

ライフサイエンスデータベース統合推進事業
基盤技術開発プログラム
研究開発課題
「データベース統合に関わる基盤技術開発」

研究開発終了報告書

研究開発期間：平成23年4月～平成26年3月
研究代表者：小原雄治
(大学共同利用機関法人情報・システム研究機構ライフサイエンス統合データベースセンター、
センター長)



§ 1 研究開発実施の概要

(1) 実施概要

《チームとしてまとめた研究開発の実施内容や経緯、得られた成果等の研究全体概要を、簡単にわかりやすく、概ね1ページ程度でまとめてください。それぞれの研究グループの研究が、どのように全体の成果につながったのかを中心に記載してください。》

生命科学において、急増するデータの解析、管理、共有は、研究の基盤に関わる最重要課題であるにもかかわらず、国として戦略的に取り組まれて来なかった。その間、データベースの構築と運用は、個々の研究者やプロジェクトに任されてきた。しかし、データベースの維持管理には様々なコストがかかるため、プロジェクトの終了とともにデータベースが消失したり、共有可能なメタデータが付与されていないなど多数の問題が、データの解析や再利用を妨げてきた。このため、本来、価値のあるデータベースも国際的な評価を得にくいなどの状況も招いてきた。

統合データベースプロジェクトは、これら日本の生命科学系データベースが抱える問題を解決し、データベースの価値を高め、ユーザーの利用環境を飛躍的に向上させることを目的とするプロジェクトである。平成 18 年度より開始された文科省統合データベースプロジェクトでは、国内のデータベース産生プロジェクトの調査に始まり、データベースのカタログ化、横断検索、アーカイブの構築、各種データベースの利用サービスの開発等、主として幅広いデータベースへのアクセスを広げ、データ共有を研究コミュニティに広める活動を行ってきた。

平成 23 年度からは、NBDC のライフサイエンスデータベース統合推進事業がスタートし、国内の分野別データベースの本格的な統合が始まった。その中で基盤技術開発プログラムが負う課題は、データベースの統合を可能にする技術を開発し具体的に統合を実現すること、研究者のデータベース利用環境を、従来型の one by one の解析から多種類の解析を組み合わせた高度なものにすること、断片化した知識の検索からより集約された知識の取得を可能にするサービスを開発することである。世界的には次世代シーケンスデータに代表されるようなビッグデータ解析の重要性が叫ばれるようになり、そのためのデータベース側の対応も必須となった。

そこで我々は、それまでの文科省プロジェクトの成果を背景に、大規模集中型の統合ではなく、セマンティック・ウェブ技術を利用した「フェデレーション型」のデータベース統合を提案し、基盤技術開発に採択された。具体的には、DBCLS、CBRC、京都大学の 3 機関が共同し、DDBJ や PDBj など国内の拠点データベース、統合化推進プログラムの分野別データベースの協力の元、次に示す 7 課題を実施した。

- (1) データベースの RDF による統合化
- (2) 解析プラットフォームによる統合利用環境の整備
- (3) インターネットを活用した高度検索技術の開発
- (4) RDF 化に資するオントロジー、辞書、コーパスの整備、標準化
- (5) 大規模ゲノム配列データの利用技術開発
- (6) 情報統合化・知識発見のためのキュレーション支援
- (7) 統合データベースに関わるコンテンツの作成、整備

代表者を務める DBCLS グループが統合化の技術開発を担当、CBRC は保有するバイオインフォマティクスの解析サービスを生かし、課題(2)の解析プラットフォームの開発を担当、京大グループは KEGG のバックボーンである DBGET/LinkDB システムを利用し、課題(1)(3)の RDF 化や課題(4)のオントロジー開発、メタゲノム・メタメタボローム等を対象に課題(5)を担当した。

3 年間の技術開発の結果、セマンティック・ウェブ技術によるゲノム情報の統合データベース TogoGenome を完成させた。また、それをベースに生物学で用いられる可視化パターンを他のデータベース開発でも容易に再利用できるようモジュール化した TogoStanza を開発した。開発に際してはプロジェクト内で閉じるのではなく、BioHackathon での国際連携、国内の開発

者の連携を重視し、オントロジーの開発や RDF の設計、SPARQL 検索技術の検討をはじめとする、標準化と技術連携に努めた。これにより、セマンティック・ウェブ技術の利用とノウハウの蓄積という点で、国内外のライフサイエンス分野に留まらない波及効果を生むに至った。また、大規模利用技術開発としては、NGS データの急増にいち早く対応し、SRA を始めとする大型データベースの検索技術の開発とサービスの公開を行った。またデータベースのユーザーを対象とした、利用促進のための日本語コンテンツを継続して開発し生命科学分野の幅広い評価を受けている。

(2) 研究開発成果のデータベース等
別紙1を参照

§ 2. 研究開発構想(および構想計画に対する達成状況)

(1) 当初の研究開発構想

研究開始時、データベース統合化の最終的な目標は、知識発見につながる整理された、意味のある情報を提供することであり、そのためには、個々の研究者や技術者が扱えない規模の多様かつ大量のデータベースの中から、必要とするデータ、情報をストレスなく入手できる環境、およびコンテンツを整備開発することが必要であると設定した。また、更なる知識発見を促進するために、研究で得られた知見が速やかに統合データベースに格納され、効率よく共有できる環境を構築する必要があるとした。開始当初、セマンティック・ウェブ技術が未開拓であったため、初年度を調査に当て、プログラムオフィサーの指導の下、アジャイル開発手法も取り入れつつオープンな研究開発に努めたところ、ほとんどの目標はサービスやデータベースとして実現した。サービスの NBDC への移管実装に関しては、計算機の調達や運用コスト、セキュリティの問題、今後の開発の継続状況などの点を勘案して、サービス別に対応すべき問題として残った。

(2) 新たに追加・修正など変更した研究開発構想
(DBCLS グループ)

募集要項に医療用画像等に関わる画像データの利用技術(管理・検索・標準化・定量的解析等)を開発することが求められていたため、統合化推進プログラムのヒト脳疾患画像データベース統合化研究と連携し、脳画像データベースの利用技術開発を実施する予定であったが、公開データの不足と、DBCLS の医療系の個人情報を含むデータにアクセスするための権限の問題から計画を中止した。医療系データへのアクセスの問題は、今後の統合プロジェクトの中心課題でもある、ヒトの個人ゲノムデータへのアクセスについても生じるため、研究開発にそのようなデータにアクセスする必要がある場合には、対策を講じる必要がある。

(京大グループ)

オーソログクラスタ(OC)を遺伝子機能予測へと応用する際に、OC 間の距離を系統プロファイル(生物種の保存度)で定義することが有効であるとされており、本研究でも系統プロファイル間の距離を計算した。しかしながら、全ゲノムが決定される生物種が増加するに従い、生物種系統の偏りが大きくなるという問題が判明した。そこで、生物種系統の偏りを考慮して類似度を定義する手法の開発と遺伝子機能予測への応用を計画している。

LinkDBと反応オントロジーのより高度な RDF 検索として、Stanza のさらなる応用を検討している。本研究課題では、RDF 化を実現するところまでであったが、最近の SPARQL 検索応用技術の発展への対応を新たに計画し、一部は実現している。

(3) 達成状況
なし

＜以下は例示＞

細い黒点線が当初計画、太い黒実線が実際の実施状況
別紙線表を参照

(4) 研究開発の今後の展開について

近年、データへのアクセス性の向上により、国内及び国家間の情報の流れを改善するための視点やイノベーションを進化させていくという認識が高まっている。その流れを受けて、特定のデータが、著作権、特許などの制限なしで、全ての人が望むように利用・再掲載できるようにするオープンデータという概念が国際的に押し進められている。平成 25 年には、日本政府も G8 サミットでオープンデータ憲章に合意し実行計画を発表した。我々の取り組んでいるライフサイエンス分野の RDF によるリンクド・オープン・データ (LOD) の開発やバイオハッカソンの取り組みは、この新潮流に即しており、ライフサイエンス分野のオープンデータで国際的なイニシアチブを取る事ができる。

§ 3 研究開発実施体制

(1) 研究チームの体制について

- ・「○」は研究代表者又は主たる共同研究者
- ・「*」は本事業研究開発費から人件費を支出した者(専任RAを除く)
- ・「■」は本事業専任RA(博士課程で年間 200 万円程度の給与)

① 「研究代表者:小原雄治」DBCLS グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
小原雄治	DBCLS	センター長	総括	H24.4～H26.3
米澤明憲	DBCLS	センター長	総括	H23.4～H24.3
大久保公策	遺伝学研究所	教授(兼務)	研究開発	H23.4～H26.3
岡本忍	DBCLS	特任研究員 (特任准教授)	研究開発	H23.4～H26.3
川本祥子	DBCLS	特任研究員 (特任准教授)	総括補助	H23.4～H26.3
金進東	DBCLS	特任研究員 (特任准教授)	研究開発	H23.4～H26.3
坊農秀雅	DBCLS	特任研究員 (特任准教授)	研究開発	H23.4～H26.3
箕輪真理	DBCLS	特任研究員 (特任准教授)	総括補助	H23.4～H26.3
山口敦子	DBCLS	特任研究員 (特任准教授)	研究開発	H23.4～H26.3
小野浩雅	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H26.3
片山俊明	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H26.3
狩野芳伸	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H24.3
川島秀一	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H26.3
河野信	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H26.3
内藤雄樹	DBCLS	特任研究員 (特任助教)	研究開発	H23.4～H26.3

仲里猛留	DBCLS	特任研究員 (特任助教)	研究開発	H23.4~H26.3
山本泰智	DBCLS	特任研究員 (特任助教)	研究開発	H23.4~H26.3
Kocbek Simon	DBCLS	特任研究員	研究開発	H24.7~ H25.10
飯田啓介	DBCLS	特任研究員	エディター	H23.4~H26.3
王悦	DBCLS	特任技術専門 員	研究開発	H23.4~H26.3
大田達郎	DBCLS	特任技術専門 員	研究開発	H23.4~H26.3
呉紅艶	DBCLS	特任技術専門 員	研究開発	H23.4~H26.3
藤枝香	DBCLS	特任技術専門 員	モデリング	H23.4~H26.3
鐘ヶ江弘美	DBCLS	技術補佐員	キュレーター	H24.4~H26.3
中村春木	大阪大学蛋白質 研究所	客員教授	アドバイザー	H23.4~H26.3
金城玲	大阪大学蛋白質 研究所	客員准教授	研究開発	H23.4~H26.3
岩崎渉	東京大学大気海 洋研究所	業務協力者	研究開発	H23.4~H26.3
小笠原理	遺伝学研究所	業務協力者	研究開発	H23.4~H26.3
山崎千里	産業技術総合研 究所	研究員	研究開発	H23.4~H25.3

②「研究分担者:産総研 CBRC」グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
浅井潔	生命情報工学 研究センター	センター長	総括	H23.4~H26.3
福井一彦	同上	主任研究員	ツール開発&検 証	H23.4~H26.3
田代俊行	同上	テクニカルスタッ フ	ツール開発	H23.4~H26.2

③「研究分担者:五斗進」京大グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
五斗進	京都大学 化学研究所	准教授	総括、LinkDB の統合利用環境 への応用	H23.4~H26.3
時松敏明	同上	特定助教	新規分野データ 活用技術の開発	H23.4~H26.3
小寺正明	同上	特定助教	反応オントロジ ーの整備	H23.4~H26.3
Diego Diez	同上	特定研究員	LinkDB の統合 利用環境への応 用	H23.4~ H23.12

