトゲノム解析センター 同上 独立行政法人産業技術総合研究所 生命情報工学研究センター 同上 同上 同上 同上 同上 同上 同上 同上 同上 同上 奈良先端科学技術大学院大学情報科学研究科 同上 同上 同上 同上 同上 同上	○大久保 公策 高木 利久 山口 敦子 山本 泰智 中尾 光輝 全 弘宇 大内田 賢太 片山 俊明 川島 秀一 浅井 潔 野口 保 諏訪 牧子 光山 統泰 Horton Poul 広川 貴次 福井 一彦 藤渕 航 堀本 勝久 田代 俊行 松本 裕治 原 一夫 呂 嘉 新保 仁 渡邊 陽太郎 浅原 正幸 原(鈴木)郁美

注1. ◎:課題代表者、○:サブテーマ代表者

注2. 本業務に携わっている方は、全て記入。

別表2 平成21年度に於ける実施体制

研究項目	担当機関等	研究担当者
(2)ヒト統合DBの開発・運用	情報・システム研究機構 国立遺伝学研究所	大久保 公策
①ヒト統合DBの開発・運用	同上	同上
	同上	小笠原 理
	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	三橋 信孝
	同上	藤枝 香
	同上	坂東明日佳
	大阪府立成人病センター研究所	加藤 菊也
	九州大学生体防御医学研究所	林 健志
②多型知識表現技術の開発	情報・システム研究機構 国立遺伝学研究所	菅原 秀明
(3)モデル生物・産業応用生物統合DB	同上	同上
①モデル生物・産業応用生物統合DB	情報・システム研究機構 国立遺伝学研究所	大久保 公策
	東京工業大学大学院生命理工学研究科	黒川 顕
	自然科学研究機構 基礎生物学研究所	内山 郁夫
	農業生物資源研究所 基礎研究領域	伊藤 剛
②植物及び植物関連微生物のゲノム情報データベース統合と高度化	かずさDNA研究所植物ゲノム基盤研究部	中村 保一
	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	岡本 忍
	かずさDNA研究所植物ゲノム基盤研究部	中尾 光輝
	同上	藤澤 貴智
	同上	藤代 維一
	同上	小原 光代
3. 統合データベース支援	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	○川本 祥子
(1)ポータル整備・運用、広報、普及啓発	同上	同上
	同上	坊農 秀雅
	同上	河野 信
	同上	仲里 猛留
	同上	箕輪 真理
	同上	高祖 歩美
	同上	小野 浩雅
	同上	山口 瑤子
	同上	大村 蓉子
	同上	千葉 温子
	同上	横山 貴央
(2)データベース受入れ	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	西川 哲夫
	同上	島中 秀樹
	同上	中村 春木
	同上	金城 玲

注1. ◎:課題代表者、○:サブテーマ代表者

注2. 本業務に携わっている方は、全て記入。

別表3 平成21年度に於ける実施体制

研究項目	担当機関等	研究担当者
(3)人材育成		
①DB構築者の養成	東京大学大学院新領域創成科学研究科	森下 真一
	同上	中谷 洋一郎
	同上	吉村 淳
②DB高度利用者の養成	お茶の水女子大学大学院人間文化創成科学研究科	瀬々 潤
	同上	由良 敬
	同上	鷹野 景子
	同上	小口 正人
	同上	加藤 毅
	同上	渡辺 知恵美
	同上	近藤 るみ
	同上	金子 聡子
③アナテータ・キュレータの養成	長浜バイオ大学バイオサイエンス学部	池村 淑道
	同上	白井 剛
	同上	高 橋 健一
	同上	大島 一彦
	同上	依田 隆夫
	同上	塩生 真史
	同上	阿部 貴志
	同上	齊藤 美保子
	同上	上原 啓史
	同上	井口 八郎
	同上	米森 ゆかり
	同上	清水 信義
	同上	和田 千恵子
	奈良先端科学技術大学院大学	池田 俊
	同上	金谷 重彦
	山形大学工学部	木ノ内 誠
	京都大学大学院 医学研究科	平賀 壯太
	弘前大学 農学生命科学部	武藤 あきら
	所属機関無し	山田 優子
	首都大学東京	駒野 照弥
	岡山大学大学院	香川 弘昭
	奈良先端科学技術大学院大学	大橋 武久
	協和メディックス株式会社	伊藤 菁菫
	慶應義塾大学先端生命科学研究科	金井 昭夫
	慶應義塾大学大学院	菅原 潤一
	独立行政法人産業技術総合研究所	渡辺 公綱
4. プロジェクトの総合的推進	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	◎○高木 利久

注1. ◎:課題代表者、○:サブテーマ代表者

注2. 本業務に携わっている方は、全て記入。

整備実績一覧

(1) 保有データ情報

(1-1) データの種類

①生物種	外部 DB や文献から得たデータは全生物種に及ぶ その他に、グループ内に植物関連のゲノム配列データ、微生物の完全長+ドラフトゲノム(合計 1,353 種)および並びにメタゲノム解析で塩基配列が解読された多数の微生物種データ、ヒトの多型データを保有
②試料・ライブラリ等の種類、数	シアノバクテリア 38 種、根粒菌 19 種、ミヤコグサ 日本人健常人 200 人の DNA
③測定方法	ゲノム塩基の配列決定、DNA アレイ法による SNP アレルタイプ決定、tRNA 遺伝子発見プログラムなど
④データの内容	塩基配列、塩基配列上の予測遺伝子の番地、配列類似等の解析情報、遺伝子機能予測結果 ゲノムワイド確定ハプロタイプ、専門家が精査した tRNA データなど
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	国内外 DB カタログ：国内 519、海外 288 の DB 注釈データ 横断検索：検索対象 250 サイトの検索用インデックスデータ 生命科学系データベースアーカイブ：国内 28 の DB データ 生命科学系プロジェクト主要プロジェクト一覧：国内 67 プロジェクトのデータ ISND 総覧と検索：国際 DNA 配列バンク ISND の 1 億 1200 万レコードの構造化データ 統合 TV：動画コンテンツ 274 種類 生命科学学協会検索：国内 600 の学協会情報 学会要旨統合俯瞰システム：4 学会 15,729 要旨のテキストデータ 蛋白質核酸酵素全文検索：23 年間 2,148 記事のテキストデータ WINGpro：国内外 557 の DB データ Web リソースポータル：解析ツール総数 456、解析ワークフロー総数 29、著名な論文 3 報のワークフロー KazusaAnnotation：14 生物種分のゲノム配列に対する、遺伝子注釈情報 163,173 件 JAGQC：2 x 10 ⁸ 個の SNP ジェノタイプ tRNADB-CE：18 万レコード
②データ区分	■自前 ■第三者 ■文献データ ■計算結果等の二次データ □その他（下欄に詳細を記述）

