

# 表現型で生物をつなぐ フェノーム統合データベース

「生命と環境のフェノーム統合データベース」

理化学研究所 バイオリソースセンター

梶屋啓志

# フェノタイプは生命科学の最重要情報

フェノタイプ: 遺伝因子と、環境因子の相互作用によって現れる生命の形質の変化

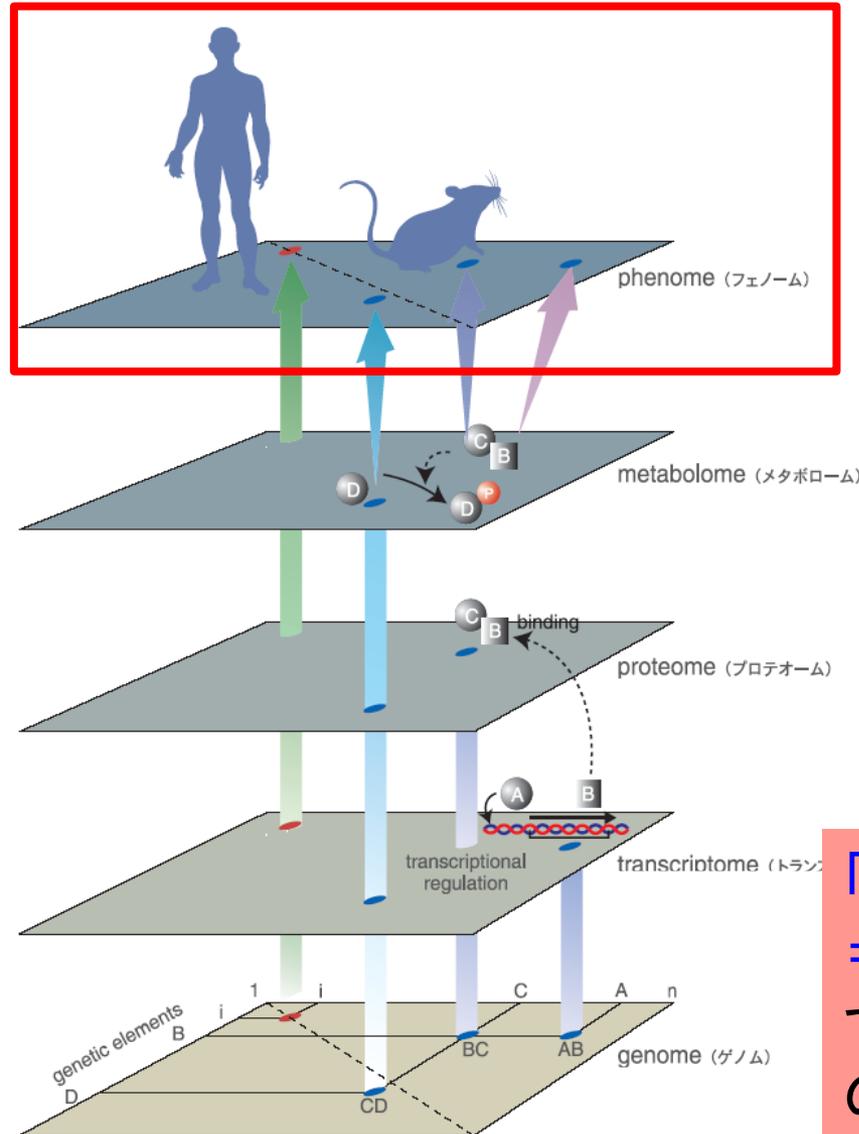
例: 疾患、先天的な異常、検査値、薬剤の効果...

フェノタイプ同士は、あらゆるレベルで相互に関係している!

- 個体レベル:  
分子パスウェイを通して互いに関係。
- 生物種横断的なレベル:  
進化的なつながり。モデル生物
- 生態系のレベル:  
生物間相互作用、環境応答



「フェノタイプのつながりを持った総体」  
=フェノームを、遺伝情報と環境の言葉  
で理解することこそが、生命原理の理解  
の応用のための重要課題!



