

ライフサイエンスデータベース統合推進事業  
統合化推進プログラム  
研究開発課題「生命と環境のフェノーム統合データ  
ベース」

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：豊田哲郎  
(理化学研究所・情報基盤センター・統合データ  
ベース特別ユニット、ユニットリーダー)



## § 1 研究開発実施の概要

### (1) 実施概要

#### 1. 「フェノタイプ記述子の体系化とプロパティの標準化」

NBDCの推進するRDF技術を用いた体系的なデータ記述を行うために、データ統合化の基礎となる識別子の体系化を行なう目的で、生物の表現型を網羅的に記述できる汎用フォーマットを開発した。表現型を示す部位、生物学的特徴の分類である形質、各形質において生物が具体的に示す特徴(定性値あるいは定量値)を、国内外で使用されているものに共通のオントロジー、またはデータベース値を代入するよう設計した。これにより、表現型データの多様性が、3つのフィールドそれぞれのオントロジーの分類体系によって吸収されるため、極めて自由度の高い記述が可能であり、様々な生物の多様な表現型を、1つのデータベースに格納する事、およびデータを分散させながらも、ひとつデータベースとして統合的に扱う事が可能となる。

NBDCで提供している `purl.jp` システムに合わせて、我々の `biolod.org` の URL をマッピングし直し、ダウンロードできるすべての RDF ファイルを `purl.jp` に変換して提供した。こうして作られた RDF データを RIKENBASE として公開するとともに NBDC に提供した。

#### 2. 「フェノーム統合化」

各種バイオリソースデータ収集のプラットフォームを構築し、マウス 5000 系統、細胞株 3600 系統、微生物株 15000 系統、シロイヌナズナ 3700 件に関連する表現型データを合計 27000 件収集し、一部に関しては常に最新情報に更新できる体制とした。ずれのデータセットも、フェノタイプ情報にそれぞれ適切なオントロジーターム、TAIR の遺伝子情報、PosMed の文献レコードなどへのリンクを付加して、統合データベース RIKENBASE に統合化・公開された。また、後述の SWAS システムにより、次世代シーケンサー (NGS) のデータなどから生成した遺伝子リストを対象とした相関解析も可能となった。

#### 3. 「フェノーム利用ワークフロー開発」

RDF で記述されたマウス表現型データを分かりやすく配信するためのインターフェース開発を行った。

データ共有の規格として近年注目されている LOD(Linked Open Data)に準拠した標準形式で生命科学関連の公開データを提供する目的で、BioLOD.org (Biological Linked Open Databases)を開発し、このようなデータにアクセスするためのプログラミングインターフェイスとして、semantic-JSON を新規に開発した。

また、ユーザがデータ相互の関係性を把握し必要な情報を検索して取り出すことを容易化するための支援システムとして、BioSPARQL(バイオスパークル; Broadly Integrated Ontological SPARQL Protocol and RDF Query Language)を開発した。

関連情報が網羅的につながった RDF のネットワークの閲覧画面として、まずはシンプルなカード型の画面によって最低限の情報を見せておき、関連情報ボタンをクリックすることにより、関連情報を次々と開いて参照できるようにした。また、体系立った RDF のリンクを用いて、関連する表現型の情報を自動検索するインターフェース (お勧め機能) も開発した。これを用いて、実験材料を検索する生物学者が、Amazon オンラインショップを操作するような感覚で、関連する表現型もつ材料を次々と閲覧し、自分の研究にふさわしい材料を選びやすくするようにした。

以前に開発した検索システム PosMed を統合して、Semantic Web の特性を活用した大規模相関解析によって SWAS (Semantic-Web Association Study)を実現した。

## § 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

本課題では、「フェノーム統合データベース」と、「フェノーム利用ワークフロー」によって、種や個体の違いを越えて同一の計測技術を適用可能である事を提案したり、種横断的な表現型の体系的な分類により分野の壁をこえて技術情報を共有する仕組みを提供する事を目指して、大まかに以下の3つに取り組む。すなわち、1)フェノタイプ記述子の体系化として、データ記述に用いるプロパティを標準化し、識別子を体系的に整理する。また、2)フェノーム統合化として、バイオリソースフェノームの統合化と、先端計測データフェノームの統合化を行う。さらに、3)フェノーム利用ワークフロー開発として、上記の標準化された体系的データベースを統合的に利用するためのシステムを開発する。

#### 1. 「フェノタイプ記述子の体系化」

##### 1-1. プロパティ標準化と評価(豊田グループ)

データ間の関係性を属性として表す「プロパティ」は、統合化の対象となる個別データベースごとにばらばらなものが用いられることから、これらを国内外で使用されているものに共通化させていくことで、データ統合化の基礎となるプロパティの標準化を行う

##### 1-2. 識別子体系化と評価(柘屋グループ)

体系的なデータ記述を行うために必要な概念の関係性を表す「オントロジー」やその「インスタンス(リソースなど)」は、統合化の対象となる個別データベースごとにばらばらなものが用いられていたり欠けていたりすることから、これらを国内外で使用されているものに共通化させていくことで、データ統合化の基礎となる識別子の体系化を行う

#### 2. 「フェノーム統合化」

##### 2-1. バイオリソースフェノーム(柘屋グループ)

主にバイオリソースに関連付けられるフェノームについて、上記のフェノタイプ記述子に対応付けながら情報の収集と整理を行い、マウス系統、細胞株、微生物株、植物株などの表現型情報や有用性情報を統合化する。研究材料としてのバイオリソースの共通利用により、再利用可能な知見の蓄積を促進する

##### 2-2. 先端計測データフェノーム(豊田グループ)

主にバイオリソースに関連付けられないフェノームについて、上記のフェノタイプ記述子に対応付けながら情報の収集と整理を行い、文献記載のフェノームデータに加えて、生命動態システムなどの先端研究分野で生み出される新しいタイプのデータを統合化する

#### 3. 「フェノーム利用ワークフロー開発」(豊田グループ)

フェノーム利用ワークフローを開発し、上記のように標準化されたデータを統合的に利用するためのシステムを開発する。複雑な RDF 等の仕様を知らない生物学研究者にもセマンティックウェブデータを活用できる機会を提供するとともに、現在のセマンティックウェブ技術の主流となっている「SPARQL エンドポイント法」に依存したデータ提供方法の欠点を補完することでセマンティックウェブデータ利用環境の発展を図る。

### (2) 新たに追加・修正など変更した研究開発構想

特に、(3)「フェノーム利用ワークフロー開発」において、構想の変更を行った。

1年目には、RDF へのアクセス方法として、SPARQL を用いる方向を考えており、1年目には、セマンティックウェブデータにアクセスするためのプログラミングインターフェイス、semantic-JSON の開発、ユーザがデータ相互の関係性を把握し必要な情報を検索して取り

出すことを容易化するための支援システムとして、BioSPARQL(バイオスパークル; Broadly Integrated Ontological SPARQL Protocol and RDF Query Language)の開発を行なった。しかし、当時は SPARQL に検索で期待通りのパフォーマンスが得られなかったため、2年目以降、より高速でユーザの望む情報が得られるような方法論を模索した。その結果、以前に開発・公開した PosMed を統合して RDF データに適用できるよう改良することで、構想を変更し、高速性、セキュリティへの配慮などの点に優れた大規模相関解析を実現できることが分かり、このシステムを利用した、データ検索のインターフェースを提供する事とした。結果として、RDF で作られたセマンティック Web データを用いた大規模相関解析 SWAS (Semantic-Web Association Study)システムの開発を実現した。