③将来の増加の見込み	開発の進行および環境の整備に伴い、随時増加の見込み
④権利関係	所有者（基本的に自前データないし使用許諾を受けたもの） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 []）
⑤その他、特記事項	公開データについては、下記の（２）、（３）、（４）項を参照願います。

（１－３）データの管理状況

①更新頻度等の管理状況、体制	国内 DB カタログや ISND 全文検索は日々更新、その他は随時更新 DB 毎に担当者がアサインして管理
②その他、特記事項	

（１－４）データベース関係

①DB 管理者数	中核機関グループ全体で延べ 19 名
②キュレータ・アナデータ数	中核機関グループ全体で延べ 15 名
③データ構造	関係データベース
④DB 管理ソフト	MySQL、Postgres、sqlite
⑤サーバの OS	Linux、Solaris、Windows XP、MacOS
⑥サーバ規模	ノード数：85、コア数：614
⑦DB へのアクセス数	本プロジェクトの成果サイトへは、平成 20 年 4 月～平成 22 年 3 月に 892,000 件
⑧独立 IP 数	上と同じく、のべ 450,000 件（月毎集計の総計）
⑨その他、特記事項	保有 DB の詳細は下記[2]の項を参照ください。

（２）データ（又はDB）の連結、統合化整備

通番	データ（又はDB）の名称	公開／未公開	概要（データの種別（生物種）・数量（kB 等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	生命科学データベースカタログ http://lifesciencedb.jp/lldb.cgi?pg=1 （英語版） http://lifesciencedb.jp/lldb.cgi?gg=dbcatalog&lng=en	公開	生命科学分野のデータベースカタログ（注釈及びメタデータ）で、データベースサイトのトップページをサムネイルで表示。データベースの稼働状況をモニターして表示。利用者からのコメントを受け付ける。掲載データベース数を増やすことが課題で、海外 DB 数の増加と記載内容の充実化を図ることが必要。その他の課題はデータベースの ID の統一化など、他のデータベースに関する情報との対応を持たせること。平成 21 年度はカテゴリの見直しや、ヘルプ情報のリンク項目を開設した。 登録 DB 数：807 件（平成 22 年 4 月現在）開始年月：平成 20 年 2 月。平成 22 年 3 月の訪問

			者数：281/訪問回数：377/ページビュー：492。
2	生命科学データベース横断検索 http://lifesciencedb.jp/dbsearch/	公開	生命科学分野のデータベースや文献の横断検索サービスで、生命科学分野の国内外のデータベースと文献を網羅的に検索。平成20年3月にサービスを開始、平成20年度は対象データベースの拡大を図り、平成21年度はインデックスの更新管理並びに、関連キーワード表示機能やBlast検索機能の開発を行った。また経産省統合DBプロジェクトとの検索連携を実現した。農芸化学会から提供を受けた学会要旨を追加した他、経産省統合DBプロジェクトとの相互検索を実現するなど、他団体や組織との連携も行っている。今後も対象データベースの充実を図るとともに、更新日によるソート機能、Blastに続く解析ツールとの連携機能を開発する。検索対象DB数：250データベース(平成22年4月現在) 開始年月：平成20年3月。平成22年3月の訪問者数：8870/訪問回数：16844/ページビュー：29152。平成21年4月に対象データベースを増やし、新しいインターフェースを公開。
3	DNA データベース総覧と検索 http://lifesciencedb.jp/ddbj/ (英語版) http://lifesciencedb.jp/ddbj-e/	公開	巨大な国際DNAバンクを高速に検索し生物種、分子種、プロジェクトに分類して表示するシステム。従来巨大な計算機を必要としていたが、バンク内部を詳細に検討し1億300万レコードを58万プロジェクトに縮退させ実質上、計算機一台での検索を可能にした。さらに、ある研究機関から発表されたある生物のゲノムデータを一括してダウンロードするというような研究単位でのデータ取得を実現した。BLASTによる配列検索、核酸関連の特許公報へのダイレクトリンク、レコードの時系列展開も可能である。英語版サービスサイトの整備も完了。今後は、データベース受入れ、横断検索など他のサービスとの統合、検索の一層の高速化、ESTの臓器分類や発現データとの統合を図る。検索対象：1億300万レコード。開始年月：平成19年3月。平成22年3月の訪問者数：3938/訪問回数：13482/ページビュー：179527/ダウンロード：8.44Gbyte。
4	統合TV http://togotv.dbcls.jp/ (英語版) http://togotv.dbcls.jp/?category=English	公開	データベースやウェブツールの使い方などを動画で発信するウェブサイト。動画コンテンツ数は150に達した(平成21年4月現在)。開始年月：平成19年7月。平成22年3月の訪問者数：3961/訪問回数：13821/ページビュー：58287/ダウンロード：150.47Gbyte。
5	生命科学学協会カタログ http://lifesciencedb.jp/lbdb.cgi?gg=tool_academy	公開	生命科学系の学協会のデータベースと検索サービス。国内の生命科学系学協会を網羅したデータベース。各学会の詳細情報やウェブサイト、学会誌などの公開情報を掲載。平成19年度は日本学術会議の資料等をもとに生命科学系の学協会をリストアップしそれぞれの情報を調査、DBカタログと同じユーザインターフェースを利用して公開している。平成21年度は各学会が出版している学会誌や大会要旨の情報調査を行った。掲載学協会数：600学会 開始年月：平成20年4月。平成22年3月の訪問者数：74/訪問回数：271/ページビュー：3051。
6	学会要旨統合検索 http://lifesciencedb.jp/lbdb	公開	学会要旨集の検索サービス。著者名や施設名の表記揺れを吸収する辞書を搭載し、研究者の研究履歴をたどることが可能で、キーワードのトレンドを表示させることもできる。要旨集