# 本プロジェクトのねらい

フェノタイプと関連データの情報共有と標準化を、  
研究分野の垣根を超えて実現する



(RDFによる  
メタデータ空間)

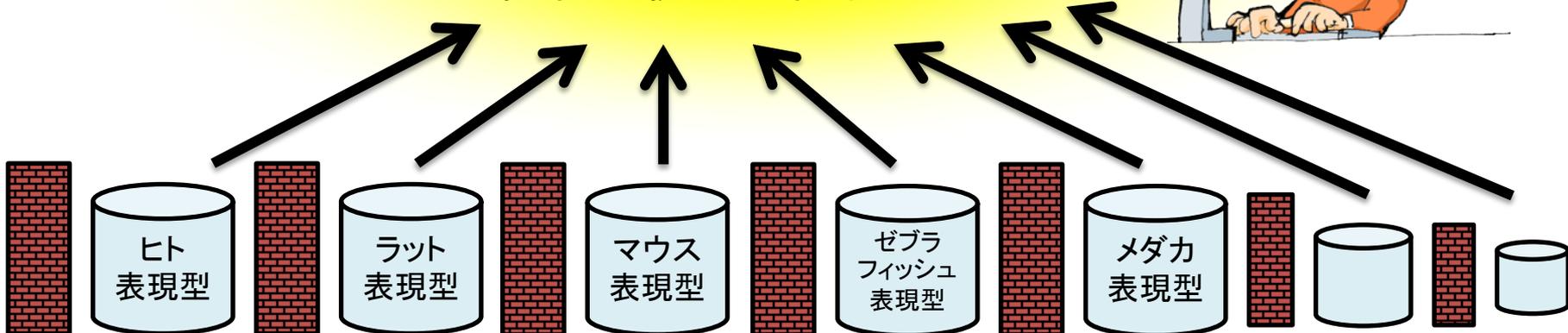


自由に一括  
ダウンロード

イノベーションの推進

疾患研究の新たなモデル系  
薬物応答に影響する変異／多型  
創薬の新たなターゲット候補

統一フォーマット  
標準化・統合化・体系化

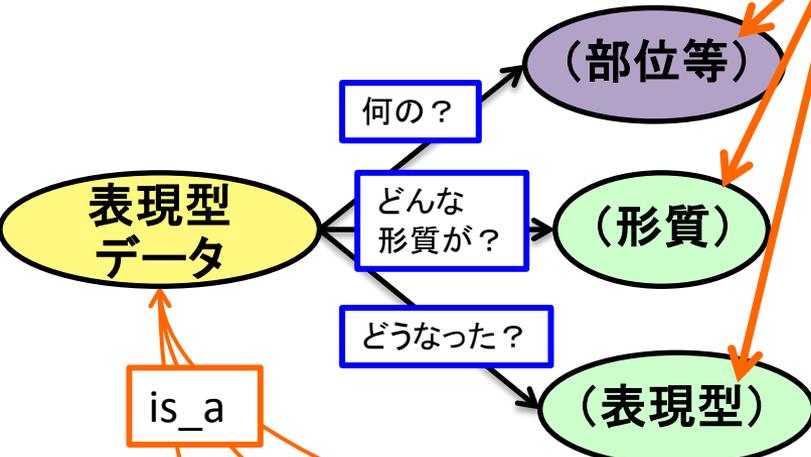


# 第1期プロジェクトの成果(基盤技術開発): RDFによるフェノタイプの統一スキーマの開発

表現型データ汎用RDFスキーム(抜粋)  
(OBOコンソーシアム提案の改良型)

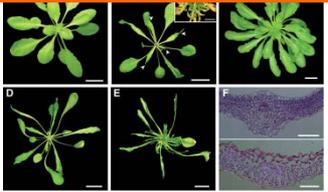
生物種に対応した  
オントロジーで代入

何の?  
どんな  
形質が?  
どうなった?