### (3) 達成状況

研究開発項目	H23年度	H24年度	H25年度	変更点
1. フェノタイプ記述子の体系化 ・プロパティ標準化と評価(豊田) ・識別子体系化と評価(梶屋)	←-----→ ←-----→			
2. フェノーム統合化 ・バイオリソースフェノーム(梶屋) ・先端計測データフェノーム(豊田)	←-----→ ←-----→			
3. フェノーム利用ワークフロー開発 ・インタフェース開発(豊田) ・連携支援システム開発(豊田) ・ダウンロードサイト ・推論検索機能の普及化	←-----→ ←-----→	←-----→ ←-----→	←-----→ ←-----→	早期完了 新規追加
4. NBDC への成果の受け渡し			←-----→ ←-----→	追加

## 1. 「フェノタイプ記述子の体系化」

### 1-1. プロパティ標準化と評価

標準プロパティとしては、NBDC で提供している purl.jp システムに合わせることで、我々の biolod.org の URL をマッピングし直し、ダウンロードできるすべての RDF ファイルを purl.jp に変換して提供した。また、統合対象のフェノームデータは当初は独自に定義したプロパティセットを利用した標準化されていたが、システムバイオロジーのポータル

サイト **BioGateway** の開発グループによって提供されている標準プロパティセットである **biorel** がオミックスデータ間の関連性を表現するのに有用であることが分かり、これらのプロパティへの対応付けを進めることにした。現在までに、**BRC** 細胞リソースデータベース、**BRC** 表現型アノテーションデータベース、文献キュレーションによる植物フェノームデータベースのプロパティ標準化作業を完了した。

## 1-2. 識別子体系化と評価

生物の表現型を網羅的に記述できる汎用フォーマット (**RDF** モデル) を開発した。**RDF** は、「主語」「述語」「目的語」の3つ組を基本構造とする。我々が提案する表現型の **RDF** モデルでは、表現型データを主語として、表現型を示す部位、生物学的特徴の分類である形質、各形質において生物が具体的に示す特徴 (定性値あるいは定量値) という3種類の述語を持ち、それぞれの述語に、国内外で使用されているものに共通のオントロジー、またはデータベース値を代入するよう設計した。これらの主語、述語、目的語それぞれについて、**RDF** の標準に従って **ID** を付し、同じく **RDF** によるデータ統合を推進する国際的な生物オントロジーの標準化コンソーシアム、**Open Biomedical Ontology (OBO)** コンソーシアムの提案する形式との相互関係 (概念間の親子関係) を定義した。これにより、本形式は、柔軟な表現力を持ちながら、**OBO** との互換性を保つことができる。また、細胞株に関しては、細胞株の国際的な相互利用を目指し、ミシガン大学、**Sanger** 研究所、**EBI** 等との連携により、**Cell line Ontology (CLO)** コンソーシアムを設立し、**RDF** 準拠の標準スキーマを開発した。

## 2. 「フェノーム統合化」

### 2-1. バイオリソースフェノーム

バイオリソースに関連付けられるフェノームについて、上記のフェノタイプ記述子に対応付けながら情報の収集と整理を行い、マウス系統、細胞株、微生物株、植物株などの表現型情報や有用性情報を統合化するために、「基盤データ収集のための、バイオリソースデータ収集プラットフォーム(マウス)構築」を構築した。これを用いて、マウス 5000 系統、細胞株 3600 系統、微生物株 15000 系統および、表現型データ 18000 件、遺伝子等関連情報 50000 件を収集し、上記 **RDF** フォーマットに落とし込むとともに、月ごとに最新情報に更新できる体制とした。微生物表現型に関しては、**DBCLS** で策定中の、**Microbial Phenotype Ontology (MPO)** による特性記述に準拠した記述を行なった。これらの情報は、下記に述べるフェノーム利用のために開発したインターフェース(画面)を用いて、配信を行なっている。

### 2-2. 先端計測データフェノーム

シロイヌナズナフェノームについて、以下の2種類のデータセットを統合化した。ひとつは、理研内で開発されたシロイヌナズナ変異株、もう一つは、文献キュレーションにより収集したシロイヌナズナの表現型変異情報である。前項で述べた標準化方針に基づいてそれぞれ標準化し、フェノーム統合データベースを新たに作成して収納した。2014年2月時点で収納されたフェノーム情報は、変異株フェノームが 3763 件、文献フェノームが 823 件である。いずれのデータセットも、フェノタイプ情報にそれぞれ適切なオントロジーターム、**TAIR** の遺伝子情報、**PosMed** の文献レコードなどへのリンクを付加して、統合データベース **RIKENBASE** に統合化・公開された。

また、当プロジェクトの成果を生かした連携システムとして、統合化されたシロイヌナズナ・マウス・細胞系統の遺伝子・発現プロファイルなどのデータベースを利用して人工プロモーターの配列設計ができる **web** アプリである「**promoter CAD**」を開発した。

## 3. 「フェノーム利用ワークフロー開発」

**BioLOD.org (Biological Linked Open Data)** を開設し、データ共有の規格として近年注目されている **LOD (Linked Open Data)** に準拠した標準形式で生命科学関連の公開データ提供を行なっている。

こうして統合されたデータにアクセスする為の論理は **BioLOD.org** によって提供される。さらに、統

合データベースプロジェクトを通じて統合化されたデータをはじめとするセマンティックウェブデータに探索的にアクセスするためのプログラミングインターフェイス semantic-JSON を新規に開発した。このインターフェースを介して、公開・非公開データの複雑なつながりをアクセス権に応じて探索できる。

一方、ユーザがデータ相互の関係性を把握し必要な情報を検索して取り出すことを容易化するための支援システムとして、BioSPARQL(バイオスパークル; Broadly Integrated Ontological SPARQL Protocol and RDF Query Language)を開発した。BioSPARQL(バイオスパークル; Broadly Integrated Ontological SPARQL Protocol and RDF Query Language)は、RDF/OWL データ構造を解析して適切な SPARQL クエリを容易に構築するためのフレームワークである。既存のクエリ生成補助ツールとは異なり、BioSPARQL は RDF/OWL データを論理的に解析して、適切なクエリのひな形をグラフィカルにユーザーに提示する。生成されたテンプレートに必要なに応じてキーワードを入力することで、ユーザーは実行可能な SPARQL クエリを容易に得られる。

さらに、セマンティックウェブベースのデータを利用して、高速性、セキュリティへの配慮などの点に優れた大規模相関解析を実現する方法を模索し、結果として SWAS (Semantic-Web Association Study)システムの開発を実現した。これは、以前に開発・公開した PosMed を統合して RDF データに適用できるよう改良したものである。現在、ゲノム領域を対象として遺伝情報の相違を検定して病気の原因遺伝子等を見つけ出す GWAS という手法が盛んに利用されているが、SWAS ではこの発想をセマンティックウェブデータに適用し、研究対象の表現型などのキーワードと相関の高いデータ群を検定により見つけ出す。以前から、フェノームを起点として多様なオミックスデータを横断的に検索するシステムを整備する必要性は研究コミュニティで指摘されており、SWAS 解析システムを整備はそのニーズを実現するものと考えている。