中川善一	同上	特定研究員	反応オントロジーの整備	H23.4~H26.3
守屋勇樹	同上	特定研究員	新規分野データ活用技術の開発	H23.4~H26.3
武藤愛	同上	技術補佐員	反応オントロジーの整備	H24.10~ H24.12
西村陽介	同上	D3, D4	LinkDBの統合利用環境への応用	H24.7~H26.3

(2)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

- ・東京農工大学天竺桂弘子講師とのカイコアノテーションパイプライン構築。
- ・東京工業大学近藤科江教授との低酸素下での遺伝子発現データ解析手法の構築。
- ・東京大学武田憲彦特任助教とのストレス刺激下での遺伝子発現データを中心とする統合オミックス解析手法の構築。
- ・熊本大学荒木准教授・水口研究員との TogoTable 開発。
- ・産業技術総合研究所糖鎖医工学研究センター成松センター長・澤木研究員・鹿内研究員との GlycoProtDB の RDF 化。
- ・創価大学木下准教授との RINGS の RDF 化。
- ・立命館大学川寄教授・新潟大学奥田准教授との GlycoEpitope の RDF 化。
- ・野口研究所山田研究員との WURCS の開発。
- ・Macquarie University (オーストラリア) Matthew P Campbell 研究員との UniCarbKB の RDF 化。
- ・Justus-Liebig-University Giessen (ドイツ) Thomas Lütke 研究員との MonosaccharideDB の RDF 化。
- ・University of Georgia (アメリカ) Rene Ranzinger 研究員との GlycomeDB の RDF 化。
- ・N.D. Zelinsky Institute of Organic Chemistry (ロシア) Philip Toukach 研究員との BCSDB の RDF 化。
- ・株式会社 Preferred Infrastructure との塩基配列データベース高速検索法の研究開発。
- ・製品評価技術基盤機構バイオテクノロジーセンター(NITE)市川博士らとのキュレーション支援システムの開発。
- ・国立遺伝学研究所の藤澤研究員にテキストマイニングを用いた、ゲノムアノテーションにおけるタンパク質命名の自動化と質的な改善に関して技術支援をした。
- ・大阪府立成人病センター 井上正宏教授との新規がん細胞培養法に関する研究
- ・日本大学 加野浩一郎教授・沖嘉尚助手との体細胞の脱分化機構に関する研究
- ・遺伝学研究所中村教授との文献からの遺伝子機能アノテーション支援と TogoStanza によるゲノムデータ可視化
- ・基礎生物学研究所内山助教との MBGD 微生物オルソログデータベースの RDF 化支援と TogoStanza による可視化
- ・東京工業大学黒川教授との微生物生育環境、表現型オントロジー、培地オントロジーの協同開発
- ・製品評価技術機構市原博士と NBRC 菌株カルチャーコレクションデータの RDF 化
- ・理化学研究所大熊博士、榎谷博士と JCM 菌株カルチャーコレクションデータの RDF 化
- ・国立環境研究所河地博士と NIES 菌株カルチャーコレクションデータの RDF 化
- ・Swiss Institute of Bioinformatics (スイス)の UniProt グループ Jerven Bolleman 研究員、The James Hutton Institute (イギリス)の Peter Cock 研究員、国立遺伝学研究所 藤澤貴智研究員らとの塩基配列・アミノ酸配列の位置座標記述オントロジー FALDO の研究開発
- ・国立遺伝学研究所 藤澤貴智研究員と国際塩基配列データベース INSDC (DDBJ) の RDF 化とオントロジー開発

- Ontario Institute for Cancer Research (カナダ)の Joachim Baran 研究員、Fondazione Istituto Nazionale Genetica Molecolare (イタリア)の Raoul Bonnal 研究員、Parco Tecnologico Padano (イタリア)の Francesco Strozzi 研究員と、GFF, GVF などゲノム配列アノテーションファイルの RDF 化のためのソフトウェア BioInterchange の開発
- Ontario Institute for Cancer Research (カナダ)の Joachim Baran 研究員、Fondazione Istituto Nazionale Genetica Molecolare (イタリア)の Raoul Bonnal 研究員、Parco Tecnologico Padano (イタリア)の Francesco Strozzi 研究員、Illumina 社(アメリカ)の Mayank Tyagi 研究員、イルミナ株式会社の癸生川絵里研究員と、イルミナ社の次世代シーケンサー用クラウド環境を扱うための BaseSpace Ruby SDK の開発
- 東北大学メディカル・メガバンクの荻島創一博士とのセマンティック・ウェブによるアルツハイマー病関連データベースの構築
- European Bioinformatics Institute (イギリス)の James Malone 研究員、Simon Jupp 研究員とのデータベース RDF 化に関する共同研究
- European Bioinformatics Institute (イギリス)の Nick Juty 研究員、Camille Laibe 研究員、Sarala Wimalaratne 研究員との Identifiers.org による RDF 用の URI 標準化のための共同研究
- Stanford University (アメリカ)の Michel Dumontier 研究員、MAASTRO clinic (オランダ)の Scott Marshall 研究員、バイオサイエンスデータベースセンターの坂東明日佳研究員らとのデータベースのメタデータ記述の RDF 化に関する標準化
- UC Berkeley (アメリカ)の Robert Buels 研究員との RDF 化されたゲノムアノテーションの JBrowse での可視化に関する研究開発

(京大グループ)

- JAMSTEC 高見研究員とのメタゲノム・ゲノムの機能解析システム構築。
- RIKEN-BRC 榎屋研究員・村田研究員との BRC バイオリソースデータへの LinkDB 拡張。

§ 4 研究実施内容及び成果

4. 1 研究課題名:生命科学分野におけるデータベース統合化のための基盤技術開発 (DBCLS グループ)

研究開発実施内容及び成果

DBCLS は研究代表グループとして提案課題全体を担当するとともに、NBDC と協力して、BioHackathon, や講習会の開催を実施した。以下、各課題の 3 年間の実施内容について報告する。

1) データベースの RDF による統合化

- 有用 DB のサードパーティとしての RDF 化

NBDC が管理運用するサービス「生命科学系データベースアーカイブ」に対し、より高度な検索を実現可能とするために、NBDC と協力して RDF 化を行った。また、RDF 化したアーカイブでのユースケースを NBDC に提示していただき、それに合った検索を可能とするために、主に以下の二点の作業を行った。(1) 生物種は NCBI Taxonomy、化合物は InChI 記述をアーカイブ内の標準記述として与えた。(2) リテラルで与えられた ID を URI で書き直してデータに加えた。これらの成果により、生物種、化合物、あるいはデータベース ID でアーカイブ内の複数のデータをつなぐ検索が可能となった。作成した RDF データの SPARQL エンドポイントおよびそれ上で動く 5 つのユースケースは NBDC のサイト <http://dba-rdf.biosciencedbc.jp/sparql> で公開中である。

•TogoDB の RDF 化機能開発

TogoDB は、論文のサプリメントや研究者が Excel など管理しているような表形式の有用なライフサイエンスのデータを容易にデータベース化し公開するためのシステムである。データベース化されたデータを RDF に変換し統合的に利用しやすくするために、TogoDB に RDF/XML と Turtle 形式への変換機能を実装した。データ更新の都度、リリース機能を利用することにより、データベースのメタデータとコンテンツが自動的に RDF 化できるようになった。また生成された RDF は個別の SPARQL エンドポイントに格納され SPARQL 言語による検索も可能となった。さらに、生成される RDF をより良いものとするために、データベースごとにオントロジーを定義できるように拡張した。これを用い、カラムごとに既存のオントロジーや個別に定義したオントロジーを利用して property や Class を指定できるようになり、きちんと意味づけされた RDF を生成できるようになった。



•DB メタデータ RDF 蓄積

RDF は元々ウェブ上に分散したリソースについて、統一した枠組みで記述するものである。そのため、RDF データの検索システムはフェデレーション型を取ることが自然である。当初はその検索をサポートするために、どこにどのようなデータがあるかを DB メタデータとして記述し、その RDF を蓄積する方法を取ろうとした。しかし、後述する「RDF を蓄積可能な DB の調査・整備開発」において、フェデレーション型の検索システムは現時点では実用レベルに達していないことが明らかになった。そのため、本プロジェクトでは、まずデータ集中型のシステムあるいは特定のデータソースを用いたシステムを採用し、フェデレーション型の検索システム構築は次期プロジェクトにおいて中心課題とすることとした

•自然文からの SPARQL 生成システム開発

RDF データの活用のため必須的な SPARQL クエリ言語は複雑で一般人は使い難い。その問題を解決するため、自然文から SPARQL クエリを自動生成するシステムである LODQA を開発した。今は OMIM のデータに題して比較的簡単なクエリの生成が可能になっている。そのシステムは <http://lodqa.org> で α 公開中である。

•RDF を蓄積可能な DB の調査・整備開発

セマンティック・ウェブ技術を用いたデータベース統合において、RDF データを格納するトリプルストアの性能はシステム全体の性能に影響を及ぼす。そこで、生命科学データに適したトリプルストアを選択するために、Cell cycle ontology(11M トリプル)、Allie(94M トリプル)、PDBj(590M

トリプル)、UniProt(4B トリプル)、DDBJ(8Bトリプル)の5つのデータセットに対し、OWLIM-SE、Virtuoso、4store、Bigdata、Mulgara の5種のトリプルストアのロード時間、クエリ応答時間の測定などを行った。性能比較の結果、Cell cycle ontology には OWLIM-SE, 4store が、Allie には OWLIM-SE, Virtuoso, 4store が、UniProt と DDBJ には Virtuoso が比較的良好な性能を示すことが分かった。すなわち、おおざっぱに述べて、100M 前後のトリプル数およびそれ以下では OWLIM-SE, 4store が、100M 前後のトリプル数およびそれ以上では Virtuoso が性能的に向いていることが示された。結果の詳細については、<http://kiban.dbcls.jp/togordf> で公開している。

2) 解析プラットフォームによる統合利用環境の整備

RDF 基盤を利用して複数のツールを組み合わせた解析をスムーズに行うため、解析プラットフォームとして DBCLS で運用中である DBCLS Galaxy (<http://galaxy.dbcls.jp/>) に RDF 入出力のフレームワーク SADI 対応機能を追加した。この追加により、CBRC が提供する SADI 対応の解析ツール群をはじめとした SADI 対応ツール群を DBCLS Galaxy からスムーズに利用可能となった。さらに、その有効性を確認するため、京都大学の API とも連携し、タンパク質とそのリガンドを入力とし、リガンドの類似化合物群を求め、類似化合物群とタンパク質とのドッキングシミュレーション結果を出力するワークフローを DBCLS Galaxy 上で構築した。

DBCLS+京大+CBRC連携でのワークフロー構築例

- 京都大学の simcomp サービスでシアル酸の類似化合物を検索
- ノイラミニダーゼと類似化合物をRDFデータとしてCBRCに送信

ノイラミニダーゼの立体構造を指定
シアル酸の構造をSMILESで指定
京都大学の類似化合物検索を実行
類似化合物の計算結果一覧
CBRCのサービスにRDFデータを送信