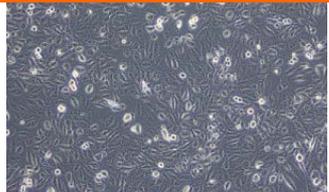


表現型アノテーション	系統	mammalian phenotype ontology	測定部位	測定対象の形質	測定結果	何の?/形質
<a href="#">phenotype of M100919</a>	M100919	赤血球細胞数の増加	-	has number of	has extra parts of type	erythrocyte
<a href="#">phenotype of M100451</a>	M100451	短足症	foot	size	decreased size	-
<a href="#">phenotype of M101156</a>	M101156	光受容体外節の短縮	photoreceptor outer segment	length	decreased length	-
<a href="#">phenotype of M101152</a>	M101152	聴覚障害	sensory perception of sound	rate	decreased rate	-
<a href="#">phenotype of M100646</a>	M100646	血中インスリンレベル上昇	blood	concentration of	increased concentration	insulin
<a href="#">phenotype of M100856</a>	M100856	短指	limb digit	length	decreased length	-
<a href="#">phenotype of M100702</a>	M100702	インスリンの血中濃度減少	blood	concentration of	decreased concentration	insulin
<a href="#">phenotype of M100392</a>	M100392	高血糖	blood	concentration of	increased concentration	glucose
<a href="#">phenotype of M100210</a>	M100210	高血糖	blood	concentration of	increased concentration	glucose

植物表現型特性情報



培養細胞特性情報



KOマウス表現型情報



(マウス表現型の例)