また、RDF で記述されたマウス表現型データを分かりやすく配信するためのインターフェース開発を行った。RDF のデータは、関連情報が網羅的につながったネットワーク型を示すが、これをシンプルに表を用いたカード型の画面によって、最初に最低限の情報を見せておき、関連情報ボタンをクリックすることにより、関連情報を次々と開いて参照できるようにした。また、RDF のリンクを用いて、関連情報を自動検索するインターフェースも開発した。これを用いて、1つのマウス系統から関連する表現型を示す、Amazon のオンラインショップを模した「お勧め機能」を作成し、ユーザーの利便をはかった。

#### (4) 研究開発の今後の展開について

本プロジェクトで開発したシステムは、フェノタイプの膨大な文献とバイオリソースを含む幅広いリソースを、セマンティックウェブ技術を用いて統合し、「つながり検索」によって蓄積する知見を高度に活用する可能性を広げた。RDF 型データを活用することで「分かりやすい」データ表現を実現し、異常の起こる部位や内容による柔軟な検索が可能となった。同時に、高いオープン性、相互運用性をもつシステムが構築できた。これらのことは、国際連携による研究の促進にも役立つと考えている。フェノームを起点として多様なオミックスデータを横断的に検索するシステムの整備の必要性は研究コミュニティでも指摘されており、本プロジェクトの成果はその要請に応えるものとも言える。また、ユーザー評価に寄せられたコメントからも、デモンストレーション会場における来場者との会話でも、細胞株を扱う研究者の当プロジェクトへの関心が高いことが伺え、細胞リソースのフェノタイプ情報の需要が今後見込まれる。細胞株の国際的な相互利用では、Cell line Ontology (CLO)を通した、国際的な細胞レポジトリ連携や、標準化などが期待される。

### § 3 研究開発実施体制

(1) 研究チームの体制について

①「研究代表者:豊田 哲郎」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
豊田 哲郎	理化学研究所・情報基盤センター・統合データベース特別ユニット	ユニットリーダー	プロパティ標準化と評価、先端計測データフェノーム、フェノーム利用ワークフロー開発	H23.4～H26.3
小林 紀郎	同上	センター上級研究員	フェノーム利用ワークフロー開発	H23.4～H26.3
蒔田 由布子	同上	センター研究員	先端計測データフェノーム、フェノーム利用ワークフロー開発	H23.4～H25.12
吉田 有子	同上	センター技師	フェノーム利用ワークフロー開発	H23.4～H25.9
西方 公郎	同上	センター研究員	先端計測データフェノーム、フェノーム利用ワークフロー開発	H23.4～H26.3
土井 考爾	同上	センター研究員	プロパティ標準化と評価、先端計測データフェノーム	H23.4～H26.3
高橋 聡史	同上	テクニカルスタッフ I	先端計測データフェノーム、フェノーム利用ワークフロー開発	H23.4～H26.3
川口 修司	同上	センター研究員	先端計測データフェノーム	H23.4～H26.3
下山 紗代子	同上	リサーチアソシエイト	プロパティ標準化と評価	H23.4～H26.3
望月 芳樹	同上	センター技師	先端計測データフェノーム、フェノーム利用ワークフロー開発	H23.4～H25.12
石井 学	同上	テクニカルスタッフ I	フェノーム利用ワークフロー開発	H23.4～H26.3

松嶋 明宏	同上	テクニカルスタッフ I	フェノーム利用 ワークフロー開 発	H23.4～H26.3
飯田 慶	京都大学大学院医学研究科医学研究 支援センター	特定助教	先端計測デー タフェノーム	H23.4～H26.3
GIFFORD DAVID JONATHAN	理化学研究所・情報 基盤センター・統合 データベース特別ユ ニット	テクニカルスタッフ I	フェノーム利用 ワークフロー開 発	H23.4～H26.3

## ②「研究分担者:梶屋 啓志」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
梶屋 啓志	理化学研究所・バイオリ ソースセンター・マウス 表現型知識化研究開発 ユニット	ユニットリーダー	データ解析・ア ノテーション・デ ータベース設 計	H23.4～ H26.3
田中 信彦	同上	開発研究員	データ構造の 評価	H23.4～ H26.3
脇 和規	同上	テクニカルスタッ フ I	データ変換、作 成	H23.4～ H26.1
高月 照江	同上	テクニカルスタッ フ II	データ変換、作 成	H23.4～ H26.3
斎藤 実香子	同上	パートタイマー	データ変換、作 成	H23.4～ H26.3
松庫 義弘	同上	パートタイマー	アノテーション、 データ変換、作 成	H24.4～ H25.9
高山 英紀	同上	パートタイマー	アノテーション、 データ変換、作 成	H25.7～ 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

- 細胞株の国際的な相互利用を目指し、ミシガン大学、Sanger 研究所、EBI 等との連携により、Cell line Ontology (CLO)コンソーシアムを設立し、RDF 準拠の標準スキーマを開発した。
- 微生物フェノームについては黒川グループと基本スキーマを共有し、特性情報の標準化と最新リソース情報とゲノム・メタゲノム情報の連結を、連携して進めている。
- マウスフェノームについては、アジアマウス変異リソース連合 (Asian Mouse Mutagenesis and Resource Association: AMMRA) 国際マウス表現型解析コンソーシアム (International Mouse Phenotyping Consortium: IMPC) との連携により、RDF ベースの情報共有について議論を行っている。
- NBDC の研究開発プログラム「生命動態システム科学のデータベースの統合化」とは、生命動態と表現型情報の相互利用について議論し、生命動態プロジェクトのデータをフェノームデータとして、本プロジェクトで扱う試みを行なっている。
- 植物フェノームについては、シロイヌナズナ変異体の表現型・遺伝子の変異情報を統合



化しているが、これらの情報は、理研植物科学センターに所属するそれぞれの変異株の開発者から提供を受けたものである。

- また、**promoterCAD** の開発も理研植物科学センターバイオマス工学連携部門 合成ゲノミクス研究チームとの協同によって実現した。
- シロイヌナズナのオミックス情報の国際的な共有化を主導している **The International Arabidopsis Informatics Consortium** は文献・フェノタイプ・バイオリソース間の検索システムの整備を重視しており、その意味から我々が開発した **PosMed** を高く評価している。本事業では、**PosMed** をさらに発展させた **SWAS** により、フェノームを中心としたさらに効果的な検索システムが実現しており、このような研究コミュニティからの要請にこたえる成果となっている。

## § 4 研究実施内容及び成果

### 4.1 研究課題名: RDF モデルによる大規模フェノーム統合 (独立行政法人理化学研究所・情報基盤センター・統合データベース特別ユニット)

#### 研究開発実施内容及び成果

##### 課題1「フェノタイプ記述子の体系化:プロパティ標準化と評価」

RDF での体系的なデータ記述を行うためのプロパティを国内外で使用されているものに共通化させ、URLの標準化を進めた。具体的には、まず、NBDCで提供している **purl.jp** システムに合わせて、我々の **biolod.org** の URL をマッピングし直し、ダウンロードできるすべての RDF ファイルを **purl.jp** に変換して提供した。