DBCLS+京大+CBRC連携でのワークフロー構築例

- CBRCの AutoDock サービスでドッキングシミュレーションを実行
- 結果をRDFで取得・可視化

ドッキングシミュレーションの結果
CBRCのサービスにRDFデータを送信

3) インターネットを活用した高度検索技術の開発

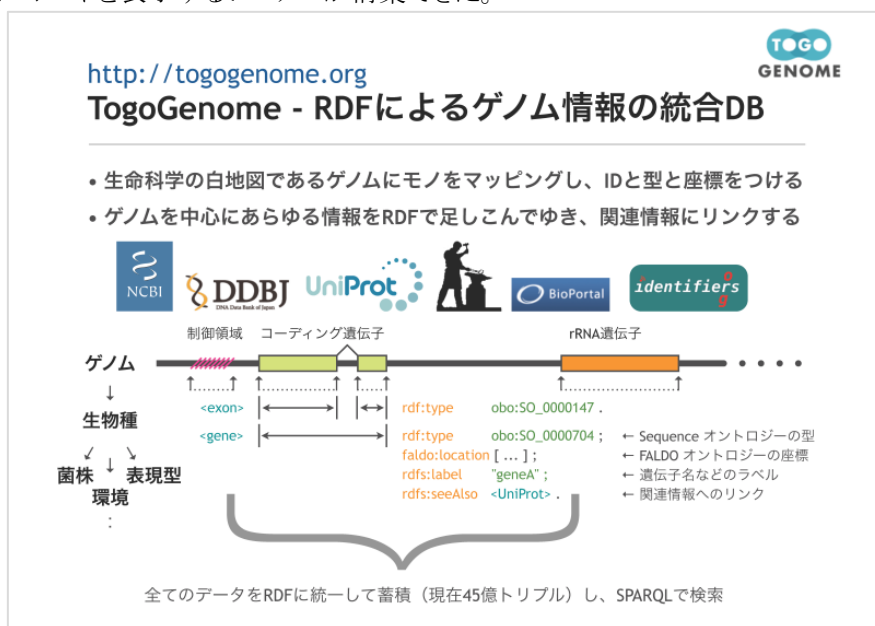
・既開発サービスの機能向上

文部科学省統合データベースプロジェクトで開発された既開発サービス、データベースカタログや横断検索、DB アーカイブについて高機能化や RDF 化を実施した。DB カタログについては経産省統合 DB プロジェクト並びに BioDBC Core との協力により RDF 化の項目やオントロジー等の検討を行い、最終年度に BioHackathon での連携を中心に NBDC により第一バージョンの RDF 化カタログが構築された。今後はユースケースに合わせた拡張を検討していく。横断検索については、3 年間のサーバー運用の支援を DBCLS にて行った。また、インデックスに対するオントロジーや辞書のマッピングを行い、検索結果表示の質の向上を検討した。アーカイブの RDF 化については、課題(1)において、アーカイブの公開システムである TogoDB の RDF 関連機能を開発するとともに、NBDC と協力しデータの RDF 化を実現しエンドポイントを公開した。開発と運用を NBDC と DBCLS の間でどのように分担していくかは今後の課題である。

・可視化技術等の活用

データ科学としての生物学・医科学の情報の多くは、生物種ごとのゲノム座標に紐づけて整理される。そこで、ゲノム情報の分野において、研究課題「1) データベースの RDF による統合化」の成果を活用し、情報検索と可視化のフレームワークを構築した。TogoGenome (<http://togogenome.org/>)は、生物種とゲノムに関連する多種多様な情報を集約し、ゲノム情報の統合的で新しい検索システムを構築するとともに、TogoStanza サーバーによる再利用可能なデータと可視化パーツを提供することで、オープンデータの促進と新たなデータベース開発のコスト削減にも貢献する。

TogoGenome では、これまで統一的に扱うことが困難であった様々なデータ型や対象の異なる多様なデータセットを一元化するため、ゲノム情報のためのオントロジー（整理された共通語彙）と RDF（世界標準のデータ形式）データモデルの設計を行い、すでに RDF 化が完了しているタンパク質配列データベース UniProt や遺伝子オントロジーGO などのアノテーション情報をデータレベルで統合した。また、毎年主催している開発者会議 BioHackathon での国際連携にもとづき、配列座標・生物種・表現型・疾病・環境・培地など様々な標準オントロジーを新規に開発して統合した。この統合によって、全ゲノム情報に対し複数のオントロジーをファセットとして用いた効率的な絞り込み検索が可能となり、検索結果からは環境・生物種・遺伝子などの意味単位ごとに集約したレポートを表示するシステムが構築できた。



- 有用分野でのプロトタイピング、統合化推進プログラムとの連携（H23-H25）
ライフサイエンス分野では、データベースで確立された「生物学的意味ごとに繰り返し利用される情報可視化パターン」が存在する。その可視化パターンを他のデータベース開発でも容易に再利用できるようモジュール化し TogoStanza として提供した。TogoStanza サーバーは、SPARQL エンドポイントから RDF データを検索し、遺伝子機能アノテーションを表示する表や、インタラクティブな操作が行えるグラフなど目的に適した可視化を行うことができるフレームワークである。すでに微生物メタゲノム DB の Microbedb.jp や藍藻ゲノム DB の CyanoBase など外部のデータベースと連携して多数の TogoStanza が開発され相互に利用されている。
- 一方で、コンピュータが得意ではない実験研究者が、複数のデータベースが RDF 形式でつながっているという LOD ネットワークを容易に利用可能とするために、TogoTable を開発した。これは、ウェブ上で解析ソフトやエクセル表などの複数行にわたる表形式のデータに RDF の LOD ネットワークからさまざまなアノテーション情報を取得して元の表形式データに付け加えるというものである。当初は熊本大学との共同研究によりプロテオミクスデータに対してアノテーション情報を付加するために開発されたが、最終的にはプロテオミクス分野に限らず一般的なライフサイエンスデータに適応できるシステムとなった。
- 統合化推進プログラムとの連携では、上述の微生物メタゲノム DB を構築する際にオントロジー作成、データの RDF 化、TogoStanza の提供で連携した。また、糖鎖のグループとも糖鎖構造を表現するユニークな ID の開発、オントロジー作成、データの RDF で連携した。化合物のグループとも、共通の構造の表記法やオントロジーを使うなどで連携した。

4) RDF 化に資するオントロジー、辞書、コーパスの整備、標準化

- 言語資源リポジトリ開発

生命科学データの RDF 化に資する重要な言語資源としては(1)データ分類の基準になる「オントロジー」、(2)データ表現のための言葉を定義する「辞書」、(3)文献中情報を表す「アノテーション」がある。これら言語資源の管理のため、オントロジーは世界的に生命科学オントロジーのレポジトリとして認められている BioPortal を活用して、辞書と文献アノテーションは PubDictionaries と PubAnnotation という独自のプラットフォームを開発した。そのための標準フォーマットも開発した。オントロジーの場合は BioPortal の活用の他、BioPortal オントロジーの可視化、再利用を助けるシステムとして OntoFinder、OntoFactory、OntoCloud とのシステムを開発した。これらプラットフォームの上、生命科学データの RDF 化のため必要なオントロジー、辞書、文献アノテーションの整備を実施した。

- 自然言語処理、テキストマイニング技術の整備

辞書と文献アノテーションの開発のため必要な自然言語処理・テキストマイニング技術を整備した。具体的には文章認識(BioSentencer)、構文解析(enju)、用語認識(PubDictionaries)のツールを収集及び開発した。特にこれらのツールは辞書と文献アノテーションの管理プラットフォームである PubDictionaries と PubAnnotation システムと REST API で連動できるようにして一つの繋がったシステムとして動くように開発した。

5) 大規模データの利用技術開発

- 遺伝子発現のリファレンスデータセットの整備 (H23-H25)

「RefEx (Reference Expression dataset)」(<http://refex.dbcls.jp/>)を開発、公開した。「RefEx」は、4つの異なる実験手法(EST、GeneChip、CAGE、RNA-seq)によって得られた40種類の正常組織における遺伝子発現データを統合し並列に表現することで、測定手法間の比較とともに各遺伝子の発現量を直感的に比較することが可能なリファレンス(参照)データセットである。論文を読んでいて見かけた馴染みのない遺伝子がどの組織・臓器で発現しているのか、どんな特徴があるのかを論文の記述ではなく、バイアスの少ない実際の測定データから研究者自身の目で簡単に確認することができ、「知りたいことが予め決まっている」ことを検索するだけでなく、遺伝子発現解析などで見出された不詳な遺伝子群の関係性を調べるツールとして「探索的に何らかの知見を得られる」ことを志向している。また、RefEx は生命科学データの共有および再利用の活用例のひとつであり、RefEx で使用したデータはダウンロードおよび再利用することが可能であり、データ駆動型研究のためのツールとしてだれでも自由に使うことができる。

- 開発技術の連携統合(H25)

前年度までに「メタ情報による大規模ゲノム配列データの整理・再利用促進技術開発」として開発した次世代シーケンサー(NGS)からのデータに対する分類にもとづいて「DBCLS SRA」(<http://sra.dbcls.jp/>)サイトの大幅な改良を行った。「DBCLS SRA」(旧 SRAs・Survey of Read Archives)では、DDBJ/EBI/NCBI により収集された NGS データを、メタデータの記載事項をよりどころに実験目的(ゲノム解読、発現解析、メタゲノム等)や測定機器、対象生物種によって分類し、公共データベース中の目的の NGS データを簡単に検索、ダウンロードし、再利用を行えるようにしたウェブサービスである。H25年度は、特に、NGSによるトランスクリプトーム(RNA-seq)データを解釈するための技術開発として、メタデータだけでなく配列データそのものの解析にも踏み込んで QC 値などの統計値



を事前に網羅的に計算し、データ再利用の際のリファレンスとして役立つように整備を行った。また、NGS データと、そのデータにもとづいて発表された論文とを対応づける技術を活用し、公共 NGS のメタデータに記載の情報だけでなく、関連論文の情報も含めての検索を行えるように技術の連携統合を行った。また、遺伝子や転写産物をさまざまなキーワードからすばやく検索し、その結果をわかりやすく提示することができる遺伝子検索エンジン GGRNA を開発・公開し、NCBI RefSeq に収録された全生物種に対応した。このサービスにより、遺伝子名や各種の ID、タンパク質の機能や特徴などのキーワードだけでなく、短い塩基配列やアミノ酸配列から遺伝子を高速に検索することが可能になった。さらに、短い塩基配列の高速検索に特化した検索エンジン GGGenome を開発・公開した。

6) 情報統合化・知識発見のためのキュレーション支援

・キュレーション支援システムの開発

生命科学分野の略語情報を検索するシステム Allie について、日本語対訳を充実させるとともに、表記上のゆれを吸収して同義表現をまとめる技術を開発した。さらに、収められているデータを RDF を用いた表現として生成し、SPARQL エンドポイントを立ち上げ、DBpedia やライフサイエンス辞書などの他の RDF 化されているデータベースとのリンクを生成した。

文献情報管理・推薦システム TogoDoc については iOS 上で動作するクライアントを開発することで、スマートフォンから効率的に推薦文献を確認することが出来るようになった。また、国際公共ゲノムデータバンクに新規ゲノム情報を登録する際に必要となる、タンパク質名の定義(デフィニション)をするキュレーション作業について、テキスト処理技術を用いることで作業支援するツール TogoAnnotator を開発し、DNA Data Bank of Japan (DDBJ)でのアノテーションパイプラインで利用される作業が進められているほか、製品評価技術基盤機構バイオテクノロジーセンター(NITE)におけるキュレーション作業の軽減を目指した共同研究が開始された。