フェノタイプの「ワンストップショップ」を実現する技術

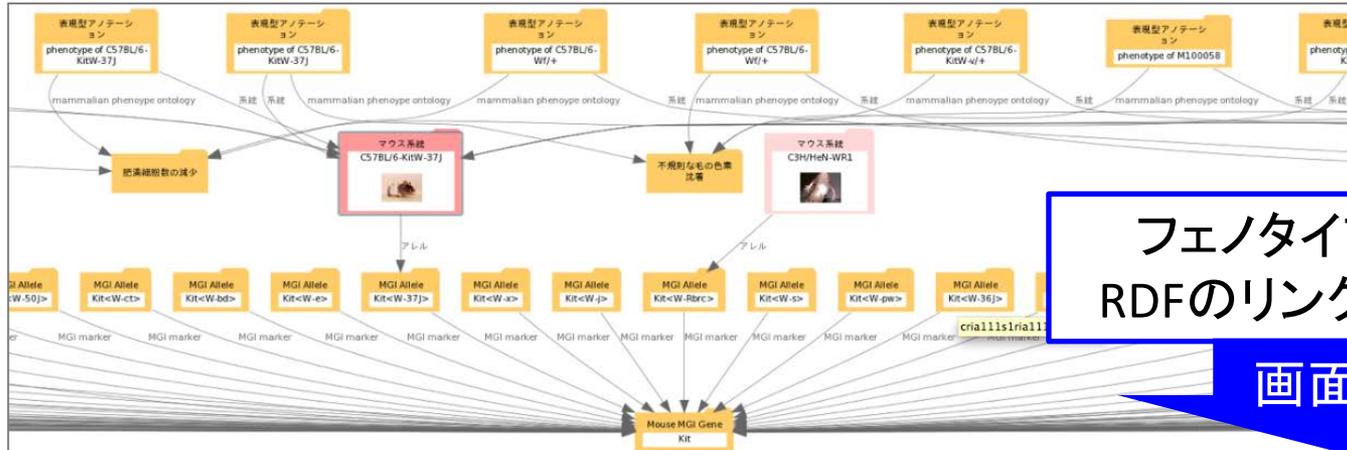
# フェノーム汎用スキーマで可能になったこと

## 表現型という『複雑で自由度の高い』情報に対して・・・

- スキーマの統一化による「ワンストップショップ」構築の可能性が高まった。  
⇒異なる生物の異なるフェノタイプを、ひとつのデータベース(あるいは連携可能なデータベース群)で、コストを抑えて管理できる。
- 部位、異常の内容、それぞれのオントロジー語句検索が可能になり、語彙表現の揺らぎに影響されない、表現型の多次元分類が可能に。  
⇒「長骨の長軸方向の成長が阻害されるKOマウス」等を、語彙表現のゆらぎを超え、特徴に基づいて検索する事が可能になった。
- 同じく、オントロジー・アノテーションを用いている、外部データベースの成果との相互利用が可能に  
⇒ Mouse Genome Informatics (MGI: ジャクソン研)  
Phenome Net (ケンブリッジ大)
- 種内、および種横断的な表現型類似性の推論が可能に。  
⇒「お勧めマウスリソース」提案機能

# フェノタイプの類似性を自動探索して「お勧めマウス」を提示

研究者が探したいマウスとフェノタイプが類似する他のマウスを分かりやすく提示



研究者に新たな気づきを与える

フェノタイプに関する  
RDFのリンクを自動探索

画面生成

## 同じ表現型を示すマウス系統 (56) (各系統の示す表現型の一致度が高い順番に提示)

<p>7 個のmammalian phenotype ontologyが一致 <b>C57BL/6-Wf/+</b></p>  <p>共通の mammalian phenotype...</p> <ul style="list-style-type: none"> <li>肥満細胞数の減少</li> <li>薄い毛色</li> <li>腹部の白斑</li> <li>大球性貧血</li> <li>赤血球数の減少</li> </ul>	<p>4 個のmammalian phenotype ontologyが一致 <b>C57BL/6-Wsh/Wsh</b></p>  <p>共通の mammalian phenotype...</p> <ul style="list-style-type: none"> <li>肥満細胞数の減少</li> <li>被毛色異常</li> <li>赤血球数の減少</li> <li>可変性体部斑点</li> </ul>	<p>3 個のmammalian phenotype ontologyが一致 <b>C57BL/6-KitW-37J</b></p>  <p>共通の mammalian phenotype...</p> <ul style="list-style-type: none"> <li>肥満細胞数の減少</li> <li>薄い毛色</li> <li>白色スポット</li> </ul>	<p>3 個のmammalian phenotype ontologyが一致 <b>C3.Cg-Kitl&lt;Sl-pan&amp;g...</b></p>  <p>共通の mammalian phenotype...</p> <ul style="list-style-type: none"> <li>薄い毛色</li> <li>大球性貧血</li> <li>白色スポット</li> </ul>
---	--	---	--

類似度高

類似度低

フェノタイプ推薦機能でバイオリソースの付加価値が向上

# フェノーム統合の第2段階へ

## 第1期(終了)

## 第2期

主要な  
統合対象  
データ

マウス系統:約5000系統  
培養細胞:約3600株  
微生物株:約15000株  
(理研バイオリソースセンター)

シロイヌナズナ変異株  
:3700件環境資源科学科  
学研究センター)

**【提供データ拡充】**提供するフェノーム情報の質と量の拡大

分子フェノタイプ:

マウスCreドライバー系統 マーカー遺伝子の発現情報:約50系統 (熊本大)

ゼブラフィッシュのジーントラップ系統 マーカー遺伝子の発現情報:約500系統 (遺伝研)

単純フェノタイプ:

マウス網羅的表現型解析データ:約100系統(理研)  
ラット網羅的表現型解析データ:約1000系統(京大)  
メダカ表現型:約500系統(基生研)

複合フェノタイプ:

疾患オントロジー:約5000疾患(リンク)  
国際的な標準病名ICD10:約12000疾患(リンク)

NBDC各グループと連携、RDF技術に基づく相互利用を実現

データ  
収集方法

データベースより直接収集  
(バイオリソースセンター等)