現時点で世界の研究機関が様々な標準プロパティセットを公開しているが、それぞれを精査したところ、システムバイオロジーのポータルサイト **BioGateway** の開発グループによって提供されている標準プロパティセットである **biorel** がオミックスデータに対応する標準プロパティセットとして最も内容が充実しており、フェノームデータと関連オミックスデータ間の関連性を表現するのに有用であることが分かった。また、**biorel** は、**Cell Cycle Ontology (CCO)**、**Gene Expression Knowledge Base (GeXKB)**などの公開サイトで既に使われている。そこで、これまでに統合化を進めてきたフェノームデータを対象に、これまで独自に定義して使用してきたプロパティを **biorel** のプロパティに対応付ける作業を行い、現在までに、**BRC 細胞リソースデータベース**、**BRC 表現型アノテーションデータベース**、**文献キュレーションによる植物フェノームデータベース**のそれぞれについて、標準化を完了した。また、標準化されたプロパティを用いてデータのアップロードができるシステムを開発し、組織間でのデータ受け渡しに利用できる環境を構築した。

##### 課題2「フェノーム統合化:先端計測データフェノーム」

文献キュレーションによって収集したシロイヌナズナの様々な表現型変異に関する記述を一つ一つキュレーションし、2014年2月の段階で、823件の表現型情報を6種類の公開オントロジー (**PO**, **GO**, **TO**, **ChEBI**, **PATO**, **EO**)を用いて標準化し統合化した。並行して、理研内で開発されたシロイヌナズナのバイオリソースとして、シロイヌナズナ変異体 3763種類の表現型情報 14631件をオントロジーによって標準化した上で統合化し、それぞれ「理研シロイヌナズナ変異株フェノームデータベース」「文献由来植物フェノームデータベース」として公開した。さらに、これらのシロイヌナズナフェノーム情報をとりまとめる「シロイヌナズナフェノームデータベース」を新設し、GUIを工夫して、データベースの内容をより簡単に概観できるようにするとともに、共通的なインターフェースでバイオリソースフェノーム・文献フェノーム双方の検索が容易に行えるようにした。また、植物の部位に関する標準語彙である **Plant Ontology (PO)**の語彙体系に準拠した2階層構造の検索メニューシステムを新設した。本データベースでは、収録されているフェノタイプ情報のそれぞれが **PO** の term へリンクするよう RDF を構築しているため、このような階層メニューを導入することで、たとえば「葉の異常」に該当する表現型を幅広く概観することも、「ロゼット葉が細長い」といった、より限定

的な表現型情報を見ることも、簡単にできるようになった。

### 課題3「フェノーム利用ワークフロー開発」

データ共有の規格として近年注目されている LOD(Linked Open Data)に準拠した標準形式で生命科学関連の公開データを提供する目的で、BioLOD.org (Biological Linked Open Data)を開発した。本サイトの検索機能は、結果を様々な観点からランク付けして示す仕様となっており、目的に適した良質なデータセットをユーザーが容易にダウンロードできる。また、データは様々な形式でダウンロードでき、生物情報学的研究への適用も容易である。2014年2月現在で、216 データベース、1 千万件のインスタンスのバイオデータが統合されている。

また、こうして統合データベースプロジェクトを通じて統合化されたデータをはじめとするセマンティックウェブデータにアクセスするためのプログラミングインターフェイスとして、semantic-JSONを新規に開発した。このインターフェイスを介して、公開・非公開データの複雑なつながりをアクセス権に応じて探索できる。

一方、ユーザーがデータ相互の関係性を把握し必要な情報を検索して取り出すことを容易化するための支援システムとして、BioSPARQL(バイオスパークル; Broadly Integrated Ontological SPARQL Protocol and RDF Query Language)を開発した。本システムは、BioLOD で扱っている RDF/OWL データ構造を解析して適切な SPARQL クエリ(検索対象となるテーブルやデータの抽出条件、並べ方などを指定する文字列)を構築し提供するためのフレームワークである。

さらに、以前に開発した検索システム PosMed を統合して、Semantic Web の特性を活用した大規模相関解析を実現し、これを SWAS (Semantic-Web Association Study)と名付けた。現在、ゲノム領域を対象として遺伝情報の相違を検定して病気の原因遺伝子等を見つけ出す GWAS という手法が盛んに利用されているが、SWAS ではこの発想をセマンティックウェブデータに適用し、研究対象の表現型などのキーワードと相関の高い遺伝子などのデータ群を統計検定により高速に見つけ出す。本システムは、Medline, OMIM, パスウェイ情報、共発現情報や、統合化されたバイオリソースデータなどを「ドキュメントデータ」として利用する。検索キーワードとドキュメント間の対応表を作成し、一方で、遺伝子と文書の対応づけデータを用意し、検索キーワードと統計的に有意な関連性の認められる遺伝子やバイオリソースを高速に推論検索することができる。また、それぞれの研究に有用なバイオリソースを、たとえ直接的にフェノタイプ情報が付随していなくても推論検索によって見出すことができるため、バイオリソースの有用性を高めることにもつながる。

平成 25 年度はさらに PosMed システムを更新し、ファイルアップロード機能を追加した。これにより本システムは、次世代シーケンサー(NGS)から生み出されるデータに基づく関連性解析にも適用できるようになった。NGS の解析から得られる遺伝子のリスト(たとえば癌細胞と正常細胞の比較から発現量に変化が見られた遺伝子のリスト)を TSV もしくはエクセル形式のファイルとして用意する。このファイルを PosMed システム上にアップロードし、所定のキーワード(たとえば“cancer”)を指定して検索を試みると、フェノタイプ情報を利用した推論検索により、変異遺伝子とフェノタイプ情報が推論的に結び付けられ、リストされた中から高い関連性が示唆される遺伝子が提示される。このようにして、NGS データに基づいて表現型変異の原因遺伝子を in silico で絞り込むことが可能となる。

平成 25 年度には、マウスフェノームデータ、シロイヌナズナフェノームデータなどを最新版に置き換え、Mammalian Phenotype Ontology, Plant Ontology, 国際疾病分類(ICD-10)などのデータを新たに統合化し推論検索に使えるようにした。

また、最終年度は、統合化されたフェノームと関連情報をより効率よく検索し、わかりやすく提示するためのシステム改良を進めた。RDF で記述されたマウス表現型データを分かりやすく配信するためのインターフェイスとして、関連情報を自動検索するインターフェイスも開発した。RDF のデータは関連情報が網羅的につながったネットワークを構成しているが、まずはシンプルに表を用いたカード型の画面によって最低限の情報を見せておき、関連情報ボタンをクリックすることにより、関連情報を次々と開いて参照できるようにした。また、RDF のリンクを用いて、関連情報を自動検索するインターフェイスも開発した。これを用いて、1つのマウス系統から関連する表現型を示す、Amazon のオンラインショップを模した「お勧め機能」を作成した。

同じ仕組みを用いて、シロイヌナズナバイオリソースデータにおいても、変異体を検索した際に、ヒットした変異体と同一の遺伝子がノックアウトあるいは強制発現している他の変異体が同時にリストアップされるようにした。これにより、表現型情報から直接検索することが難しいが研究に役立つ可能性の高いシロイヌナズナ変異体(具体的な観察情報が付随していない変異体も多数あるのが現状である)を効率よく探索できるようになった。

さらに当プロジェクトの成果を生かした連携システムとして、遺伝子発現する組織やタイミングの調節を可能にする人工プロモーターの配列設計ができる web アプリである「**promoter CAD**」を開発した。**Promoter CAD** では、統合化されたシロイヌナズナ・マウス・細胞系統の遺伝子・発現プロファイルなどのデータベースが利用され、利用者の設定した遺伝子発現条件に応じて、関連性が強いと予想されるモチーフ配列を組み合わせる配列設計を行う。研究者はプログラミングなどの専門知識がなくてもこのアプリを容易に利用でき、また、自由にデータを追加したり、プログラムをカスタマイズすることもできる。