・協働キュレーション作業運用技術の整備

情報資源のデータベースへの整理格納と維持更新には、情報の内容を特徴づける最適な視座の提供や、さまざまな粒度での人手による永続的な協働作業が必要である。今回のプロジェクトでは、コンピュータ支援による協働作業手法(CSCW computer-supported cooperative work)の研究分野で開発されているツールを組み合わせ、統合化推進課題との連携において、文献キュレーション、オントロジー開発、データマッピングの作業を支援した。リファレンス微生物の文献による遺伝子アノテーションや TogoTV の作成で、ソフトウェアプロジェクト管理ツール(Trac)を応用してタスク管理を行なう手法を開発し実際の作業支援を行なった。

7) 統合 DB に関わるコンテンツの作成、整備

・チュートリアル動画の作成

「統合 TV」(<http://togotv.dbcls.jp/ja/>) を作成・公開した。「統合 TV」は生命科学分野の有用なデータベースやツールの使い方を動画で紹介するもので、平成 17 年 7 月のスタート以来、750 本以上の番組を作成・公開しており、平成 25 年度は 1 月末までに 83 本を公開した。平成 26 年 1 月の一意な訪問者は 9,431 であった。研究規模の拡大、研究形態の多様化に伴って、データベースやウェブツールの利用方法が複雑化・専門化しているなか、研究者や技術者、それらを志す初学者が必要とする情報やツールに到達するための体系的なガイドとして、さらにはそれらを使いこなし活用するための動画チュートリアルとして他に比類の無い日本語コンテンツである。

・日本語コンテンツの作成

「ライフサイエンス 新着論文レビュー」(<http://first.lifesciencedb.jp/>) および「ライフサイエンス 領域融合レビュー」(<http://leading.lifesciencedb.jp/>) を作成・公開した。「新着論文レビュー」は、トップジャーナルに掲載された日本人を著者とする生命科学分野の論文について、論文の著者自身の執筆による日本語の解説(レビュー)を、だれでも自由に閲覧・利用できるよう、いち早く公開するもので、平成 22 年 9 月のスタート以来、600 本以上のレビューを公開しており、平

成 25 年度は 1 月末までに 114 本を公開した。平成 26 年 1 月の一意な訪問者数は 29,521 であった。「領域融合レビュー」は、日本分子生物学会、日本蛋白質科学会、日本細胞生物学会、日本植物生理学会との協力のもと、生命科学において注目される分野・学問領域における最新の研究成果について、第一線の研究者の執筆による日本語の総説(レビュー)を、だれでも自由に閲覧・利用できるよう、無料で公開するもので、平成 24 年 9 月のスタート以来、24 本のレビューを公開しており、平成 25 年度は 1 月末までに 12 本を公開した。平成 26 年 1 月の一意な訪問者数は 6,393 であった。ともに、Web にはなかなか存在しない、高い信頼性をもつ良質な日本語コンテンツとして、最新の研究成果あるいは研究動向についての情報源としてだけでなく、有用なデータベースとしての機能をもはたしている。

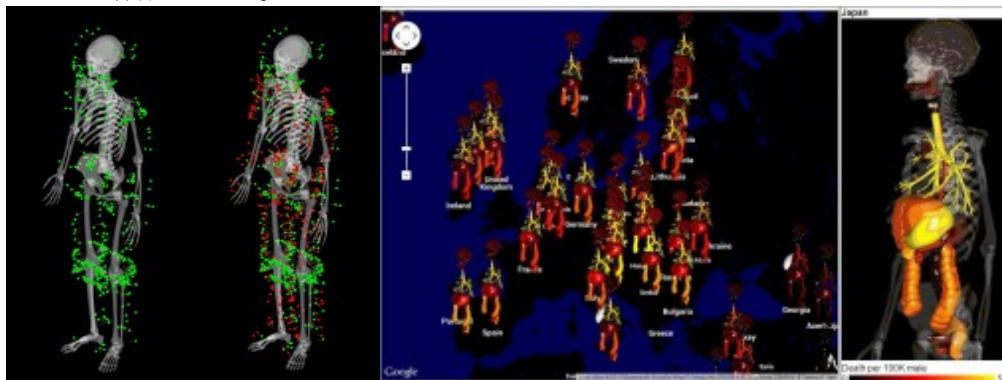
・生物並びに臓器形状 3D データの作成

デジタル人体模型 BodyParts3D は 2012 年 6 月に version 4.0 を公開した。概念名称をファイル名とするデータの管理方法を廃止し、(1)オブジェクトファイル名称、(2)ファイル名と解剖概念の対応表、(3)概念間の階層関係(オントロジー)の 3 つによって管理を行う方式に変更、これにより新旧オブジェクトが同時に利用可能になった(対応表：<http://lifesciencedb.jp/bp3d/get-info.cgi?version=4.0&tree=isa&cmd=upload-all-list&title=list%20of%20obj%20files>)。このために第三者がオブジェクトファイルの形状から概念に対応付けする対応表作成ステップを補助する webGL を用いた web application を構築して使用することにより、概念表現の誤りが起きる割合が低減した。データレジストリを検索する web-application (Bodyparts3Dsite)では概念ツリー、身体位置の二つからドリルダウンが可能になり、名称検索に加えオブジェクトの体積や重心の位置による並べ替えが可能となった。

map editor の機能拡充として、身体アドレスの発行を実装した。editor のインタラクティブ画像上のクリックでクリックされた表面上の点を pinURL として提示することで、身体アドレスを利用者間で交換可能にする機能を追加した。例えば下の URL はブラウザからリクエストすると上右図のような画像(三尖弁上の鏝)を返すが URL 中に概念 ID およびミリ単位での鏝の空間座標を含むので Web 環境以外でも鏝の位置を表現することができる。これによって“あなたは(私は)下の URL の場所に異常があります。”等身体部位についての専門家非専門家間のメール等でのコミュニケーションが可能になった。

[http://lifesciencedb.jp/bp3d/API/animation?{"Part":{"PartID":"FMA7240"},"Pin":{"PinSize":30,"PinX":-6,"PinY":-124,"PinZ":1221,"PinArrowVectorX":0.4}}](http://lifesciencedb.jp/bp3d/API/animation?{)

次に mapAPI 利用の拡充を行った mapAPI は一般の地図 API と同様に、作図だけでなく複数で行う地図作成(collaborative mapping)や地図の色付けによって解剖区分別頻度データの可視化等に利用でき、また他人の地図上の印を自分の地図に重ねる map overlay 等の再利用も容易となる。このような解剖地図の利用は世界でも例がないためにあまり気づかれていないが、例示する意味でいくつかの web application を作成し、公開した。る意味でいくつかの web application を作成しました。



Collaborative mapping 左中：共同作業用サイトにて作成 http://lifesciencedb.jp/bp3d/collaborative_map/
 WHO 部位別がん死亡統計右:部位別癌統計を解剖概念にマップし地理 API 上に表示 <http://lifesciencedb.jp/bp3d/cancermap/>

4.2 研究課題名:解析プラットフォームによる統合利用環境の整備(産業技術総合研究所 生命情報工学研究センターCBRC グループ)

研究開発実施内容及び成果

ライフサイエンス分野では様々な解析ツールやソフトウェアが存在しており、効率よく解析するにはその使い易さに加え、多種多様な大量のデータからの知識抽出のためデータベースとの連携が必要である。我々は解析ツールの統合化として、DB に蓄積された多種多様なデータと解析ツールを連携させ必要な情報を効率よく入手でき活用することを目的とし、高度なインターネット技術を用いた解析プラットフォームの基盤技術開発(ノード化)、利用環境の整備(RDF 化)及び解析結果の統合化されたレポート表示機能(可視化)を行う。

・解析プラットフォームの基盤技術開発

KNIME のプラットフォーム上にて動作する新規解析ノードを、組み合わせ型として RNA 解析、タンパク質解析及び分子シミュレーション・ワークフローとして公開した。KNIME ノードとしては合計 28ノードの開発を行い、その内訳としては外部解析ツールが18ノードであり、所内独自開発の解析ツールとして 6 ノード及び4つの DB 連携ノードを開発し公開した。

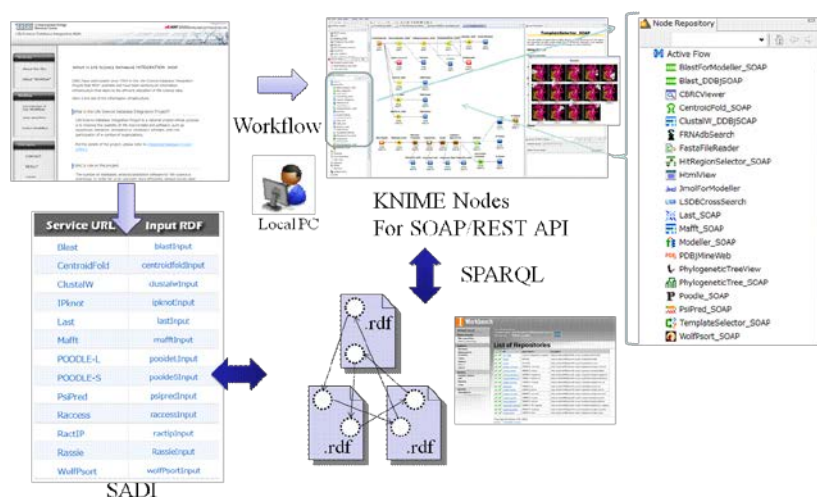
・利用環境の整備(RDF 化)

セマンティック技術に対応し、RDF 入出力機能を備えた SADI (Semantic Automated Discovery and Integration) サービスによる 14 個の解析ツールを公開している。ここでは、オントロジーとして SIO (Semantic science Integrated Ontology)を利用したサービスを公開している。また KNIME の DB 連携ノードでは SPARQL ノードを公開しており所内・外に構築したトリプルのエンドポイントへの検索を可能としている。

・レポート表示機能(可視化)

KNIME の新規機能を利用し、PhylogeneticTree (DNA, RNA, Protein)ワークフローの解析結果を自動でワードファイルにレポートサービスを行う機能を公開した。

本研究で開発した解析ツール等は WEB サイト(togo.cbrc.jp)にて公開している。また利用者によりワークフローや SADI を使い易くするため、日本語と英語版のインストール及びユーザ・マニュアルを公開した。



4.3 研究課題名:データ統合と新規分野データ活用のための基盤技術開発(京大グループ)

研究開発実施内容及び成果

データベース検索のセマンティック化に必要とされる要素技術として LinkDB の開発を行うとともに、様々な解析ツールを整備しメタゲノムなど新規な大規模データへと応用した。また、遺伝子や化合物に比べて整備が遅れている反応オントロジーを整備した。

1) LinkDB システムの統合利用環境への応用

LinkDB は DB エントリー間の関係を定義した DB である。その統合利用環境への応用として、まず研究代表者グループと協力し、RDF 化を実現した。具体的には、RDF 記述形式の一つである Turtle と N-triple を用いて、LinkDB の全データ (2014 年 2 月で約 25 億リンク) を RDF 化した。この形式を検索結果からダウンロードできるようにするとともに、日次更新においても RDF 化に対応した。さらに、RDF 質問言語の一種である SPARQL で検索できるようにした。具体的には、Virtuoso と Stanza を使ったインタフェースを構築した。

ウェブ検索に関しては、REST 形式での LinkDB API も整備するとともに、データベース関連図から二つのデータベースを指定してリンク情報を取得する機能を公開した。また、LinkDB で用いている各種データベースのリンク用 URI リストを整備した。LinkDB 検索対象の拡張に関して、統合プロジェクトの PGDBj や RIKEN のバイオリソースなどの国内外の遺伝子、化合物データベースとのリンクを増やした。