	.cgi?gg=tool_tproc		の提供を受けた遺伝学会と進化学会については全文検索全文表示を行っている。非公開の分子生物学会については、索引に対する検索のみ行っており要旨本文は表示しない。昨年度には生物物理学会から要旨データの提供を受けた。学協会カタログや横断検索との重複を避けるため現在は更新を中断している。検索対象：4学会 15729 要旨。開始年月：平成 20 年 4 月。平成 22 年 3 月の訪問者数：45/訪問回数：67/ページビュー：77。
7	蛋白質核酸酵素全文検索 http://lifesciencedb.jp/pne/	公開 (更新分一部未公開)	蛋白質核酸酵素バックナンバーの全文検索サービス。国内で出版社により発行されている総説誌は、研究者にとって重要な研究発表の場であるとともに、無くてはならない良質な情報源でもある。共立出版の協力により蛋白質核酸酵素のバックナンバー電子化と公開が可能となり、すでに 1985～2007 年分を公開中である。残念ながら本雑誌は平成 22 年 1 月号をもって休刊となったため、休刊号までの総説をアーカイブし今後も提供を続ける予定である。収録論文：1985 年～2007 年(全文表示分)～2009 年(要旨表示分)。開始年月：平成 20 年 6 月。平成 22 年 3 月の訪問者数：2434/訪問回数：3887/ページビュー：7604。
8	MotDB http://MotDB.dbcls.jp/	公開	DBCLS を担っていく人材であるアノテータ、キュレータ、システムデータベース管理者向けの教材を提供するサイト (MotDB は Master of the database の略)。中核機関で独自に作成した教材 PDF ファイルもダウンロード可能。講習会開催のたびに、情報が追加されている。開始年月：平成 19 年 3 月。平成 22 年 3 月の訪問者数：1205/訪問回数：2298/ページビュー：15489/ダウンロード：1.29Gbyte。
9	遺伝子発現バンク (GEO) 目次 http://okubolab.genes.nig.ac.jp/bodymap_i/ (英語版) http://lifesciencedb.jp/geo-e/	公開	ヒト遺伝子の解剖学的な発現パターンデータの統合サイトを構築した。発現パターンは、測定法毎に異なる場合があることが知られており、ここではできるだけ客観的な発現パターンの解釈を可能にするために、5 種類の発現データ、即ち iAFLP、GeneChip、EST、NCBI の SAGEmap によるタグマップ、SAGE データの独自タグマップに基づく発現データを表示可能にした。平成 22 年 3 月の訪問者数：1334/訪問回数：2614/ページビュー：9652/ダウンロード：885.65Mbyte。
10	生命科学系データベースアーカイブ http://dbarchive.lifesciencedb.jp/	公開	アーカイブデータベース一覧を提供するサービス。各登録データベースに対して作成したデータベースとデータのメタデータ、利用許諾と共に、アーカイブデータのダウンロードが可能である。データの一括ダウンロードと共に、検索して絞り込んだデータのダウンロード機能も備えている。また、データベースの寄託を公募するためのページを作成し、その中で「寄託データベース利用許諾作成マニュアル」、「データベース利用許諾用語定義解説」、「データベース公開に関する覚書書式」の作成を行い公開している。アーカイブデータベース一覧には、現在国内の生命科学系データベース 28 件登録されている。開始年月：平成 21 年 3 月。平成 22 年 3 月の利用状況 (訪問者数：682、訪問回数：1227、ページビュー：4704、ダウンロード：2.93GB)
11	生命科学系主要プロジェクト一覧	公開	日本国内の生命科学系関連の主要なプロジェクトに関して詳細な情報を収集しまとめたもの。1990 年代以降の 4 省 (文科省、経産省、厚労省、農林水産省) の生命科学に関連する 67

	http://togodb.dbcls.jp/lbdb_project (英語版) http://togodb.dbcls.jp/lbdb_project_en		の主要プロジェクトについて、プロジェクトの期間、概要、予算、代表研究者、公開の状況、成果、文献等に関する情報が閲覧できる。開始年月：平成20年3月。昨年度まで「ゲノム・ポストゲノム主要プロジェクト一覧」として公開していたものの後継版。
12	TogoProt http://lifesciencedb.jp/dbsearch/togoprot/	試験的公開	蛋白質関連データベース統合検索システム。ユーザは日本語または英語のキーワードで蛋白質を検索し、得た蛋白質リストから蛋白質を選ぶと、その蛋白質に関連づけられる各種のデータをデータベース横断的に集めて表示することができる。蛋白質の代わりに蛋白質ファミリーを選べば、ファミリーのいずれかのメンバーに関連づけられる各種データを表示できる。平成21年12月からプロトタイプを公開し、平成22年3月までに40のデータベースを収録した。4月から正式版を公開予定。
13	日本の生命科学データベース政策 http://lifesciencedb.jp/sciencepolicy/	公開	これまでにライフサイエンスのデータベースに関わる検討がいろいろなところで行われており、一元的に俯瞰できる状態ではなかったため、本プロジェクトに関与しそうな報告書、会議資料（ウェブページ情報等含む）等を1箇所に集約し、時系列に整理した。開始年月：平成21年3月。4月から外部公開した。
14	WINGpro http://wingpro.lifesciencedb.org/ http://wingpro.lifesciencedb.jp	公開	ライフサイエンス分野の内外のデータベースの情報を収集、整理、分類、431DBを収録。ディレクトリからは、データベースの構築法による分類と生物種および対象による分類でデータベースを一覧可能。平成19年3月30日正式公開。平成19年7月2日よりユーザが記事を投稿および編集できる機能を公開。全文検索、分類一覧が可能。入力方式、一括データ取得、tipsなど記述。ユーザによる登録を促すことが課題。分類一覧の自動作成を開発。平成22年3月の利用状況（訪問者数：1229、訪問回数：2419、ページビュー：11641）
15	Web リソースポータルサイト for バイオインフォマティクス http://tools.lifesciencedb.jp/	公開	実験データの解析や公的データの加工に使用する解析ツールや環境を案内するサイト。ベンチメソッドオントロジーの一部を利用。解析ツール総数456、解析ワークフロー総数29、著名な論文3報のワークフローを収録、一括ダウンロードも可能。平成19年3月30日正式公開。教科書に近いサービスであるので、多くの研究者が行う実験フローをいち早く提供するなど工夫が必要。平成22年3月の利用状況（訪問者数：123、訪問回数：171、ページビュー：310）
16	Jabion Genome Viewer : HAL http://www.biportal.jp/genome/	公開	HAL（生物種 human）（ヒトゲノムに潜む遺伝子の位置と構造、その機能などの注釈情報を提供。）を既存のゲノムビューアである Jabion Genome Viewer に受け入れたテストサイト。H20.1.7 公開。ヒトゲノムに潜む遺伝子を発見する信頼性の高いコンピュータアルゴリズムにより予測した遺伝子数多数を公開。
17	KazusaAnnotation http://a.kazusa.or.jp/	公開	高度情報集積型データベースのソーシャルブックマーク型サービスで、独自開発のソーシャルブックマークによるオープンアノテーションを可能にした遺伝子情報統合データベースゲノムデータベースに対するマニュアルキュレーションによる情報抽出を行い、163,173件の