以上のシステムについて継続的に GUI の改善とシステムのパフォーマンス改善をすすめ、チュートリアル文書も整備して当該ページからアクセスできるようにした。また、平成 25 年 5 月に開催された第 60 回日本実験動物学会総会、6 月にオーストラリアで開催された 25th International Conference on Arabidopsis Research、12 月に神戸で開催された第 36 回日本分子生物学会年会などでデモンストレーションを行った。これらを通じて、フェノーム統合データベースおよび関連システムの有用性や具体的な応用方法などの研究者への周知を図った。

#### 課題 4 「NBDC への成果の受け渡し」

本事業で統合化されたデータは現時点ですべて NBDC が策定した pur1.jp に対応する形で RDF 化して公開しており、外部に容易に移設可能である。これをもって NBDC 側に成果を受け渡しできる体制が整った。

#### 4. 2 研究課題名: バイオリソース・フェノーム統合化

(理化学研究所バイオリソースセンター・マウス表現型知識化研究開発ユニット)

研究開発実施内容及び成果

#### 課題1「フェノタイプ記述子の体系化: 識別子体系化と評価

平成 23 年度は、「上位オントロジーに基づくフェノタイプ識別子の、OWL による体系化」として、生物に普遍的な表現型データ形式を、国内で独自開発された上位オントロジー、**Yet Another More Advanced Top-level Ontology (YAMATO)** および、バイオ分野で広く使われている **Basic Formal Ontology (BFO)** それぞれの下位概念として定義した。このデータ形式策定にあたっては、ドメインオントロジーとして、**OBO コンソーシアム** のオントロジーを主に採用し、統合化プログラムの各プロジェクトとの議論を行い、相互運用性を確保する方向で調整を進めた。「識別子体系化作業用プラットフォームの開発」として、バイオリソースの表現型を、上記のドメインオントロジーを用いてアノテーションするシステムを作成し、最終的に **BioLOD** へエクスポートする **OWL 化ワークフロー** を確立した。

平成 24 年度は、このワークフローを用いて、特にマウスを対象として、実質的な表現型データ作成を行なった。リソースデータベース、および関連論文を参照するキュレーション作業を行い、「識別子体系化作業」として、マウスの表現型データ約 1000 件、細胞の特性データ約 5300 件を作成し、平成 25 年度は、微生物データ 14250 件と、マウス特性データ 10129 件を追加した。細胞株に関しては、細胞株の国際的な相互利用を目指し、ミシガン大学、**Sanger 研究所**、**EBI** 等との連携により、**Cell line Ontology (CLO)** コンソーシアムを設立し、**RDF 準拠の標準スキーマ**を開発した。

平成 25 年度は、微生物データ 14250 件と、マウス特性データ 10129 件を追加するとともに、**DBCLS** および、黒川グループの提案する微生物リソースの基本情報を記述する **RDF フォーマット**

ト、**Microbial Culture Collection Vocabulary (MCCV)**の評価を行い、これを用いてデータ記述を開始した。また、微生物表現型を記述する**Microbial Phenotype Ontology (MPO)**についても評価を行い、本プロジェクトにおける記述との相互運用性を確保しながら、これを用いた表現型データ記述を行なうことを決定した。

## 課題2 「フェノーム統合化：バイオリソースフェノーム」

平成 23 年度は、「基盤データ収集のための、バイオリソースデータ収集プラットフォーム（マウス）構築」として、マウス用プラットフォーム構築を完了し、月毎更新ワークフローを確立した。

平成 24 年度は、バイオリソースデータ収集プラットフォームを培養細胞についても拡張した、「基盤データ収集のための、バイオリソースデータ収集プラットフォーム（細胞株）」を作成し、これを用いて、「細胞株リソースデータ収集作業」として、約 3600 株の情報を収集、データベース化し、月ごとに最新情報を配信している。マウスにおいても月ごとのデータ更新を続行するとともに、識別子体系化作業において作成した表現型データとリソースデータの結合を行い、公開した。

平成 25 年度は、これまで作成したフェノームデータの定期的なアップデートとともに、細胞株フェノームに対する環境情報となる培養条件情報の拡充、微生物株（基本情報およびフェノーム情報）の統合化作業を実施した。

微生物表現型に関しては、DBCLS で策定中の、Microbial Phenotype Ontology (MPO) による特性記述に準拠した記述を行なった。これらの情報は、豊田グループ開発のフェノーム利用のために開発したインターフェース(画面)を用いて、配信を行なっている。これらのデータの可視化については、豊田グループと綿密な議論を行い、RDF のデータリンクを介して、表現型が関連するバイオリソースを提示できるようなインターフェース開発の協力および、評価を行った。

また、医薬基盤研究所 SAGACE データベースにおいて、当データベースで作成する RDF ファイルからの情報のクローリングを行なうようにし、RDF を介した培養細胞情報の統合連携を行なった。

## § 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 6 件）

1. 著者、論文タイトル、掲載誌 巻、号、発行年、DOI

1. Kobayashi N, Ishii M, Takahashi S, Mochizuki Y, Matsushima A, Toyoda T. “Semantic-JSON: a lightweight web service interface for Semantic Web contents integrating multiple life science databases”, Nucl Acids Res, 39:W533-540, 2011, DOI:10.1093/nar/gkr353
2. Baerenfaller K, Bastow R, Beynon J, Brady S, Brendel V, Donaldson S, Forster M, Gifford D, Grotewold E, Gutierrez R, Huala E, Jaiswal P, Joshi H, Kersey P, Liu L, Loraine A, Lyons E, May S, Mayer K, MacLean D, Meyers B, Mueller L, Muller R, Muller H.M, Ouellette F, Pires J. C, Provart N, Staiger D, Stanzione D, Taylor J, Taylor C, Town, C.J, Toyoda T, Vaughn M, Walsh S, Ware D, Weckwerth W. “Taking the Next Step: Building an *Arabidopsis* Information Portal”, Plant Cell, 24:2248-56, 2012, DOI:10.1105/tpc.112.100669
3. Makita Y, Kobayashi N, Yoshida Y, Doi K, Mochizuki Y, Nishikata K, Matsushima A, Takahashi S, Ishii M, Takatsuki T, Bhatia R, Khadbaatar Z, Watabe H, Masuya H, Toyoda T. “PosMed: ranking genes and bioresources based on semantic web association study”, Nucl Acids Res, 41:W109-W114, 2013
4. Cox R. S, Nishikata K, Shimoyama S, Yoshida Y, Matsui M, Makita Y, Toyoda T.