2) メタゲノム・メタメタボローム等新規分野データ活用技術の開発

遺伝子機能自動アノテーションサーバ KAAS (KEGG Automatic Annotation Server) において、メタゲノムアノテーションで必要となる生物種情報を考慮するようにした。その結果、アノテーション精度が向上した。これを用いて、日本人腸内細菌メタゲノム、欧州 META-HIT Project、米国 Human Microbiome Project のデータをアノテーションし KEGG MGENES として公開した。また、KAAS の結果から KEGG MODULE などを用いて機能アノテーションを支援するツールを開発し、MAPLE (Metabolic and physiological potential evaluator) として公開した。

機能未知遺伝子も含めて KEGG GENES に登録された全遺伝子を自動でオーソログ分類した KEGG OC を公開した。また、オーソロググループを含む複数のゲノムレベルデータを統合して、遺伝子ネットワークを予測するツール GENIES を整備し、サンプルデータを増やすなどして、使い勝手を向上させた。

反応オントロジーの整備で開発した反応パターンデータベース RCLASS を用いて、新規反応予測のための方法論とシステムを開発した。具体的には、メタボロームデータから得られる多数の化合物間を結ぶ反応経路を予測するための方法論を開発した。また、基質と生成物のペアから酵素分類番号を予測する E-zyne を拡張し、OC で定義されたオーソログクラスまで予測できるようにするとともに、反応パターンの表現方法も改良した。これらを応用することにより、メタメタボローム情報とメタゲノム情報との統合が可能となる。

3) 反応オントロジーの整備

官能基の変換パターンや反応中心の化学構造に基づく反応オントロジーを開発し公開するとともに、化合物アライメントの結果から類似反応を検索できるようにした。また、その応用として類似パスウェイ検索ツール PathSearch を開発した。さらに、パスウェイ上で保存された反応モジュールの抽出に応用し、その成果は KEGG RMODULE データベースとして公開された。基質—生成物ペアのデータベース RPAIR、反応パターンデータベース RCLASS を継続的に更新して整備している。

RCLASS における反応パターンの定義には、反応中心回りの数原子を用いているだけであるが、新規反応予測や類似反応検索には、それ以外の原子情報も用いた方がよいことが多い。そこで反応予測に有効な新しい化合物構造のフィンガープリント KCF-S (KEGG Chemical Function and Substructure) を開発し、反応経路予測や E-zyne に応用した。

§ 5 成果発表等

(1)原著論文発表 (国内(和文)誌 0 件、国際(欧文)誌 12 件)

1. 著者、論文タイトル、掲載誌 巻、号、発行年、DOI

1. Toshiaki Katayama, Mark D. Wilkinson, Kiyoko F. Aoki-Kinoshita, Shuichi Kawashima, Yasunori Yamamoto, Atsuko Yamaguchi, Shinobu Okamoto, Shin Kawano, Jin-Dong Kim, Yue Wang, Hongyan Wu, Yoshinobu Kano, Hiromasa Ono, Hidemasa Bono, Simon Kocbek, Jan Aerts, Yukie Akune, Erick Antezana, Kazuharu Arakawa, Bruno Aranda, Joachim Baran, Jerven Bolleman, Raoul JP Bonnal, Pier Luigi Buttigieg, Matthew P Campbell, Yi-an Chen, Hirokazu Chiba, Peter JA Cock, Kevin B Cohen, Alexandru Constantin, Geraint Duck, Michel Dumontier, Takatomo Fujisawa, Toyofumi Fujiwara, Naohisa Goto, Robert Hoehndorf, Yoshinobu Igarashi, Hidetoshi Itaya, Maori Ito, Wataru cIwasaki, Matúš Kalaš, Takeo Katoda, Taehong Kim, Anna Kokubu, Yusuke Komiyama, Masaaki Kotera, Camille Laibe, Hilmar Lapp, Thomas Lütteke, M. Scott Marshall, Takaaki Mori, Hiroshi Mori, Mizuki Morita, Katsuhiko Murakami, Mitsuteru Nakao, Hisashi Narimatsu, Hiroyo Nishide, Yosuke Nishimura, Johan Nystrom-Persson, Soichi Ogishima, Yasunobu Okamura, Shujiro Okuda, Kazuki Oshita, Nicki H Packer, Pjotr Prins, Rene Ranzinger, Philippe Rocca-Serra, Susanna Sansone, Hiromichi Sawaki, Sung-Ho Shin, Andrea Splendiani, Francesco Strozzi, Shu Tadaka, Philip Toukach, Ikuo Uchiyama, Masahito Umezaki, Rutger Vos, Patricia L Whetzel, Issaku Yamada, Chisato Yamasaki, Riu Yamashita, William S York, Christian M Zmasek, Shoko Kawamoto, and Toshihisa Takagi. BioHackathon series in 2011 and 2012: penetration of ontology and Linked Data in life science domains. *Journal of Biomedical Semantics*, 5:5, 2014. (PMID: 24495517) (doi: 10.1186/2041-1480-5-5)
2. Kiyoko F Aoki-Kinoshita, Jerven Bolleman, Matthew P Campbell, Shin Kawano, Jin-Dong Kim, Thomas Lütteke, Masaaki Matsubara, Shujiro Okuda, Rene Ranzinger, Hiromichi Sawaki, Toshihide Shikanai, Daisuke Shinmachi, Yoshinori Suzuki, Philip Toukach, Issaku Yamada, Nicolle H Packer, and Hisashi Narimatsu. Introducing glycomics data into the Semantic Web. *Journal of Biomedical Semantics*, 4:39, 2013. (PMID: 24280648) (doi: 10.1186/2041-1480-4-39)
3. Takeru Nakazato, Tazro Ohta, Hidemasa Bono. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One*. 8(10):e77910, 2013 (PMID: 24167589) (doi: 10.1371/journal.pone.0077910)
4. Hiroko Tabunoki, Hiromasa Ono, Hiroaki Ode, Kazuhiro Ishikawa, Natsuki Kawana, Yutaka Banno, Toru Shimada, Yuki Nakamura, Kimiko Yamamoto, Jun-ichi Satoh, Hidemasa Bono. Identification of key uric acid synthesis pathway in a unique mutant silkworm *Bombyx mori* model of Parkinson's disease. *PLoS One*. 8(7):e69130, 2013 (PMID: 23894418) (doi: 10.1371/journal.pone.0069130)
5. Kiyoko F Aoki-Kinoshita, Hiromichi Sawaki, Hyun Joo An, Jin Won Cho, Daniel Hsu, Masaki Kato, Shin Kawano, Toshisuke Kawasaki, Kay-Hooi Khoo, Jaehan Kim, Jin-Dong Kim, Xianyu Li, Thomas Lütteke, Shujiro Okuda, Nicolle H Packer, James C Paulson, Rahul Raman, René Ranzinger, Huali Shen, Toshihide Shikanai, Issaku Yamada, Pengyuan Yang, Yoshiki Yamaguchi, Wantao Ying, Jong Shin Yoo, Yang Zhang, and Hisashi Narimatsu. The Third ACGG-DB Meeting Report: Towards an international collaborative infrastructure for glycobioinformatics. *Glycobiology*, 23(2), 144-146, 2013. (PMID: 23271684) (doi: 10.1093/glycob/cws167)
6. Mio Hirayama, Daiki Kobayashi, Souhei Mizuguchi, Takashi Morikawa, Megumi

- Nagayama, Uichi Midorikawa, Masayo M. Wilson, Akiko N. Nambu, Akiyasu C. Yoshizawa, Shin Kawano, and Norie Araki. Integrated proteomics identified novel activation of dynein IC2-GR-COX-1 signaling in NF1 disease model cells. *Molecular & Cellular Proteomics*, 12(5), 1377-1394, 2013. (PMID: 23358504) (doi: 10.1074/mcp.M112.024802)
7. Takatomo Fujisawa, Shinobu Okamoto, Toshiaki Katayama, Mitsuteru Nakao, Hidehisa Yoshimura, Hiromi Kajiya-Kanegae, Sumiko Yamamoto, Chiyoko Yano, Yuka Yanaka, Hiroko Maita, Takakazu Kaneko, Satoshi Tabata, Yasukazu Nakamura. CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Research*. 1:42(1):D666-70., 2014. (PMID: 24275496) (doi: 10.1093/nar/gkt1145)
 8. S. Möller, E. Afgan, M. Banck, P. J. A. Cock, M. Kalas, L. Kajan, P. Prins, J. Quinn, O. Sallou, F. Strozzi, T. Seemann, A. Tille, R. V. Guimera, T. Katayama, and B. Chapman, "Sprints, Hackathons and Codefests as community gluons in computational biology," *EMBNet.journal*, vol. 19, pp. 40–42, 2013.
 9. Satoshi Yamasaki, Takatsugu Hirokawa, Kiyoshi Asai, Kazuhiko Fukui, Tertiary Structure Prediction of RNA-RNA Complexes Using a Secondary Structure and Fragment-Based Method, *J. Chem. Inf. Model.*, 2014, 54 (2), pp 672-682, DOI: 10.1021/ci400525t
 10. Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Toshiaki Tokimatsu and Susumu Goto, "Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets", *Bioinformatics*, vol. 29, pp.i135-i144, 2013 (PMID:23812977) (DOI:10.1093/bioinformatics/btt244)
 11. Hiroaki Iwata, Sayaka Mizutani, Yasuo Tabei, Masaaki Kotera, Susumu Goto, Yoshihiro Yamanishi, "Inferring protein domains associated with drug side effects based on drug-target interaction network", *BMC Syst. Biol.* vol. 7(S6), S18, 2013 (DOI:10.1186/1752-0509-7-S6-S18)
 12. Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Yuki Moriya, Toshiaki Tokimatsu, Minoru Kanehisa and Susumu Goto, "KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics", *BMC Syst. Biol.*, vol. 7(S6), S2, 2013 (DOI:10.1186/1752-0509-7-S6-S2).

(2)その他の著作物(総説、書籍など)

1. 河野信、HUPO-PSI(Proteomics Standards Initiative)ミーティング参加に関するレポート、日本プロテオーム学会通信 No.150、2013年8月22日 (<http://www.jhupo.org/letter/pdf/No150.pdf>)
2. 飯田啓介. 新しい日本語 Web コンテンツ, 「新着論文レビュー」と「領域融合レビュー」. 情報管理, 56, 148-155 (2013) doi: 10.1241/johokanri.56.148 (<http://dx.doi.org/10.1241/johokanri.56.148>)

(3)国際学会発表及び主要な国内学会発表

- ① 招待講演(国内会議 4件、国際会議 1件)
(主要な国際会議への招待講演の前に*を付記してください)

1. 飯田啓介、新しい日本語 Web コンテンツ、「新着論文レビュー」と「領域融合レビュー」、2013年7月度(第314回)OUG ライフサイエンス分科会、文京シビックセンター、7/18/2013
2. *Toshiaki Katayama、International collaborations for semantic data integration and interoperability、Bioinformatics Week in Odaiba (BiWO 2013)産業技術総合研究所生命情報工学研究センター、9/11/2013~9/13/2013
3. 坊農秀雅、生命科学分野の大規模データ利用技術開発の現状と今後の展開、第3回 SPARC Japan セミナー 2013、国立情報学研究所、10/25/2013
4. 小寺正明、化学構造に基づいたメタボロームスケール酵素反応ネットワーク予測、第8回メタボロームシンポジウム、福岡、2013年10月4日
5. 小寺正明、昆虫種横断検索データベースの開発、第36回日本分子生物学会年会、神戸、2013年12月4日

② 口頭発表(国内会議 19件、国際会議 7件)