			情報を蓄積し、公開している。今後は、改善を加えながら運用し、情報の集積によるゲノムを基盤とした生物学研究情報の統合を図る。アクセス数 167,153 (平成21年度)。
18	KazusaNavigation http://navi.kazusa.or.jp/	公開	高度情報集積型データベースのポータルサービスで、ソーシャルネットワーク型のElggをベースに開発。現在は公開運用しながら情報蓄積を続けている。学会、研究会情報などを定期的に更新している。また、かずさ内部のキュレーション作業の情報共有などに有効活用している。今後、さらに、研究者コミュニティに利用をよびかけ、情報交換の促進による知識の統合を図る。登録者数80名、アクセス数58,243 (平成21年度、Wikiと共通)
19	KazusaWiki http://wiki.kazusa.or.jp/	公開	高度情報集積型データベースのデータとりまとめサイトで、Wikiタイプのデータ共有サイトである MediaWiki をベースに開発。現在は公開運用しながら情報蓄積を続けている。ワークショップの告知、既存データベースの受入などにも活用している。今後は改善を加えながら運用し、研究者による情報とりまとめサイトとして発展させていき、研究情報の統合を図る。提供者側からの情報入力もすすめていく予定である。登録者数、アクセス数 27,165。
20	「健康への貢献遺伝子データベース」「持続可能型社会への貢献遺伝子データベース」 http://dbcls.nagahama-i-bio.ac.jp/	公開	人類の健康に役立つ遺伝子を学生が各自で想定して、世界で進行している環境生物試料に関する大規模シーケンシング (メタゲノム解析) で得られた大量な塩基配列を対象に、公的データベースにおいて未だ機能が特定されていない 1,700 万件以上の遺伝子候補に着目して、想定した遺伝子の探索を学生実習として行った。他大学からの探索結果も、相互に情報交換を持ちながら追加し、集合知の形成を目標に学生等の個人名入りで、2007年6月から公開している。自習可能なテキストを作成し、公開も行っている。月平均アクセス数は、35,000件である。登録データの精査を行うために、シニア世代の専門家のキュレーション体制の拡充を図りつつ、高品質なデータの収集を目指す。
21	「エキスパートがキュレートした tRNA データベース」 http://trna.nagahama-i-bio.ac.jp/cgi-bin/trnadb/index.cgi	公開	微生物の完全長とドラフトゲノム (合計 1356 種)、メタゲノム解析由来塩基配列 (1,700 万件)。ウイルスゲノム (151 種) を対象に、3 種の予測プログラムを用いて網羅的な tRNA 遺伝子探索を行い、172,073 件の tRNA 遺伝子情報 (生物種・アミノ酸タイプ・アンチコドン・位置・塩基配列) を 2008 年 7 月より提供している。月平均アクセス数は、156,000 件である。21 年度はウイルス・イネを対象としたデータ登録も行った。今後は、オルガネラ・真核生物を対象にデータ登録を行っていく予定である。公開開始年月: 2008.5 月～
22	JAGQC http://crane.gen.kyushu-u.ac.jp/JAGQC/JAGQC_top.html	公開	我国の GWAS 解析データの信頼度を第三者として評価・公開し、同データを関連分野の研究者が利用する際の参考指標となることを目的として作成されている。平成 22 年 1 月に大幅な改訂を行い公開とした。現在は標準日本人のジェノタイプデータ 200 例をモデルデータとして、GWAS での複数の Q.C. 解析を行い、解析結果を画像として、また、解析結果全データと解析過程の LOG FILE をダウンロード可能なフラットファイルとして WEB 上に公開している。現在はモデルデータを用いた結果が公開されており、外部からのアクセス、及びダウンロードは実質的には無い。今後、ライフサイエンス統合 DB の GWAS db と連携を強化により広く宣伝する予定である。