- “PromoterCAD: Data-driven design of plant regulatory DNA” Nucl Acids Res 41:W569-74, 2013, DOI:10.1093/nar/gkt518
5. Nishikata K., Cox R. S., Shimoyama S., Yoshida Y., Matsui M., Makita Y., Toyoda T., “Database construction for PromoterCAD: synthetic promoter design for mammals and plants”, ACS Synth Biol, 2013, 10.1021/sb400178c
  6. Suzuki T., Furuse T., Yamada I., Motegi H., Kozawa Y., Masuya H., Wakana S. “Pheno-Pub: a total support system for the publication of mouse phenotypic data on the web”, Mamm Genome, 24:473-483, 2013

## (2)その他の著作物(総説、書籍など)

1. 榎屋啓志. “表現型データの国際標準化とデータの利用、論文ができてしまう！疾患モデルマウス表現型解析指南” 348-359, 2011
2. 榎屋啓志. “バイオインフォマティクス” 生物機能モデルと新しいリソース・リサーチツール (series モデル動物利用マニュアル), 第2部第2章第3節 623-630, 2011
3. 豊田哲郎, 榎屋啓志. “フェノーム統合データベース” 細胞工学/Cell Technology, 31(2):240-243, 2012
4. 豊田哲郎. “SciNetS: セマンティックウェブ技術を活用した創薬のための情報基盤” SNR News, 23:2-5, 2012
5. Masuya H. “Roles and applications of biomedical ontologies in Experimental Animal Science”, Exp Anim, 61:365-373, 2012
6. 榎屋啓志. “第6章 生物学関連のオントロジーの動向” オントロジーの普及と応用, 115-130, 2012
7. 土井考爾, 榎屋啓志, 豊田哲郎. “フェノーム・バイオリソースを中心とした統合データベース開発の現状” 日本バイオインフォマティクス学会ニュースレター, 第26号, 2013
8. 榎屋啓志. “第11章 マウスのゲノムについて知ろう” マウス実験の基礎知識【第2版】, 147-156, 2013
9. 榎屋啓志. “第15章 コンピュータでマウスを研究しよう” マウス実験の基礎知識【第2版】, 199-209, 2013
10. 榎屋啓志. “生物学的特性情報を扱うバイオリソースデータベース～研究の幅を広げる、提案型のデータベースへ～” ビオフィリア, 9巻2号, 2013

## (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 5件、国際会議 5件)

### (国内)

1. 榎屋啓志. “マウスでの表現型情報の国際的統合の動向について”, 第58回日本実験動物学会総会, 東京, 2011年5月26日
2. 榎屋啓志. “マウス表現型の情報統合へ向けた研究”, 第26回モロシヌス研究会, 東京, 2012年6月15日
3. 豊田哲郎. “公共データのRDF化と公開を簡単に行う方法”, 第1回 LOD チャレンジデー「Linked Open Data 入門」, 東京, 2012年8月25日
4. 豊田哲郎. “公共データのRDF化と公開を簡単に行う方法”, 第2回 LOD チャレンジデー「Linked Open Data 入門」, 名古屋, 2012年9月1日
5. 榎屋啓志. “上・中位オントロジーからつくる表現型統合データベース”, ライフサイエンス統合データベースセンター勉強会, 東京, 2012年2月18日

### (国際)

1. Masuya H. “Recent progress of IT in JMC and AMPC”, 1st Asian Mouse Phenotyping Consortium Council Meeting, Seoul, 13 June 2011

2. Toyoda T. “Future Biology is Semantic Information Science”, 4th International SWAT4LS Workshop Semantic web applications and tools for life sciences, London, 7-9 Dec. 2011
3. Masuya H. “MUSDB: LIMS for Japan Mouse Clinic”, IMPC Informatics Workshop, Washington DC, 25 May 2012
4. Masuya H. “Development of information technologies for the integration of mouse phenotype data”, Infrafrontier/IMPC Korea Meeting, Jeju Korea, 27 Sep. 2012
5. Toyoda T. “POSMED: ANOTHER GATEWAY TO THE AIP DATABASES FROM LITERATURE”, the 24th International Conference on Arabidopsis Research (ICAR2013), Sydney, 29 June 2013

② 口頭発表(国内会議 14 件、国際会議 10 件)

1. 発表者(所属)、タイトル、学会名、場所、月日

(国内)

1. 榎屋啓志. “遺伝学オントロジーを基盤にしたバイオリソース情報統合”, 日本遺伝学会第 83 回大会, 京都, 2011 年 9 月 22 日
2. 脇和規, 榎屋啓志. “BioResource Proposer: the annotation database of mouse biological resources using KEGG MEDICUS and biomedical ontologies”, 日本分子生物学会第 34 回年会, 横浜, 2011 年 12 月 14 日
3. 西方公郎, 石井学, 吉田有子, 小林紀郎, 高橋聡史, 望月芳樹, 松嶋明宏, 田中芳幸, Gifford David Jonathan, 土井考爾, 原田えりみ, 蒔田由布子, 豊田哲郎. “BioLOD.org: 生命科学系 LOD データベースと既存アプリケーションによる可視化”, 第 26 回セマンティックウェブとオントロジー研究会, 横浜, 2011 年 12 月 14 日
4. 榎屋啓志. “ヒト疾患と対応したマウス表現型リソースデータベースの活用”, 平成 23 年度 第 3 回データベース講習会@つくば「創薬研究における統合データベースの活用」, つくば, 2012 年 3 月 9 日
5. 榎屋啓志, 田中信彦, 脇和規, 高月照江, 齋藤実香子, 小幡裕一. “理研 BRC IMPC(国際マウス表現型解析コンソーシアム) 参画計画3 -表現型解析情報の公開と統合-“, 日本実験動物科学・技術 九州, 別府, 2012 年 5 月 24 日
6. 榎屋啓志, 溝口理一郎. “遺伝子のオントロジー”, 2012 年度 人工知能学会 全国大会(第 26 回), 山口, 2012 年 6 月 5 日
7. 榎屋啓志, 溝口理一郎. “遺伝子の意味を情報モデルとして記述する”, 日本遺伝学会第 84 回大会, 福岡, 2012 年 9 月 24 日
8. 田中信彦, 茂木浩未, 鈴木智広, 金田秀貴, 三浦郁生, 山田郁子, 古瀬民生, 小林喜美男, 土岐秀明, 井上麻紀, 美野輪治, 野田哲生, 若菜茂晴, 榎屋啓志. “ヒト疾患の新規モデルマウスを探索するためのワークフローの開発”, 第 60 回日本実験動物学会総会, つくば, 2013 年 5 月 15 日
9. 榎屋啓志, 古崎晃司, 高田豊行, 真下知士, 若菜茂晴, 大江和彦, 溝口理一郎. “マウス・ラット表現型から疾患を推論するソフトウェアの開発”, 第 60 回日本実験動物学会総会, つくば, 2013 年 5 月 15 日
10. Masuya H, Tanaka N, Suzuki T, Wakana S. “Examinations of phenotyping data in Japan Mouse Clinic”, 第 60 回日本実験動物学会総会, つくば, 2013 年 5 月 17 日
11. 榎屋啓志, 古崎晃司, 大江和彦, 溝口理一郎. “コンテキストに依存した定性値を扱う生物表現型統合データベースの試作”, 2013 年度 人工知能学会 全国大会(第 27 回), 富山, 2013 年 6 月 6 日
12. 西方公郎, 小林紀郎, 国島直樹, 明恒次郎, 吉良聡, 桂一茂, 赤坂領吾, 寺田貴帆, 白水美香子, 横山茂之, 榎屋啓志, 豊田哲郎. “Integration of RIKEN protein experiment and structure database”, 第 86 回 日本生化学会大会, 横浜, 2013 年 9 月 11-13 日

13. 豊田哲郎. “フェノタイプ情報の高度な活用によるバイオリソースの付加価値創造”, トーゴの日シンポジウム 2013, 東京, 2013 年 10 月 5 日
14. 梶屋啓志. “マウス・ラットの表現型から関連疾患を探すデータベース”, 第 7 回ラットリソースリサーチ研究会, 京都, 2014 年 1 月 31 日