1. Shin Kawano、Database Integration toward Semantic Web: Developments and Applications of RDF databases、HUPO-PSI meeting 2013、University of Liverpool (UK)、2013年4月15日-2013年4月17日
2. Shin Kawano、Database Integration Activities in Japan、ProteomeXchange meeting 2013、University of Liverpool (UK)、2013年4月18日-2013年4月19日
3. 川島秀一、岡本忍、微生物表現型オントロジーおよび LOD の開発、第27回人工知能学会年会、富山国際会議場、2013年6月4日-2013年6月7日
4. 藤原 豊史、山本 泰智、生命科学分野の学術文献情報からなるリンクデータの構築、第27回人工知能学会年会、富山国際会議場、2013年6月4日-2013年6月7日
5. Shuichi Kawashima、Integration of Microbial Data Resources with Semantic Web Technologies、BioHackathon 2013 symposium、2013年6月23日
6. 川島秀一、微生物データ統合を目指したオントロジーの開発 MCCV と MPO の紹介、ラン藻ゲノム交流会、東京大学駒場キャンパス、2013年7月13日
7. 山本泰智、岡本忍、宮沢せいは、市川夏子、藤田信之、テキスト処理技術を用いたキュレーション支援、ラン藻ゲノム交流会、東京大学駒場キャンパス、2013年7月13日
8. 岡本忍、片山俊明、川島秀一、藤澤貴智、TogoGenome/TogoStanza: データ統合と再利用性向上のためのゲノム情報デザイン、ラン藻ゲノム交流会 2013、東京大学駒場キャンパス、2013年7月13日
9. Toshiaki Katayama, Piotr Prins, Raoul Bonnal, Francesco Strozzi, Naohisa Goto、BioRuby updates - power of modularity in the community-based open source development model、The 14th Annual Bioinformatics Open Source Conference (BOSC 2013)、Berlin, Germany、2013年7月19日-2013年7月20日
10. Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Toshiaki Tokimatsu and Susumu Goto、Supervised learning of enzymatic reaction likeness for de novo reconstruction of metabolic pathways、International Workshop on Bioinformatics and Systems Biology 2013、2013年7月31日-2013年8月2日
11. Tazro Ohta, Hiromasa Ono, Yuki Naito, Takeru Nakazato, Hidemasa Bono、Technology development of database integration to make sense of big data in lifescience、QMB 2013、Rydges Hotel Queenstown, New Zealand、2013年8月25日-2013年8月30日
12. 小林大樹、平山未央、水口惣平、森川崇、長山慈、緑川宇一、ウィルソン-森藤政代、南部-新堀晶子、吉沢明康、河野信、荒木令江、融合プロテオミクスによる神経線維腫症 1 型 (NF1) 病態モデル細胞内の活性化シグナル Dynein IC2-GR-COX-1 経路の同定、第86回日本生化学会大会、パシフィコ横浜、2013年9月11日-2013年9月13日
13. 片山俊明、データベース統合の実現に向けて 基盤技術開発 (I)、トーゴーの日シンポジウム 2013、時事通信ホール、2013年10月4日-2013年10月5日

14. 岡本 忍、データベース統合の実現に向けて 基盤技術開発 (II)、トーゴの日シンポジウム 2013、時事通信ホール、2013年10月4日-2013年10月5日
 15. Toshiaki Katayama, Takatomo Fujisawa, Hiroshi Mori, Shuichi Kawashima, Shinobu Okamoto, TogoGenome: Semantic Web-based genome database、生命医薬情報学連合大会 (JSBi 2013)、タワーホール船堀、2013年10月29日-2013年10月31日
 16. Masaaki Kotera, Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets、JSBi2013 年会、東京、2013年10月29日
 17. Lucky Runtuwene, Shuichi Kawashima, Kaori Noguchi, Eri Kawakami, Akinori Tokunaga, Yutaka Suzuki, Takashi Kobayashi, Yuki Eshita, Potential Novel Anti-viral Proteins in *Aedes aegypti*、第 66 回日本寄生虫学会南日本支部大会、第 63 回日本衛生動物学会南日本支部大会 合同大会、大分大学医学部、2013年11月2日-2013年11月3日
 18. 片山俊明、ライフサイエンスにおけるセマンティック・ウェブのアプリケーション開発、バイオインフォマティクスセミナー in 新潟、新潟大学駅南キャンパスときめいと、2013年11月13日、
 19. 川島秀一、生命科学データの RDF 化、バイオインフォマティクスセミナー in 新潟、新潟大学駅南キャンパスときめいと、2013年11月13日
 20. 飯田啓介、新たな日本語 Web コンテンツ「ライフサイエンス 新着論文レビュー」、第 36 回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
 21. Masaaki Kotera, アブイニシオ代謝経路再構築へ向けた反応ネットワーク予測、「整合性マシナリー:生物活性物質構造多様性創出システムの解明と制御」第 6 回公開シンポジウム、千葉大学、2013年12月6日
 22. Shinobu Okamoto, Shuichi Kawashima, Takatomo Fujisawa and Toshiaki Katayama, TogoStanza: JavaScript framework for reusable data visualization in the biological context, "The 5th International Symposium on Languages in Biology and Medicine", 東京大学福武ホール、2013年12月12日-2013年12月13日
 23. 坊農秀雅、公共遺伝子発現データベースからの集合知による低酸素応答遺伝子探索、第 11 回がんとハイポキシア研究会、東北大学片平さくらホール、2013年12月13日-2013年12月14日
 24. 河野信、TogoTable: RDF を利用したアプリケーション、理研・糖鎖インフォマティクス若手の会合同セミナー、理化学研究所、2014年2月12日
 25. 山口敦子、古崎晃司、呉紅艶、小林紀郎、SPARQL Builder: 生物学研究者が SPARQL を使いこなすための補助ツール、第 32 回セマンティック・ウェブとオントロジー研究会、オリーブアン小豆島、2014年3月5日-2014年3月6日
 26. Lucky R. Runtuwene, Shuichi Kawashima, Kaori Noguchi, Eri Kawakami, Akinori Tokunaga, Yutaka Suzuki, Sumio Sugano, Kenta Nakai, Ryuichiro Maeda, Chihiro Sugimoto, Tomohiko Takasaki, Ichiro Kurane, Takashi Kobayashi and Yuki Eshita, Utilization of CCL-125 for analyzing innate immunity in *Aedes aegypti*、第 66 回日本衛生動物学会大会、岐阜大学、2014年3月21日-2014年3月23日
- ③ ポスター発表(国内会議 28 件、国際会議 13 件)
発表者(所属)、タイトル、学会名、場所、月日
1. Yasunori Yamamoto, Shinobu Okamoto, Seiha Miyazawa, Natsuko Ichikawa, Shuji Yamazaki, Nobuyuki Fujita, Curation Assistance System Using Natural Language Processing Technologies、Biocuration 2013、2013年4月7日-2013年4月10日
 2. Takatomo Fujisawa, Toshiaki Katayama, Shinobu Okamoto, Hiroshi Mori, Yasunori Yamamoto, Yasukazu Nakamura, Development of ontologies for the INSDC / DDBJ resource、The 15th Genomic Standards Consortium meeting、Bethesda, MD (USA)、2013年4月22日-2013年4月23日

3. Toshiaki Tokimatsu, Classification of Plant Type III Polyketide Synthase-related Secondary Metabolites for Predicting Their Biosynthetic Pathways, International Workshop on Bioinformatics and Systems Biology 2013, 京都, 2013年7月31日-2013年8月2日
4. Masaaki Kotera, Supervised Learning of Enzymatic Reaction Likeness for de novo Reconstruction of Metabolic Pathways, International Workshop on Bioinformatics and Systems Biology 2013, 京都, 2013年7月31日-2013年8月2日
5. Yuki Moriya, KEGG OC: Automatically Constructed Comprehensive Ortholog Clusters Based on Taxonomic Relation, International Workshop on Bioinformatics and Systems Biology 2013, 京都, 2013年7月31日-2013年8月2日
6. Ai Muto, Reaction Modules: Conserved Sequences of Reactions in Metabolic Pathways, International Workshop on Bioinformatics and Systems Biology 2013, 京都, 2013年7月31日-2013年8月2日
7. 川島秀一, セマンティック・ウェブ技術を利用した遺伝子発現解析データの解釈について, 第3回 NGS 現場の会, 神戸ポートアイランド, 2013年9月4日-2013年9月5日
8. 片山俊明, Joachim Baran, 癸生川絵里, BaseSpace Ruby SDK の開発, 第3回 NGS 現場の会, 神戸ポートアイランド, 2013年9月4日-2013年9月5日
9. Shin Kawano, Akiyasu C. Yoshizawa, Sohei Mizuguchi, and Norie Araki, iPEACH: integrated Protein/gene Expression Analysis CHart, HUPO 12th Annual World Congress, パシフィコ横浜, 2013年9月14日-2013年9月18日
10. Daiki Kobayashi, Mio Hirayama, Souhei Mizuguchi, Takashi Morikawa, Megumi Nagayama, Uichi Midorikawa, Masayo Wilson-Morifuji, Akiko Nambu-Niibori, Akiyasu C. Yoshizawa, Shin Kawano, and Norie Araki, Integrated proteomics identified novel activation of dynein IC2-GR-COX-1 signaling by suppression of NF1 tumor suppressor gene product, neurofibromin, in neuronal cells, HUPO 12th Annual World Congress, パシフィコ横浜, 2013年9月14日-2013年9月18日
11. Norie Araki, Takashi Morikawa, Souhei Mizuguchi, Daiki Kobayashi, Akiko Niibori Nambu, Uichi Midorikawa, Mio Hirayama, Masayo Morifuji Wilson, Shin Kawano, Hideo Nakamura, and Junichi Kuratsu, An Integrated Proteomics for Extracting Molecular Target of Malignant Gliomas, HUPO 12th Annual World Congress, パシフィコ横浜, 2013年9月14日-2013年9月18日
12. 金森茜, 稲塚歩佳, 門之園哲哉, 口丸高弘, 坊農秀雅, 近藤科江, 低酸素下での EGFR 活性化によって引き起こされる小胞体ストレスにおける MMP10 の機能, 第72回日本癌学会学術総会, パシフィコ横浜, 2013年10月3日-2013年10月5日
13. 河野信, 渡辺敦, 山口敦子, TogoTable update 2013, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
14. 坊農秀雅, 小野浩雅, 天竺桂弘子, カイコ機能アノテーションパイプラインの構築とその応用, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
15. 飯田啓介, 新しい日本語 Web コンテンツ, 「新着論文レビュー」と「領域融合レビュー」, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
16. 小野浩雅, 坊農秀雅, 遺伝子発現リファレンスデータセット RefEx, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
17. 川島秀一, 岡本忍, 片山俊明, 藤澤貴智, 森宙史, 微生物データ統合を目指したオントロジーの開発, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
18. 藤原 豊史, 山本 泰智, 論文の引用情報検索システム Colil, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年10月5日
19. 山本泰智, 岡本忍, 宮沢せいは, 市川夏子, 藤田信之, テキスト処理技術を用いたキュレーション支援, トーゴーの日シンポジウム 2013, 時事通信ホール, 2013年10月4日-2013年