(3) DB基盤システム、ツール等開発成果物の整備（※試験的、限定的公開済みのものも含む。）

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	遺伝子名称シソーラス http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	分子情報から医学文献までの遺伝子や蛋白質名称の正確な同定のための遺伝子固有名称（「遺伝子名」と一般名称（「ファミリー名」）の辞書データの構築を目的として、様々なデータベースで利用されている名称の収集と専門的キュレータによる編集を行い、遺伝子が持つ多様な名称の関係を明示した遺伝子名称シソーラスを開発した。本シソーラスは、ヒトをはじめ9種類の生物をカバーしている。
2	生物学名羅日英対応辞書 http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	研究分野でよく使われる生物種の基準として、学名に日本語一般名を対応させた生物学名日本語一般名対応辞書を開発した。対応付けは、塩基配列データベース (DDBJ) の登録エントリ数が多い生物種から順番に行った。また、標準和名が存在しない場合、その生物を説明する一般的な名称を用いた。今年度は追加作業を継続。登録件数：学名（ラテン語）と他言語での種名の対応 - 77,625 件、学名と和名の対応 - 66,346 件、そのほか、学術用語の日英対応 - 42,823 件。
3	施設名称辞書 http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1 (http://lifesciencedb.jp:32888/agency/agency.cgi)	公開 (内部公開)	日本のライフサイエンス研究を俯瞰するための重要な情報源として、各種関連学会の抄録などの報告文書があるが、これらをデータベース化する際に問題となるのが、例えば、大阪大、阪大、大阪大学大学院などといった、施設名称研究室名称の表記ゆれである。これに対応するために、施設名称辞書を開発した。すでに公開しているものは1625の名称についてまとめたものであるが、昨年度からその範囲を15,000件以上の研究機関・研究グループに広げ、ホームページ掲載情報や学会要旨データベースより収集した機関名別名情報を格納し、機関名の表記ゆれに対応した検索を可能とした。今年度も辞書の拡張を継続し、引き続き内部公開で試験運用中である。
4	動物解剖学自動分類タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	動物の臓器、組織を、大きく10のグループに分類(大分類)し、さらにそれぞれのグループを細かく分類、合計40の小分類グループに分類する。基本的には、与えられた解剖用語に対して、解剖用語辞書、病理関連語彙の分類辞書、形容詞の解剖用語辞書、一般的な臓器名称の分類辞書、の4種類の辞書を順番に検索し、上記のカテゴリに分類する。
5	植物解剖学自動分類タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	植物(維管束植物)の部位、組織を大きく6のグループに分類し、さらにそれぞれのグループを細かく分類、合計11の小分類グループに分類する。分類における検索対象としては、生物種に合わせて、種子を持たない維管束植物の解剖用語辞書、イネ科の解剖用語辞書、その他被子植物の解剖用語辞書、裸子植物の解剖用語辞書、トウモロコシ属の解剖用語辞書、フウチョウソウ目の解剖用語辞書を用いる。

6	都市名国名自動検出タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	論文やデータベースレコードに見られる国の名称の未記載や国の名称にみられる表記ゆれを吸収することを目的に、国別に分類するための辞書を開発した。
7	生物アイコン http://lifesciencedb.jp/lldb.cgi?gg=resource_icon (英語版) http://lifesciencedb.jp/lldb.cgi?gg=resource_icon&lng=en	公開	ユーザが自身のHPなどで自由に利用できるように提供されている生物のアイコン集。目的とするイラストをキーワード検索でき、生物に関する外部情報のリンクも付いている。ユーザはイラストに関するコメントを書き込むことができ、さらには同じ条件で配布可能なイラスト等を投稿することもできる。アイコン数：217。開始年月：平成20年6月。定期的にアイコン数を増やし、インターフェースを更新。ヘルプページを作製した。平成22年3月の訪問者数：433/訪問回数：530/ページビュー：851/ダウンロード：86.4Mbyte。
8	BodyParts3D http://lifesciencedb.jp/ag/pgm	公開	解剖学用語、すなわち臓器・器官・部位の名称やそれらにくくる概念をモデル人間中の3次元座標で定義した辞書。BodyParts3D内の空間関係で用語関係を表現すると、ツリー型表現（いわゆる解剖オントロジー）と違い、改変にロバスタな表現が可能。今後も用語の追加を続ける。今年度は胸部データの更新と筋肉の新規作成（774種）を行い、全パーツ数が1315種となる予定である。来年度早々に全てフェーズ2データとして公開予定である。アクセス情報はアナトモグラフィーと一緒にカウントしているため下記参照。
9	アナトモグラフィー http://lifesciencedb.jp/ag/ (http://lifesciencedb.jp/bp3d、現在公開準備中) (現バージョン英語版) http://lifesciencedb.jp/ag/index.jsp?locale=en	公開 (内部公開)	BodyParts3Dから部品を選択して自由に人体の部分のモデル図が書ける。視点固定の図譜や3Dアトラスと異なり透過度や着色、視点は自由に選択可能。今後は、解剖学教育や医学用語辞書の助けになる人体版グーグルマップを目指す。すでに公開したバージョンに加え、拡大・縮小、平行移動など自由度が高くかつ高速な図形処理が可能なバージョンを来年度早々に公開予定である。平成22年3月の訪問者数：3083/訪問回数：4613/ページビュー：180969/ダウンロード：35.7Gbyte。
10	TogoDB http://togodb.dbcls.jp/	試験的公開	データベース受け入れシステム。さまざまな形式のさまざまな種類のデータを簡単に読み込み、自動的に共通のウェブ検索インターフェースを生成する。また、自動的にウェブサービス(SOAP/WSDL API)を生成する。これらの機能により、容易に維持困難なDBを移設できる。また、本ツールのダウンロードにより、誰でもDBを構築可能である。平成20年度には新たに検索結果のダウンロード機能を追加した。平成22年3月の訪問者数：1234/訪問回数：3301/ページビュー：72543/ダウンロード：5.27Gbyte。
11	TogoWS http://togows.dbcls.jp/	公開	国内外のウェブサービスを透過的に統合するシステム。国内各サービスの稼働状況を継続的に監視し、サービスの連携に必要なデータ形式変換機能等を提供する。これにより、ユーザは統一的命名規則と統一的データ構造で国内外のウェブサービスにアクセスすることができるため、相互運用が容易となり、ワークフローの作成・蓄積を促進する。平成20年度にはキーワード検索の書式を統一した。開始年月：平成20年4月。平成22年3月の訪問者数：362/訪問回数：1547/ページビュー：37058/ダウンロード：2.25Gbyte。