(国際)

1. Masuya H, Gkoutos G, Tanaka N, Waki K, Okuda Y, Kushida T, Kobayashi N, Doi K, Kozaki K, Hoehndorf R, Wakana S, Toyoda T, Mizoguchi R. “An Advanced Strategy for Integration of Biological Measurement Data”, International Conference on Biomedical Ontology, Buffalo, 29 July 2011
2. Masuya H, Mizoguchi R, “The Ontological Representation of Genetics”, 5th Interdisciplinary Ontology Conference (InterOntology12), Tokyo, 23 Feb. 2012
3. Masuya H. “Mouse phenotype data: from data capturing to making them open”, 2012 AMMRA/AMPC Annual Meeting and Workshop on Mouse Model Research, Nanjing, 16 Mar. 2012
4. Tanaka N. “Features of control data at Japan Mouse Clinic”, 2012 AMMRA/AMPC Annual Meeting and Workshop on Mouse Model Research, Nanjing, 17 Mar. 2012
5. Masuya H, Mizoguchi R. “An Ontology of Gene”, 3rd International Conference on Biomedical Ontology (ICBO), Graz, 23 Sep. 2012
6. Masuya H, “Internationally Cooperative Development of the Informatics Infrastructure for IMPC”, IMPC International Symposium, Tokyo, 28 Sep. 2012
7. Masuya H, “A trial of the development of Linked Open Data (LOD) of bioresources”, 4th ANRRC International Meeting (ANRRC 2012), Jeju, 18 Oct. 2012
8. Masuya H, Takatsuki T, Makita Y, Yoshida Y, Mochizuki Y, Kobayashi N, Yoshiki A, Nakamura Y, Toyoda T, Obata Y. “Development of Linked Open Data for Bioresources”, 2nd Joint International Semantic Technology Conference (JIST2012), Nara, 3 Dec. 2012
9. Masuya H. “International sharing of biological information in mouse”, RIKEN BRC/Nanjing University MARC 2nd International Summer Course of the mouse research, Nanjing, 29 July 2013
10. Masuya H. “Overview of current pipeline in RIKEN BioResource Center”, IMPC Phenotyping Workshop, Toronto, 28 Oct. 2013

③ ポスター発表(国内会議 26 件、国際会議 8 件)

1. 発表者(所属)、タイトル、学会名、場所、月日

(国内)

1. 梶屋啓志. “マウス・ラットのフェノーム情報統合へ向けて”, 第 25 回モロシヌス研究会, 新潟, 2011 年 7 月 8 日
2. 土井考爾, 梶屋啓志, 豊田哲郎. “動物・植物の枠組みを超えた新たなフェノーム統合データベースの開発”, トーゴの日シンポジウム 2011, 東京, 2011 年 10 月 5 日
3. 豊田哲郎, 土井考爾, 望月芳樹, 西方公郎, 豊島裕美, 雪竹俊臣. “生命と環境の様々なフェノームデータを統合・共有する BioLOD.org”, 分子生物学会第 34 回年会特別企画「使ってみようバイオデータベース—つながるデータ, 広がる世界」, 横浜, 2011 年 12 月 13-16 日
4. 脇和規, 梶屋啓志. “BioResource Proposer: the annotation database of mouse biological resources using KEGG MEDICUS and biomedical ontologies”, 日本分子生物学会第 34 回年会, 横浜, 2011 年 12 月 14 日
5. 田中信彦, 茂木浩未, 鈴木智広, 山田郁子, 古瀬民生, 井上麻紀, 美野輪治, 野田哲生,

- 若菜茂晴, 榎屋啓志. “国際的なマウス表現型解析プロジェクトのデータの利用価値を高めるための情報システム開発へ向けて”, 日本分子生物学会第 34 回年会, 横浜, 2011 年 12 月 14 日
6. 榎屋啓志, 高月照江, 斎藤実香子, 脇和規, 田中信彦, 目加田和之, 中田初美, 西條薫, 吉田有子, 蒔田由布子, 若菜茂晴, 中村幸夫, 吉木淳, 豊田 哲郎. “Semantic Web 技術を利用した哺乳類バイオリソースのデータベース”, 日本分子生物学会第 34 回年会, 横浜, 2011 年 12 月 15 日
  7. Doi K, Harada E, Watabe H, Masuya H, Toyoda T. “Development of novel database integrating phenotypes over species for improvement of plant omics research”, 植物生理学会 第 53 回年会, 京都, 2012 年 3 月 16–18 日
  8. 田中信彦, 茂木浩未, 鈴木智広, 金田秀貴, 三浦郁生, 山田郁子, 古瀬民生, 小林喜美男, 土岐秀明, 井上麻紀, 美野輪治, 野田哲生, 若菜茂晴, 榎屋啓志. “マウス表現型の類似度解析ツールの開発:ワークフローの検討”, 日本実験動物科学・技術 九州 2012, 別府, 2012 年 5 月 25 日
  9. 田中信彦, 茂木浩未, 鈴木智広, 金田秀貴, 三浦郁生, 山田郁子, 古瀬民生, 小林喜美男, 土岐秀明, 井上麻紀, 美野輪治, 野田哲生, 若菜茂晴, 榎屋啓志. “大規模マウス表現型データの統合的解析手法の開発”, 第 26 回モロシヌス研究会, 東京, 2012 年 6 月 15 日
  10. 土井考爾, 西方公郎, 下山紗代子, 榎屋啓志, 豊田哲郎. “BioLOD:フェノームおよびバイオリソース情報の統合と共有”, トーゴーの日シンポジウム 2012, 東京, 2012 年 10 月 5 日
  11. 榎屋啓志, 高月照江, 斎藤実香子, 松庫義弘, 蒔田由布子, 吉田有子, 望月芳樹, 小林紀郎, 吉木淳, 中村幸夫, 豊田哲郎, 小幡裕一. “哺乳類バイオリソースの Linked Open Data データベースのアップデート”, 第 35 回日本分子生物学会年会, 福岡, 2012 年 12 月 12 日
  12. 田中信彦, 茂木浩未, 鈴木智広, 金田秀貴, 三浦郁生, 山田郁子, 古瀬民生, 小林喜美男, 土岐秀明, 井上麻紀, 美野輪治, 野田哲生, 若菜茂晴, 榎屋啓志. “網羅的マウス表現型解析データを用いたデータマイニング:ワークフローの検討”, 第 35 回日本分子生物学会年会, 福岡, 2012 年 12 月 13 日
  13. 土井考爾, 下山紗代子, 西方公郎, 高月照江. “BioLOD (Biological Linked Open Databases)を使ったフェノタイプ情報の共有と検索”, 第 35 回日本分子生物学会年会, 福岡, 2012 年 12 月 13 日
  14. 高月照江, 斎藤実香子, 松庫義弘, 蒔田由布子, 吉田有子, 望月芳樹, 土井考爾, 小林紀郎, 中田初美, 西條薫, 吉木淳, 中村幸夫, 豊田哲郎, 榎屋啓志, 小幡裕一. “生物学的特性情報を扱うバイオリソースデータベース”, 第 60 回実験動物学会総会, つくば, 2013 年 5 月 15–17 日
  15. 西方公郎, 小林紀郎, 国島直樹, 明恒次郎, 吉良聡, 桂一茂, 赤坂領吾, 寺田貴帆, 白水美香子, 横山茂之, 榎屋啓志, 豊田哲郎. “理化学研究所タンパク質実験・構造統合データベースの構築”, 第 13 回日本蛋白質科学会年会, 鳥取, 2013 年 6 月 12–14 日
  16. 西方公郎, Cox Robert Sidney III, 下山紗代子, 吉田有子, 松井南, 蒔田由布子, 豊田哲郎. “Database Construction for Synthetic Promoter Design Web Application (PromoterCAD), 第 53 回 生命科学夏の学校, 静岡, 2013 年 8 月 31 日
  17. 西方公郎, Cox Robert Sidney III, 下山紗代子, 吉田有子, 松井南, 蒔田由布子, 豊田哲郎. “Database Construction for Synthetic Promoter Design Web Application (PromoterCAD), 第 53 回 生物物理夏の学校, 静岡, 2013 年 9 月 6–9 日
  18. 土井考爾, 榎屋啓志, 豊田哲郎. “植物表現型情報を中心とした統合データベースの開発”, 植物学会第 77 回大会, 札幌, 2013 年 9 月 13–15 日
  19. 蒔田由布子, 小林紀郎, 吉田有子, 土井考爾, 望月芳樹, 西方公郎, 松嶋明宏, 高橋聡史, 石井学, 高月照江, 榎屋啓志, 豊田哲郎. “PosMed: フェノタイプ情報と文献の統合によるバイオリソースの推論検索”, トーゴーの日シンポジウム 2013, 東京, 2013 年 10 月 4–5 日
  20. 西方公郎, Cox Robert Sidney III, 下山紗代子, 吉田有子, 松井南, 蒔田由布子, 豊田哲郎. “Database Construction for Synthetic Promoter Design Web Application