**【データ収集技術】**生物学者向け表現型データ入力S開発

データベースからの直接収集に加え、一般ユーザー(ゲノム編集コンソーシアム)からの直接入力システムの開発

可視化

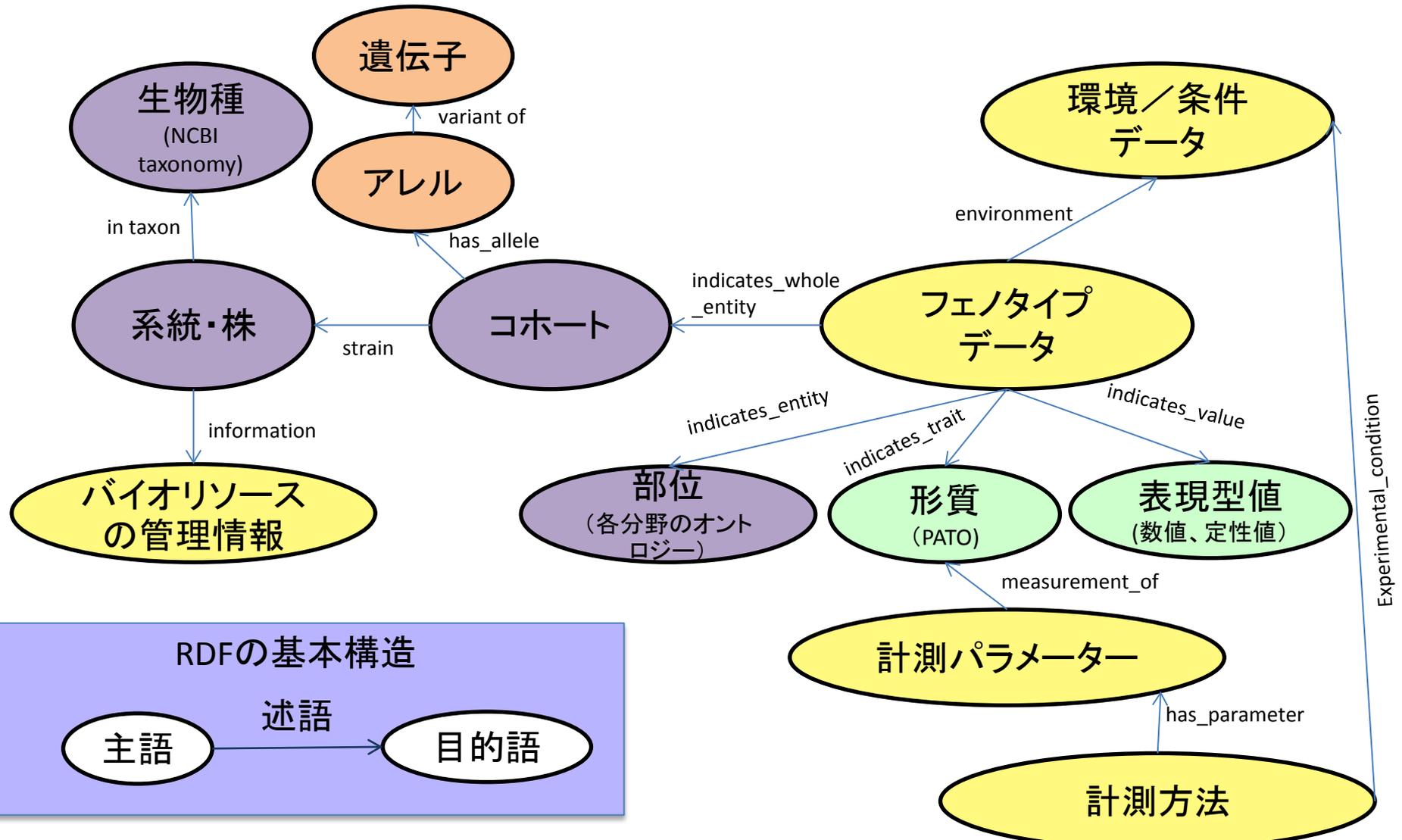
つながり検索、お勧めリソース

**【データ活用技術】**モデル生物表現型と疾患の関連づけと可視化

疾患研究者が最適な研究モデル系を探せるシステム  
(SPARQLによるインターフェース開発)