12	OpenID 認証システム http://openid.dbcls.jp/ (英語版) https://openid.dbcls.jp/?lang=en	公開	一つの ID で複数のサイトを認証できるアカウントシステム。このシステムを利用することで、ユーザはひとつの ID とひとつのパスワードを保持すればよい。これにより、各サイトで認証サービスを用意する必要がなく、サイト間のユーザ情報の集約が容易になる。開始年月：平成 20 年 4 月。平成 22 年 3 月の訪問者数：368/訪問回数：975/ページビュー：2720。
13	OReFiL http://orefil.dbcls.jp/ (英語版) http://orefil.dbcls.jp/en	公開	オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システム。信頼性や概要を容易に取得するために、資源の所在だけでなく、その資源について書かれている論文情報や、他の利用者の評価を閲覧できる。アカウントを取得することで、自ら評価を行うことも可能である。また、検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムから OReFiL を利用できる。英語版ヘルプなどを整え、環境を整備した。開始年月：平成 19 年 8 月。平成 22 年 3 月の訪問者数：412/訪問回数：909/ページビュー：9147。
14	Allie http://allie.dbcls.jp/ (英語版) http://allie.dbcls.jp/en	公開	医学生物系論文書誌情報データベース MEDLINE を対象とし、出現する略字とその正規系のペアを検索するシステム。生命科学系の論文では非常に多くの略字が使われており、同じ表記でも全く違う意味を示していることが少なくない。そこで、利用者の興味のある略字を検索語として入力することで、その使われ方を一覧表示すると共に、論文の発表年を提示する。また、検索された各略字について、その意味で使われている論文中で共起する他の略字も提示する。検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムから Allie を利用できる。英語版ヘルプなどを整え、環境を整備した。開始年月：平成 20 年 4 月。平成 22 年 3 月の訪問者数：1235/訪問回数：1849/ページビュー：6659。
15	DBCLS Galaxy http://galaxy.dbcls.jp/	公開	生命科学データに特化した、ウェブベースの対話的ツール組み合わせインターフェース。マウス操作と簡単な入力で、プログラムを書くことなく、途中結果を確認しながら、用意されたツールを対話的に組み合わせて使うことができる。DBCLS Galaxy では、TogoWS やテキストマイニングツールなど、DBCLS で開発されたツール群を組み込み、利用できるようにした。ログインは DBCLS OpenID で可能である。また、メニューの国際化対応へ協力し、メニューの日本語化を行い、DBCLS Galaxy のみならず、元の Galaxy にも反映させた。平成 22 年 3 月の訪問者数：101/訪問回数：513/ページビュー：2304。
16	ScrapParty	公開	ブラウザのプラグインによるキュレータ支援用アプリケーション。オンラインで文献やウェブサイトを閲覧しながら、重要な記述や画像、図表を集めコメントとともに簡単に DB 化するツール。平成 18 年度アノテーションの調査、試験に利用した。下記 Wired-Marker が後継アプリケーションであり、現在 ScrapParty 名での開発は終了。
17	Wired-Marker http://www.wired-marker.org	公開	情報共有ツール、アノテーション支援ツール。Firefox ブラウザのプラグインとして働く。データベース構築、アノテーションを支援するブラウザ用のプラグイン。オンラインで閲覧したページ、論文のテキストや図をマークしブックマークすることが可能。再訪問の際にも

			<p>同じ箇所がハイライトされアノテーションのエビデンスとして詳細な情報が残せる。またフォルダを自由に階層化して情報を整理することが可能。新たにキャッシュ機能が加わり、ウェブページの書き換えが起きても過去のデータを参照できるようになった。</p> <p>公開日：H19.12.19 総ダウンロード数：326,934件。</p>
18	<p>メタデータ要素レポジトリ (MDeR)</p> <p>http://mder.jst.go.jp/ (英語版)</p> <p>http://mder.jst.go.jp/index_en.html</p>	公開	<p>メタデータ要素の検索、検索結果の比較、収録メタデータのメタデータ要素の一覧表示が可能。ISO/IEC 11179 Part3(Registry metamodel and basic attributes)に準拠した形で、国際標準3種、データベース3種のメタデータ要素を収録。国際標準2種の調査終了し、搭載予定。収録内容の精査と充実によりデータ収集・データベース構築時のデータ形式等の決定に役立つサービスを目指す。公開日：平成20年3月26日。</p>
19	CBRC タンパク質構造情報ワークフロー	限定公開	<p>構造に関する予測プログラム等をGridにより効率的に分散処理し、結果を実験研究者等に提供する(プロトタイプ)。平成20年8月末よりプロジェクト関係者に限定し公開開始。後、下記20のタンパク質アノテーションワークフローに統合された。</p>
20	<p>CBRC タンパク質アノテーションワークフロー</p> <p>http://togo.cbrc.jp/wf/PAW_About.html?Eng (英語版)</p> <p>http://togo.cbrc.jp/wf/en_PAW_About.html?Eng</p>	公開	<p>立体構造未知のタンパク質に関し、構造及び機能のヒントとなる情報を幅広く実験研究者等に提供する。二次構造予測、埋れ残基予測、フォールド認識、ディスオーダー予測、膜タンパク質オールベータ・ベータシート予測、細胞内局在予測をGridにて分散処理を行う一方、データベース検索及び疎水性予測の実行を他のサーバへ依頼し、結果を取得後、全ての結果をユーザが解析し易いよう配置し出力する。平成20年8月末より一般公開開始。</p>
21	<p>CBRC タンパク質比較情報ワークフロー</p> <p>http://togo.cbrc.jp/wf/comp_prot_info_about.html (英語版)</p> <p>http://togo.cbrc.jp/wf/en_comp_prot_info_about.html?Eng</p>	公開	<p>相同なタンパク質を比較することで保存部位等構造上重要な部位を表示し、実験研究者等に提供する。マルチプルアラインメントを用い保存性が高い残基及び二次構造を表示する。平成21年3月末より一般公開開始。</p>
22	<p>CBRC タンパク質立体構造モデリングワークフロー</p> <p>http://togo.cbrc.jp/wf/modeling_about.html?Eng (英語版)</p> <p>http://togo.cbrc.jp/wf/en_modeling_about.html?Eng</p>	公開	<p>このワークフローはアミノ酸配列を入力とし、二次構造予測、ディスオーダー予測等を行い、タンパク質立体構造のテンプレートとなる構造をデータベースから見つけ出し、それを基に立体構造モデリングを行うことで、タンパク質の配列情報から3次元立体構造の構築を可能とするものである。平成21年11月に一般公開。</p>