10月5日

20. 山口敦子、大久保克彦、呉紅艶、榎田達矢、島中秀樹、セマンティック・ウェブ技術を用いた生命科学系データベースアーカイブのデータ統合、トーゴの日シンポジウム 2013、時事通信ホール、2013年10月4日-2013年10月5日
21. 福井 一彦、田代 俊行、矢葺 幸光、浅井 潔、解析ワークフローとデータベースの連携基盤構築、トーゴの日シンポジウム 2013、時事通信ホール、2013年10月4日-2013年10月5日
22. 時松敏明、ゲノムネットにおけるデータ統合と新規分野データ開発のための基盤技術開発の最新状況、トーゴの日シンポジウム 2013、時事通信ホール、2013年10月4日-2013年10月5日
23. Masaaki Kotera, Development of a Cross-species Search Database for Insect Genes, JSBi 2013 年会、2013年10月29日
24. Hiromasa Ono, Hidemasa Bono, RefEx—Reference expression dataset for tissue transcriptome, Genome Informatics 2013, Cold Spring Harbor Laboratory, 2013年10月30日-2013年11月2日
25. Susumu Goto, Identification of Metabolic Pathway Modules by Conserved Reaction Sequences and Its Application to Evolutionary Analysis, 第51回日本生物物理学会年会、京都、2013年10月30日
26. Hidemasa Bono, Tazro Ohta, Hiromasa Ono, Yuki Naito, Takeru Nakazato, TECHNOLOGY DEVELOPMENT OF DATABASE INTEGRATION TO MAKE FULL USE OF BIG DATA IN LIFESCIENCE, Functional Genomics and Systems Biology, WellcomeTrust Genome Campus, Cambridge, UK, 2013年11月21日-2013年11月23日
27. 岡本 忍、川島秀一、藤澤貴智、片山俊明、TogoGenome: RDFによる拡張可能なセマンティック・ゲノムデータベース、ラン藻の分子生物学 2013、かずさアカデミアホール、2013年11月22日-2013年11月23日
28. 川島秀一、岡本忍、微生物表現型オントロジーの開発、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
29. 河野信、渡辺敦、山口敦子、TogoTable update 2013:セマンティック・ウェブを利用したテーブル形式データへの自動アノテーション付加ツール、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
30. 坊農秀雅、仲里猛留、内藤雄樹、小野浩雅、大田達郎、統合牧場: 次世代生命科学研究室の実験場 #togofarm - a pilot farm towards the next generation research “stable” in lifescience, 第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
31. 飯田啓介、新しい日本語 Web コンテンツ、「新着論文レビュー」と「領域融合レビュー」、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
32. 小野浩雅、坊農秀雅、RefEx: 遺伝子発現解析の基準となる公共データの活用、整理および共有、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
33. 山本泰智、岡本忍、宮沢せいはい、市川夏子、藤田信之、テキスト処理技術を用いたゲノムアノテーション支援システム、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
34. 藤澤 貴智、森 宙史、岡本 忍、山本 泰智、片山 俊明、川島 秀一、谷澤 靖洋、神沼 英里、大山 彰、菅原 秀明、内山 郁夫、黒川 颯、中村 保一、Genome Refine: ゲノムアノテーションの精緻化と解析支援のためのサービス、第36回日本分子生物学会年会、神戸ポートアイランド、2013年12月3日-2013年12月6日
35. 片山俊明、藤澤貴智、森宙史、川島秀一、岡本忍、TogoGenome: セマンティック・ウェブによる次世代の統合ゲノムデータベース、第36回日本分子生物学会年会、神戸ポートアイラン

ド、2013年12月3日-2013年12月6日

36. 岡本 忍、西村邦裕、Genomic Landscape、第36回日本分子生物学会年会 サイエンスとアートの接点、神戸ポートアイランド、2013年12月3日-2013年12月6日
37. Toshiaki Katayama, Masataka Kikuchi, Soichi Ogishima、Interactive platform for semantic gene expression analysis of Alzheimer's disease、Semantic Web Applications and Tools for Life Science (SWAT4LS 2013)、Royal Society of Edinburgh, England、2013年12月9日-2013年12月11日
38. Shinobu Okamoto, Shuichi Kawashima, Takatomo Fujisawa and Toshiaki Katayama、TogoStanza: JavaScript framework for reusable data visualization in the biological context、"The 5th International Symposium
39. on Languages in Biology and Medicine"、東京大学福武ホール、2013年12月12日-2013年12月13日
40. Shuichi Kawashima, Toshiaki Katayama, Toshihisa Takagi, Shinobu Okamoto、MPO: Microbial Phenotype Ontology for Comparative Genome Analysis、"The 5th International Symposium
41. on Languages in Biology and Medicine"、東京大学福武ホール、2013年12月12日-2013年12月13日
42. 坊農秀雅、公共遺伝子発現データベースからの集合知による低酸素応答遺伝子探索、第11回がんとハイポキシア研究会、東北大学片平さくらホール、2013年12月13日-2013年12月14日
43. 小野浩雅、坊農秀雅、公共遺伝子発現データベースからの集合知による低酸素応答遺伝子探索、第11回がんとハイポキシア研究会、東北大学片平さくらホール、2013年12月13日-2013年12月14日

(4)知財出願

- ① 国内出願 (0件)
- ② 外出願 (0件)
- ③ その他の知的財産権
なし

(5)受賞・報道等

- ① 受賞(顕著な受賞の前に*を付記してください)
なし
- ② マスコミ(新聞・TV等)報道
なし
- ③ その他
なし

§6 研究開発期間中に主催した会議等

主なワークショップ、シンポジウム、アウトリーチ等の活動

年月日	名称	場所	参加人数	概要
H25年6月 23日-28日	NBDC/DBCLS BioHackathon 2013	DBCLS	77人	生命科学分野におけるセマンティック・ウェブ技術の研究開発を目的に、DBやオントロジー、アプリケーションの開発と標準化を各国の実務者が参加し合宿形式で開発を行う国際会議。NBDCと共催
H26年1月 27日-31日	国内版バイオハッカソン BH13.13	沖縄県名護市、万国津梁館	61人	国内のライフサイエンスデータベースの構築者と利用者向けの技術勉強会
H25年12月 12日-13日	The 5th International Symposium on Languages in Biology and Medicine (LBM 2013)	東京大学	43名	医学生物学の自然言語処理、オントロジー、セマンティック・ウェブなど、言語を中心とした研究の国際会議
H25年4月 17日-18日	SPARQLthon7	DBCLS	14人	生命科学分野におけるセマンティック・ウェブ技術に興味のある者が自由に参加し、オントロジーの構築・RDFの設計と生成の促進とSPARQL検索技術、可視化技術等の向上を目指すハッカソン
H25年5月 23日-24日	SPARQLthon8	DBCLS	14人	同上
H25年6月3日-4日	SPARQLthon9	DBCLS	10人	同上
H25年7月 30日-31日	SPARQLthon10	DBCLS	15人	同上
H25年8月 20日-21日	SPARQLthon11	DBCLS	16人	同上
H25年9月 17日-18日	SPARQLthon12	DBCLS	13人	同上
H25年10月 23日-24日	SPARQLthon13	DBCLS	17人	同上
H25年11月 25日-26日	SPARQLthon14	DBCLS	23人	同上
H25年12月 16日-17日	SPARQLthon15	DBCLS	16人	同上
H26年1月 16日-17日	SPARQLthon16	DBCLS	13人	同上
H26年2月 17日-18日	SPARQLthon17	DBCLS	17人	同上
H26年3月	SPARQLthon18	DBCLS		同上

H26年1月 22日-23日	AJACS 肥後	化学及血清 療法研究所	40人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年11月 6日	AJACS 蝦夷 3	北海道大学	20人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年8月 30日	AJACS 富山	富山大学	30人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年7月 30日-31日	AJACS 琉球	琉球大学	40人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年7月 12日	AJACS 岐阜	岐阜大学	30人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年5月 28日	AJACS 筑波 3	物質・材料 研究機構	30人	NBDC 主催 DBCLS 共催と して講師を派遣した
H25年5月 15日	統合データベース講演 会	東京理科大 学	100人	大学生を対象に講習会を開 催した
H25年11月 18日	塩基配列データアーカ イブをフル活用するた めの大規模データ解析技 術開発研究会	国立遺伝学 研究所	24人	DBCLS と DDBJ がさらに 連携を深めて物理的にも一 体となっていくにあたり、中 長期的な方向性について自 由に議論
H26年02月 14日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H26年01月 10日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年12月 26日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年11月 28日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年10月 24日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年09月 26日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年09月 12日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年07月 25日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年05月 23日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H25年4月 17日	チーム内ミーティング (非公開)	DBCLS	30人	研究進捗報告のためのミー ティング
H26年10月 31日	オープンサイエンスアワ ード	タワーホー ル船堀	50人	日本バイオインフォマティク ス学会の DBCLS 主催セッ ションで優秀な国産バイオ インフォマティクスのソフトウ ェア、データベース、ウェブサ イトについて投票を行い、入 賞作品について講演と表彰 を行った。

H16年10月31日	BioHack コンペティション	タワーホール船堀	50人	日本バイオインフォマティクス学会のDBCLS主催セッションで短期間のソフトウェア開発を公募し、優秀作品について講演と表彰を行った。
H26年3月20-21日	第18回オープンバイオ研究会	北陸先端科学技術大学院大学		バイオインフォマティクスにおけるオープンソースソフトウェアの技術開発に関する研究会。人工知能学会分子生物情報研究会と共催。
2013年5月22日	基盤技術開発プログラム進捗状況報告会(非公開)	DBCLS	3名	基盤技術開発プログラムに参加しているメンバーが集い、H24年度までに開発した研究内容の説明、及びH25年度に開発予定であるツール、DB、システム等について進捗報告を行った。
2013年7月2日	情報統合ワークフロー(非公開)	DBCLS	4名	DBCLSが開発中であるワークフローGALAXYについて進捗報告を行った。本会議では、二つの研究機関の連携可能な技術の洗い出しを行い、セマンティックウェブ技術におけるSADI技術やその有効利用を議論した。
2013年9月2日	情報統合ワークフロー(非公開)	DBCLS	6名	基盤技術の連携のため、我々が開発したSADI解析技術、京都大学が開発しているsimcomp及びDBCLSが開発中であるワークフローGALAXYによる3つの研究機関の連携解析について進捗報告を行った。
2013年9月19日	統合シンポ打ち合わせ(非公開)	DBCLS	5名	産総研SADI解析技術、京都大学が開発しているsimcomp及びDBCLSが開発しているワークフローGALAXYによる3つの研究機関の化合物解析の連携技術について進捗報告及び動作確認を行った。
2014年1月5日	情報統合進捗(非公開)	DBCLS	3名	基盤技術の連携のため、我々が開発したSADI解析技術やDBCLSが開発中であるワークフローを用いた研究機関の連携解析について進捗報告を行った。

2013年4月11日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年4月30日	チーム内ミーティング (非公開)	京都大学化学研究所	8人	研究進捗報告のためのミーティング
2013年5月23日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年6月6日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年6月26日	チーム内ミーティング (非公開)	京都大学化学研究所	8人	研究進捗報告のためのミーティング
2013年7月12日	チーム内ミーティング (非公開)	京都大学化学研究所	8人	研究進捗報告のためのミーティング
2013年7月26日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年8月23日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年9月6日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年9月26日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年10月11日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年11月1日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年11月14日	チーム内ミーティング (非公開)	京都大学化学研究所	10人	研究進捗報告のためのミーティング
2013年11月29日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年12月13日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2013年12月27日	チーム内ミーティング (非公開)	京都大学化学研究所	9人	研究進捗報告のためのミーティング
2014年1月29日	チーム内ミーティング (非公開)	京都大学化学研究所	7人	研究進捗報告のためのミーティング
2014年2月19日	チーム内ミーティング (非公開)	京都大学化学研究所	7人	研究進捗報告のためのミーティング