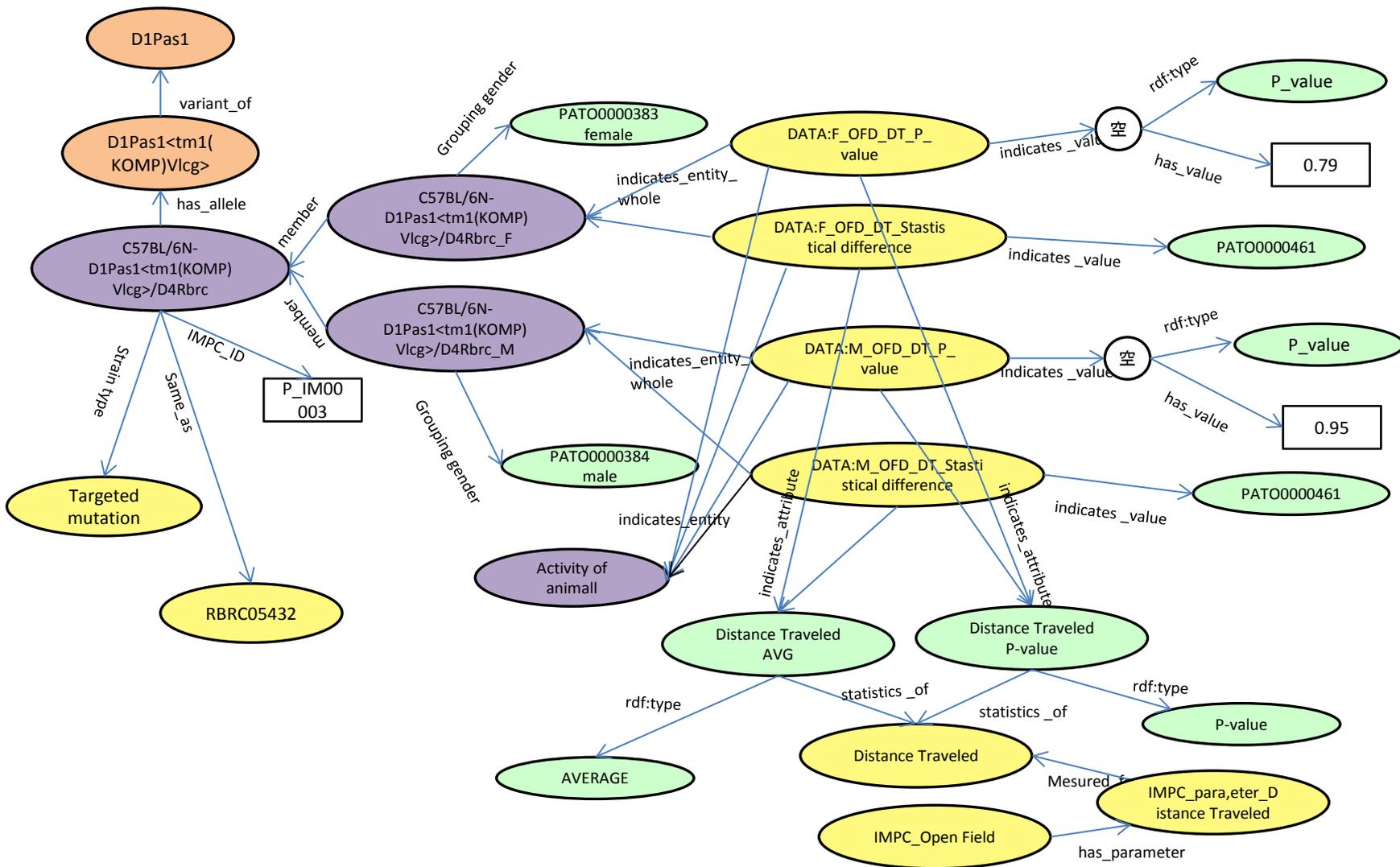
# H26年度計画：生物種の拡張と、RDFスキーマの見直し

## 生命の解析データを表現する汎用RDFスキーマ設計