23	CBRC アクティブ・ワークフローによる遺伝子ネットワーク推定システム (ASIAN-K) http://togo.cbrc.jp/active.html (英語版) http://togo.cbrc.jp/en_active.html	公開	このアクティブ・ワークフローは操作性・柔軟性に優れたプラットフォームを使用し、プログラムをインタラクティブに接続し、選択的に組み合わせる事により、独自の解析プラットフォームによるワークフロー解析を可能とするものである。ASIAN-K は、遺伝子の発現量(発現プロファイル)を入力とし、クラスタリング技術を用い遺伝子間ネットワーク推定を可能とするアクティブ・ワークフローである。平成 21 年 1 月より一般公開。
24	CBRC SOAP 対応サービス (POODLE, WOLFPSORT)	限定公開	CBRC で開発されたディスオーダー予測ツール POODLE (Prediction Of Order and Disorder by machine Learning) 及びタンパク質局在予測ツール WOLFPSORT (Protein Subcellular Localization Prediction) を SOAP 対応 (平成 19 年度に開発したプロトタイプを元に拡張・増強) としたものである。現在試験運用中。
22	genoDive Pro / Eu	限定公開	DAS 規格に準拠したゲノム情報を統合的に閲覧することが可能な DAS サーバークライアント。現在は試験配布中、ソースコードも含め完全公開し配布を行う予定。
25	遺伝子名称変換プログラム	限定公開	植物遺伝子名や遺伝子 ID の名称の食い違いを解決するための名前変換サービス。現在は試験運用中、テスト終了後、完全公開の予定。
26	なんでも DAS	限定公開	任意のウェブページに含まれる遺伝子名を抽出して、GFF 形式に変換するサーバ。このサーバを活用することで、インターネット上のあらゆる遺伝子関連情報をゲノム上にマッピングし、情報の統合的な閲覧を可能とする目的で開発した。現在は試験運用中、テスト終了後、完全公開の予定。
27	MiGAP http://migap.lifesciencedb.jp/ (英語版) http://migap.lifesciencedb.jp/en/	公開	技術革新により大量に生産されるようになった微生物ゲノムの利用価値を高めるため、アノテーションを自動的に行うパイプラインを開発し、試験的にプロジェクト内部に公開、評価を受けた。評価の意見を反映した修正作業を進め、平成 21 年 6 月に一般公開。利用のためには、OpenID の取得が必要。開始年月: 平成 21 年 3 月。平成 22 年 3 月の訪問者数: 162/訪問回数: 656/ページビュー: 44123。
28	inMeXes http://docman.dbcls.jp/im/ (英語版) http://docman.dbcls.jp/im/index.html.en	公開	医学生物系論文書誌情報データベース MEDLINE を対象とし、利用者が入力した文字列とマッチする表現を逐次的(1文字入力する度)に検索するシステム。生命科学系分野の学術文献を英語で執筆する際、既に発表されている論文の中で実際に使われている表現を容易に検索出来るれば、特に非ネイティブスピーカーが多数の日本人研究者にとって有用であると考え、構築した。マッチした表現の全 MEDLINE 中(題目・要旨)での出現頻度も表示されるほか、当該表現が使われている文や関連情報を表示させることも可能。開始年月: 平成 21 年 7 月。平成 22 年 3 月の訪問者数: 452/訪問回数: 1311/ページビュー: 15332。

29	TogoDoc http://docman.dbcls.jp/pubmed_recom/	試験的 公開	文献情報及び論文 PDF を管理し、また、特定の文献情報群に関連する論文情報を提示するシステム。日々多くの論文が発表されている昨今においては自身が興味を持つ文献情報を効率良く取得することが容易では無くなってきている。また、多くの学術雑誌がオンラインジャーナルを発行していることから論文全文を PDF ファイルで取得出来る機会が増大し、研究者の PC に多くの論文 PDF ファイルが保存されている。以上の背景から文献情報の効率的な取得、整理環境を実現すべく、そのために必要な情報は既に研究者自身の手元にある文献情報リストから得られるようなシステム TogoDoc を構築した。利用者 PC に保存されている PDF ファイルを解析し必要な書誌情報を自動的に付加するなどの機能を持つ文献情報管理ソフトウェア TogoDocClient と書誌情報のやり取りや PDF ファイルの授受を行うなどの連携機能を持つ。公開年月：平成 21 年 12 月。
30	JAGQC PIPELINE	未公開	<u>JAGQC PIPELINE</u> : GWAS 研究によって得られたジェノタイプデータを入力データとし、定評のある PINK program 及び EIGENSTRAT program を利用して、解析データの QC を一貫して行い、その結果及び解析過程を画像データ及びフラットファイルとして出力するプログラム。 <u>JAGQC DB</u> : 上記パイプラインの結果を保存し、WEB ページ上に公開するプログラム。

(4) その他の成果物 ((2)、(3) に該当しないもの)

通番	名称	公開／未公開	概要
1	統合 PJ 成果公開サイト http://lifesciencedb.jp	公開	統合 DB プロジェクトに関する中核機関の成果公開サイト。生命科学データベースカタログ、DB 横断検索、INSD 全文検索、統合 TV、OReFiL、生命科学学協会検索、学会要旨統合俯瞰システム、MotDB、LSDB ブログ等を掲載している。開始年月：平成 20 年 6 月。
2	統合 PJ 成果公開サイト(英語版) http://lifesciencedb.jp/?lng=en	公開	国外で本プロジェクトの紹介をする機会も増えてきたことから、英語での利用が可能なサービスを集めて、英語版のポータルサイトを公開した。現在紹介されているデータ・サービスは、ゲノム・ポストゲノム主要プロジェクト一覧、生命科学データベースカタログ、生物アイコン、DNA データベース総覧と検索、遺伝子発現バンク (GEO) 目次、アナトモグラフィー/BodyParts3D、Wired-Marker、TogoWS、togoTV。そのほか、プロジェクト紹介ページ、参画機関の成果ページ(英語版があるもの)へのリンクをはった。開始年月：平成 21 年 3 月。
3	事業広報サイト http://lifesciencedb.mext.go.jp/ (英語版) http://lifesciencedb.mext.go.jp/en/index.html	公開	統合データベースプロジェクトの紹介サイト