§ 7 ユーザー評価結果への対応

≪平成25年7月に実施した「NBDCにおける事業活動のユーザー評価」(<http://biosciencedbc.jp/user-hyouka-2013/result-summary>)で得られたユーザーの意見、提案等(詳細は別紙2を参照)に対し、実施済み若しくは実施予定の対応策等を具体的に記載してください。)

①実施済み

○TogoStanza

コメント:データを利用するコミュニティの造成。ユーザーと開発者を結びつける場がないと、いくら

良いツールでも発展して行かないように思います。事例が増えれば、他のユーザーも興味を持てるので、まずは具体的な課題を持っている研究者を開発グループと結びつけるのはどうでしょうか。
回答:コメントありがとうございます。ご指摘の通り、個別のユースケースを増やすことで実際に利用していただけるユーザーを増やす努力をしていきたいと考えております。具体的には、現在、微生物研究コミュニティ、菌株保存機関と密に連携をはかり、微生物の表現型や培地といった、今まで網羅的にDB化されることのなかった情報を蓄積しております。これらの情報をゲノム情報と合わせて可視化する **TogoStanza** を開発いたしました。今後、より広い生物種と生物学的な視点を取り入れて行きたいと考えております。

○生命科学文献アノテーションワークベンチ PubAnnotation

①実施済み

コメント:ローカルインストールできるなら利用したい。実働できる VM イメージなどの提供を希望します。

回答:ご提案有難うございます。ローカルインストールができるようにソースコードを公開しました:
<https://github.com/pubannotation/pubannotation>。まだドキュメンテーションが足りないですが、日本語のインストールガイドを含め充実して行きたいと思います。VM イメージの公開も進めて行きます。

②実施予定

コメント:利用する場面が想像できない。

回答:ご指摘有難うございます。色んな場面での使い方に関してのチュートリアルを作成し公開する等、実際に役に立つサービスに向け頑張っって改善して行きたいと思います。

コメント:折角日本語対応にしているのですから、説明も日本語があった方が良いと思います。文献からのデータマイニングが容易になりそうな期待感はあるので、開発者目線では無く、実際にいるんな人にあえて使ってもらって、改善していく方が良いと思いました。

回答:ご指摘有難うございます。日本語の説明、ガイドラインを充実にして行きながら、色んな場面での使い方に関してのチュートリアルも作成し公開して、情報不足で使えないことが起きないように頑張っって改善して行きます

コメント:なるほどとは思いますが、実際に使うことがあるか疑問です。

回答:好意的なコメントとご指摘ありがとうございます。色んな場面での使い方に関してのもっと工夫をして、実際に役に立つサービスに向け頑張っって改善して行きたいと思います。

コメント:使い方がわからない。基本的にはほとんどのユーザーは **PubMed** を見続けると思うので、このサイトに誘導するような仕組み(**bookmarklet** など?)が必要かと。

回答:ご指摘有難うございます。実ユーザーに役に立つシステムに向け改善して行きたいと思ます。日本語の説明、ガイドラインを充実にして行きながら、色んな場面での使い方に関してのチュートリアルも作成し公開して、情報不足で使えないことが起きないように頑張っって改善して行きます。**Bookmarklet** は素晴らしいアイデアだと思います。それを含め、他の関連サイトからも楽に辿り着ける方法を工夫して実現して行きたいと思ます。

コメント:面白そうで、かつ、役に立ちそうなプロジェクトだけに、使い方がわかりにくいのがとても惜しまれる。トップページから直感的に使えるようにしてくれると、きっと多くの人が使ってくれると思ます。「スケーラブルかつ共有可能なテキストアノテーションストレージ」と言われても一般の人にはなんのことやらわからないと思ます。

回答:好意的なコメントありがとうございます。それから使い方がよく分からない問題点に対してのご指摘も有難うございます。実ユーザーの目線からを考えて使いやすいサービスとして頑張っって改善して行きたいと思ます。

コメント:ローカルインストールできるなら利用したい。実働できる VM イメージなどの提供を希望します。

回答:ご提案有難うございます。ローカルインストールが楽に出きるように、インストールガイドラインの作成、VM イメージの公開を進めて行きます。

コメント:結果の表示が間延びして見づらいように感じてしまう。トップページからのユーザーナビゲートも直感でないように感じた。

回答:適切なご指摘有難うございます。実ユーザーの目線をもっと考え、直感的に使いやすいサービスに向け、頑張って改善して行きたいと思えます。

コメント:使い方に関して、素人には具体的なイメージが持てない。もう少し分野外のユーザーにもフレンドリーな HP が望まれる。

回答:適切なご指摘有難うございます。実ユーザーの目線をもっと考え、直感的に使いやすいサービスに向け、頑張って改善して行きたいと思えます。色んな場面での使い方に関してのチュートリアルも作成し公開して、情報不足で使えないことが起きないように改善して行きたいとおもいます。

○プラットフォーム型 Active ワークフロー

コメント:ほとんどの人が「何ができるのかわからない」と感じるのでは。

回答:CBRC では、よりユーザーに分かりやすくするため、ワークフロー上の各解析ツールの処理記述やエラーハンドリングを HP に公開した。また SADI サービスでの SIO オントロジーを用いた解析処理や SPARQL 検索に関する HELP 情報を追加した。

○データベース間のリンク情報 LinkDB

コメント:RDF でダウンロードできるのなら素晴らしいです。ぜひ利用したいです。

回答:API での取得もできるようにしました。

○統合遺伝子検索 GGRNA

コメント:出力形式を選べると、なおよいと思えます。

回答:タブ区切りテキスト形式または JSON 形式での出力に対応しました。

コメント:生物種の拡充。

回答:NCBI RefSeq の転写産物データベースに収録された全生物種を検索できるようにしました。

コメント:同義語の問題が解決していないのでは? cyclooxygenase ではヒットがあったが、"prostaglandin endoperoxide synthetase"ではヒットがなかった。これはどうでもいいことですが、名称や説明が遺伝子検索なので、遺伝子だけが出て来るのかと思ったら、mRNA も検索結果に出て来ているので、結局何を検索しているのかがよくわからなかった。

回答:NCBI RefSeq の転写産物を対象に検索しており、ヘルプページに記載しております。同義語でも検索できますが、遺伝子によってはアノテーションが充実していないために、ヒットしない場合があります。

②実施予定

○TogoStanza

コメント:まず、TogoGenome が何を目的として作られているのか分かりませんでした。検索窓や遺伝子リストがあるだけでは、これまでに利用している DDBJ/EMBL/GenBank や UniProtKB 等を使うので良いと感じてしまいます。統合サービスは初心者に対して分かりやすくないと、いけないと思えますので(経験者はある程度自分で DB をお気に入りに入っていますから)見栄えよりも使いやすさ、分かりやすさに重点を置いた方が良く思えます。

回答:率直なご指摘ありがとうございます。RDF データによるデータ統合の利点を明確に出す事により既存の DB との違いがユーザーに分かりやすくなるように、よりいっそう努力していきたいと考えております。

コメント:方向性は賛成するが、UCSC ゲノムブラウザなどと比べて、情報の integration や可視化の点で見劣りがする。

回答:ご指摘ありがとうございます。今後、国際的に RDF 化が進んでいる EBI のデータ群や Ensembl のリファレンスゲノム情報、また USCS で豊富に統合されているオミクス実験情報などもデータを RDF 化し統合することで取り込んで行く努力をしていきたいと考えております。

コメント:実際に検索してみると、検索候補はでてくるのですが、検索結果が No matching records found になるのが多いです。データ拡充はこれからだとおもうので頑張ってくださいたいです。

回答:好意的なコメントありがとうございます。ご指摘の通り、まだまだデータの網羅性が低い部分がございますので、今後データの拡充を図っていききたいと考えております。

コメント:実際の研究者がどれだけ使うか疑問です。

回答:率直なご指摘ありがとうございます。ご指摘の通り、現段階では、実験研究者が個別研究のために利用するには情報の精度と網羅性が基準に達していないことは痛感しております。一方で、TogoStanza システムは、実験生物学者だけでなく、データベース作成者、計算生物学者に対して、繰り返し作成されるデータベースのパーツを簡便に提供したり、統計解析のためのデータセットを低コストで所得できるという利便性を提供することも目指しております。今後、それぞれの利用者に、本質的な利便性を提供できるようにバランス良く開発を進めて行きたいと考えております。今後ともよろしくお願いたします。

コメント:セマンティック・ウェブ技術によるデータ統合は大事な技術の一つだと思うが、見せ方として、私が欲しい情報が抽出され、アクセスできる状況になっているのか、不明だった。あるいは、どんな情報が用意されているのか不明だった。

回答:有用なご指摘ありがとうございます。ご指摘の通り、今回の開発をとおしてセマンティック・ウェブ技術では、RDF 形式によるデータ統合は有効である一方、個々の RDF データモデルやオントロジーなどが複雑になり、データの全体像が見渡しづらいという問題点が見えてきました。今後、TogoStanza の可視化技術等を使って、統合されているデータの全体像を俯瞰できるようなアプリケーションも提供していきたいと考えております。

コメント:どう使えばいいのかわかりません。思いついたキーワードを入れてもヒットしないことが多いです。

回答:率直なご指摘ありがとうございます。RDF データのフリーワードによる検索アプリケーションは別途開発しております。TogoStanza と TogoGenome の背後にはセマンティック・ウェブ技術を利用しているので、より意味的な検索を提供したいと考えております。一方で、セマンティック・ウェブ技術自体がこの 3 年間で急速に発展した技術であるため、世界的にみても技術的なノウハウの蓄積がなく、様々な技術的な課題が山積している現実です。今後、これらの課題を解決しながら、フリーワード検索への対応を進めて行きたいと考えております。

コメント:今後の展開に期待します。

回答:好意的なコメントありがとうございます。有用なサービスになるように今後とも努力してまいります。

コメント:公開後はマニュアルや実例を充実させて欲しい。ライフサイエンスだけに留めておくにはもったいない技術と考える。科学全般に波及させることはできるか？

回答:我々のサービスに、ご期待いただきありがとうございます。ご指摘の通り、今後ヘルプページや実例の充実をはかりつつ、より広いユーザー層に利用していただけるサービスへの発展を目指して行きたいと考えております。

コメント:使い方に関して、素人には具体的なイメージが持てない。もう少し分野外のユーザーにもフレンドリーな HP が望まれる。

回答:率直なご指摘ありがとうございます。ご指摘の通り、今後、ヘルプ等の文章を充実させて利用しやすいウェブアプリケーションを目指したいと考えております。今後ともよろしく願いいたします。

○生命科学文献アノテーションワークベンチ PubAnnotation

○プラットフォーム型 Active ワークフロー

○データベース間のリンク情報 LinkDB

コメント:入力に、データベースに加えて、エントリー名が必要なのが使いづらいように感じた。

回答:データベース名を指定できるインターフェースも用意していましたが、エントリー名が必要なインターフェースに比べて目立ちませんでした。次回の更新で工夫する予定です。

コメント:KEGG にたくさんのリンクをつけただけ

回答:KEGG 以外のデータベース間のリンクも検索できるようにしていますが、より充実させたいと思います。

○統合遺伝子検索 GGRNA

コメント:検索結果の絞り込み機能。

回答:ファセット検索機能の追加を検討しています。

コメント:Galaxy から利用できるツールを Galaxy Tool Shed に登録していただけると助かります。

回答:登録を検討しています。

§ 8 その他

(1) 研究代表者として、研究開発、プロジェクト運営等について、上記以外に報告したいことがあれば、自由に記載してください。

なし

以上