(PromoterCAD), 第 51 回 日本生物物理学会年会, 京都, 2013 年 10 月 28-30 日

21. 蒔田由布子, 小林紀郎, 吉田有子, 土井考爾, 望月芳樹, 西方公郎, 松嶋明宏, 高橋聡史, 石井学, 高月照江, 榊屋啓志, 豊田哲郎. “PosMed: Prioritizing Candidate Genes and Bioresources with Literature based Direct and Inferential Search”, 2013 年年会・生命医薬情報学連合大会, 東京, 2013 年 10 月 29-31 日
22. 西方公郎, Cox Robert Sidney III, 下山紗代子, 吉田有子, 松井南, 蒔田由布子, 豊田哲郎. “人工プロモータ設計アプリケーション PromoterCAD のためのデータベース構築”, 第 36 回 日本分子生物学会年会, 神戸, 2013 年 12 月 5 日
23. 蒔田由布子, 小林 紀郎, 吉田有子, 土井考爾, 望月芳樹, 西方公郎, 松嶋明宏, 高橋聡史, 石井学, 高月照江, 榊屋啓志, 豊田哲郎. “PosMed: フェノタイプ情報と文献の統合によるバイオリソースの推論検索”, 第 36 回 日本分子生物学会年会, 神戸, 2013 年 12 月 3 日
24. 豊田哲郎, 榊屋啓志, 土井考爾, 蒔田由布子, 西方公郎, 高月照江. “生命と環境のフェノーム統合データベース”, 第 36 回日本分子生物学会年会・特別企画 「使ってみようバイオデータベースーつながるデータ, 広がる世界」, 神戸, 2013 年 12 月 3 日-6 日

(国際)

1. Waki K, Masuya H. ”BioResource Proposer: the annotation and search system of mouse biological resources using molecular network information and ontologies”, Mouse Genetics 2011, Washington DC, 24 June 2011
2. Masuya H, Takatsuki T, Saito M, Waki K, Tanaka N, Mekada K, Nakata H, Makita Y, Yoshida Y, Wakana S, Toyoda T, Yoshiki A. “Development of semantic web based database of mammalian bio-resources”, Mouse Genetics 2011, Washington DC, 25 June 2011
3. Waki K, Masuya H. “BioResource Proposer: the mouse biological resource annotation database using KEGG MEDICUS and ontologies”, CBI/JSBi, Kobe, 8 Nov. 2011
4. Kobayashi N, Toyoda T. “BioSPARQL: Ontology-based smart building of SPARQL queries for biological Linked Open Data”, 4th International SWAT4LS Workshop Semantic web applications and tools for life sciences, London, 7-9 Dec. 2011
5. Nishikata K, Toyoda T. “BioLOD.org: Ontology-based integration of Biological Linked Open Data”, 4th International SWAT4LS Workshop Semantic web applications and tools for life sciences, London, 7-9 Dec. 2011
6. Nishikata K, Cox R S, Shimoyama S, Yoshida Y, Matsui M, Toyoda T. “Database Construction for Synthetic Promoter Design Web Application (PromoterCAD)”, The 6th International Meeting on Synthetic Biology (SB6.0), London, 9 July 2013
7. Cox R S, Nishikata K, Makita Y, Matsui M, Toyoda T. “Plant PromoterCAD: Data Driven DNA Design” The 6th International Meeting on Synthetic Biology (SB6.0), London, 9 July 2013
8. Shimoyama S, Cox R S, Yoshida Y, Toyoda T. “PromoterCAD: data driven design of plant regulatory DNA”, the 24th International Conference on Arabidopsis Research (ICAR2013), Sydney, 29 June 2013

#### (4) 知財出願

- ①国内出願  
なし
- ②海外出願  
なし

- ③その他の知的財産権  
なし

#### (5) 受賞・報道等

- ①受賞  
なし
- ②マスコミ(新聞・TV等)報道(プレス発表をした場合にはその概要もお書き下さい。)  
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  
なし

### § 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

ユーザー評価 (<http://biosciencedbc.jp/user-hyouka-2013/result-summary>) に寄せられた主なコメントについては以下のように対応した。

「現状では、PubMed と Ensembl があればよいと思う。PosMed を使うメリットはない。」という意見について。

PosMed では狭義の文献データだけでなく、多様なオミックスデータ(パスウェイ、オルソログ、共発現情報など)を利用した高度な推論検索を実現している。しかし、初期のバージョンでは検索手順や利用可能なデータの種類などがユーザーにわかりにくく、従来の検索システムとの差異が見えにくい面があった。これが PubMed で十分とのユーザーコメントにつながったものと考え。そこで PosMed インターフェースの改善を図り、結果表示画面のデザインも改善した。また、チュートリアルや収録データセットの説明などの文書の整備も進めた。これにより RDF データを活用した高速推論検索という当システムの特徴がより理解しやすくなったと考える。

「理研内のバイオリソースを基にしたサービスであれば、理研の運営費交付金で実施すれば良いように思います。」という意見について。

当プロジェクトでは確かに理研内で開発したバイオリソースの情報の統合化を先行的に進めてきたが、これはあくまでも出発点である。当初より理研外のコンテンツの統合を視野に入れてシステムを設計している。現在は理研外のマウス・細胞リソース等のコンテンツの積極的な統合を始めており、こうした誤解は近い将来に解消されるものと考え。

#### ②実施予定

より幅広いリソースを網羅し横断検索できるようにしてほしいとの要望が挙がっている。これらの要望に応えられるよう今後も継続的にコンテンツの拡充を実施したい。

また、内容の正確性への疑問の指摘があった。データ入力作業は現在も作業の途上であり、当然ながら収録データのチェックと修正も合わせて継続し、信頼性を維持できるよう努力する。

## §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