4	ライフサイエンスの広場 http://www.lifescience.mext.go.jp/index.html	公開	文科省のライフサイエンス政策等の紹介サイト
5	統合データベース間の連携調査 http://charles.kazusa.or.jp/lifesciencedb/	公開	代表的モデル研究植物である「イネ」ならびに「シロイヌナズナ」のゲノムアノテーション型公開データベース各 46 サイト、25 サイトの基本項目の調査と実験生物研究者 188 名のアンケートにより調査し、それぞれのデータベース間の連携のための現状の課題と将来の DB 統合にむけた課題の整理を行った。
6	検索アルゴリズムを含めた知識情報技術の動向調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4	公開	次世代の生物情報データベース統合に必要となる知識情報技術として、検索システム、データマイニング、Web 2.0 およびグリッドコンピューティングに焦点を絞り、聞き取り調査や文献調査によって動向を調べた。
7	遺伝統計学分野における解析技術の基礎調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4	公開	遺伝統計学分野で用いられる連鎖解析、連鎖不平衡解析)、QTL 解析等の解析手法と、それぞれの手法における代表的なアルゴリズム計 8 種類の調査を行ない、その特徴および長所・短所を評価した。併せて、各手法の代表的プログラム計 15 種類に関して、実装されているアルゴリズム、動作環境などを調査し、その評価を行った。
8	臨床情報・疾患健康情報の調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4	公開	わが国の代表的コホート研究に関してその研究の背景と目的、対象地域、ターゲット疾患、特徴的な検査項目、対象人数、代表研究者、研究開始時期、資金源等について調査した。
9	ライフサイエンスプロジェクトの調査結果	試験的公開	この約 10 年間のライフサイエンス国家プロジェクトの事業内容や成果を明らかにする目的で、4 省庁のゲノム、ポストゲノム関連の主要プロジェクトの調査を行った。プロジェクト名、期間、予算、主要研究者、主要事業内容、公開状況、公開 DB、データダウンロードサイト、公開 HP、事業化関連、報告書、その他文献等に関する情報を情報ソースへのリンクをできるだけつけた形で整理した。
10	データベース受入に関するアンケート調査 http://lifesciencedb.jp/lldb.cgi?pg=0&gg=ukeire_survey	試験的公開	受入による統合化の方針策定と受入対象の候補抽出を目的として、国内データベースを対象として、データベース内容と検索方法、経歴、維持・管理状況、提供可能性及び提供時の問題点などをアンケートにより調査した。送付した 104 件中、65 件の回答が得られ、半数以上の機関が、統合 DB 事業に対して全データもしくは一部データを開示できると回答した。結果の概要および詳細報告 (PDF) をウェブ上で公開した。
11	BioHackathon2008 ホームページ http://hackathon.dbcls.jp/	公開	分散環境のままで、各種 DB をアルタイムかつ統合的に利用できる環境を構築すること目的に開催した、国内外のライフサイエンス DB のウェブサービスにおけるデータ構造や命名規則を標準化するための BioHackathon2008 の活動内容と経過を紹介するサイト。2008 年 2 月

			11日から15日まで都内で開催。海外参加者34名、国内参加者37名。
12	BioHackathon2009 ホームページ http://hackathon2.dbcls.jp/	公開	標準化技術および標準化を利用する技術をより安定的に使用可能にすることを目的として開催した BioHackathon2009 の活動内容を紹介するサイト。巨大なデータについてどうすべきか、セマンティックウェブ技術をどう利用すべきかなどの議論の報告が掲載されている。2009年3月16日から20日まで、都内および沖縄で開催、海外参加者30名、国内参加者28名。
13	データベース高度利用者養成カリキュラム http://togodb.sel.is.ocha.ac.jp/	公開	データベース高度利用者養成において利用したカリキュラム、資料の公開。
14	シンポジウム・講演会 http://symposium.lifescience.db.jp/	公開	DBCLS が主催したシンポジウムサイトをリストしたサイト。現在は2009年6月12日および10月5日開催の2つのシンポジウムについてリンクがあり、6月12日分のサイトでは、後援動画9件・スライド7件・発表ポスターPDF68件、10月5日分のサイトでは講演動画8件を掲載している。
15	外部評価 http://lifesciencedb.jp/lldb.cgi?gg=hyouka	公開	登録ユーザによるサービスへの評価内容を公開するサイト。平成19年から基本的に前年度の公開サービスへの評価を実施。平成20年度からはウェブサイト記入方式（5段階評価とフリーコメント）に変更し、サービスについての理解を深めてもらう工夫も行った。なお、いただいたご意見への対応・回答も併せて掲載している。評価者の人数：約60名（平成19年度）、58名（平成20年度）、117名（平成21年度）
16	データベース高度利用者養成カリキュラム	公開	データベース高度利用者養成において利用したカリキュラム、資料の公開。 http://togodb.sel.is.ocha.ac.jp/
17	BioHackathon2010 http://hackathon3.dbcls.jp/	公開	統合検索に向けたセマンティックウェブ技術の調査と開発を目的とした、BioHackathon2010の活動内容を紹介するサイト。2010/2/8 - 12の期間、海外32名、国内36名が参加し、LinkedDataを実現するための合意形成、セマンティックウェブ技術のフィージビリティ調査のための開発などが行われた。
18	バイオデータベース構築者養成コース演習ノート http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php?Introduction	公開	バイオデータベース構築者養成コースにおいて使用した演習用教材・演習ノートの公開。