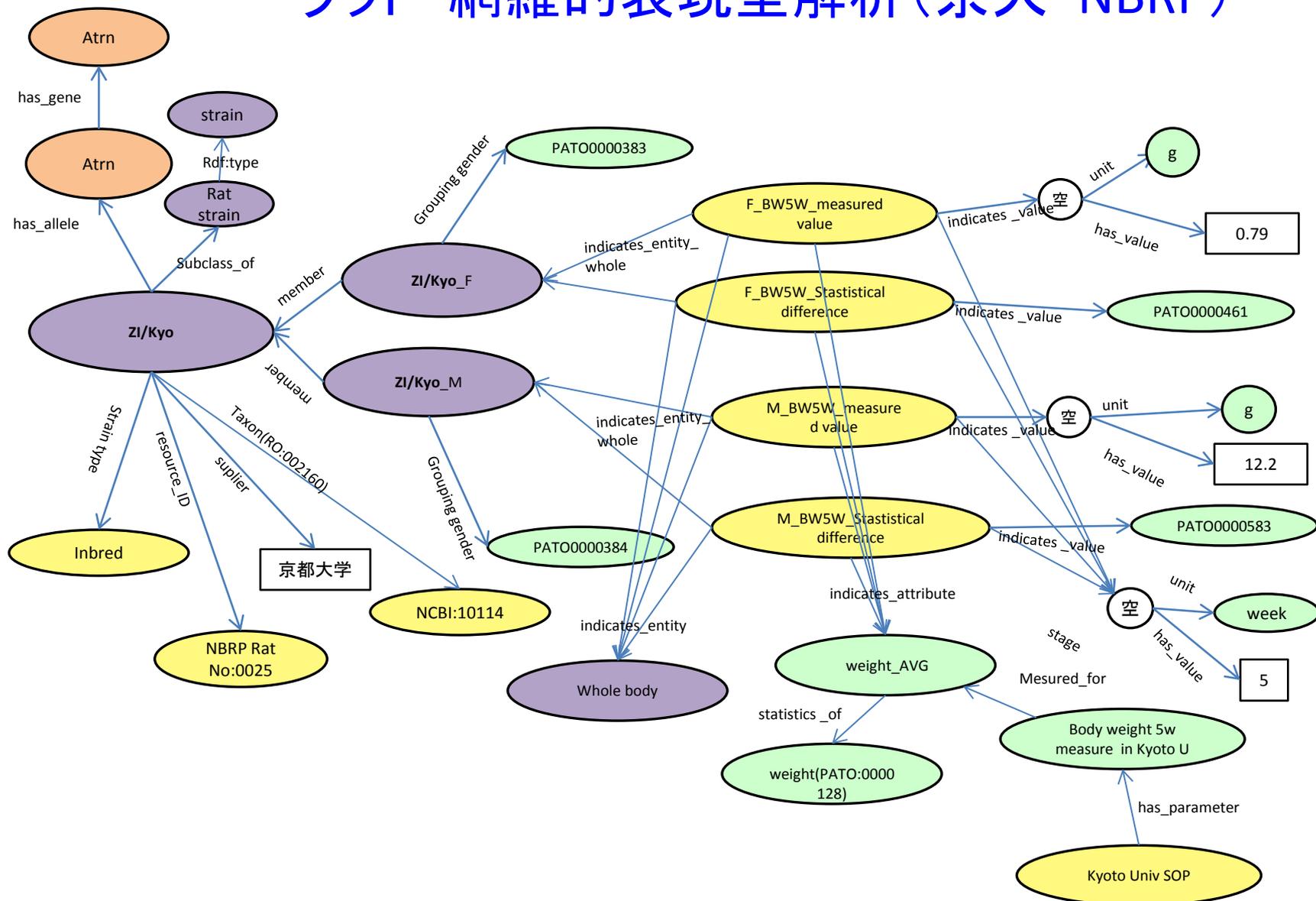
# 生物種の拡張と、RDFスキーマの見直し

## マウス・網羅的表現型解析(理研BRC)



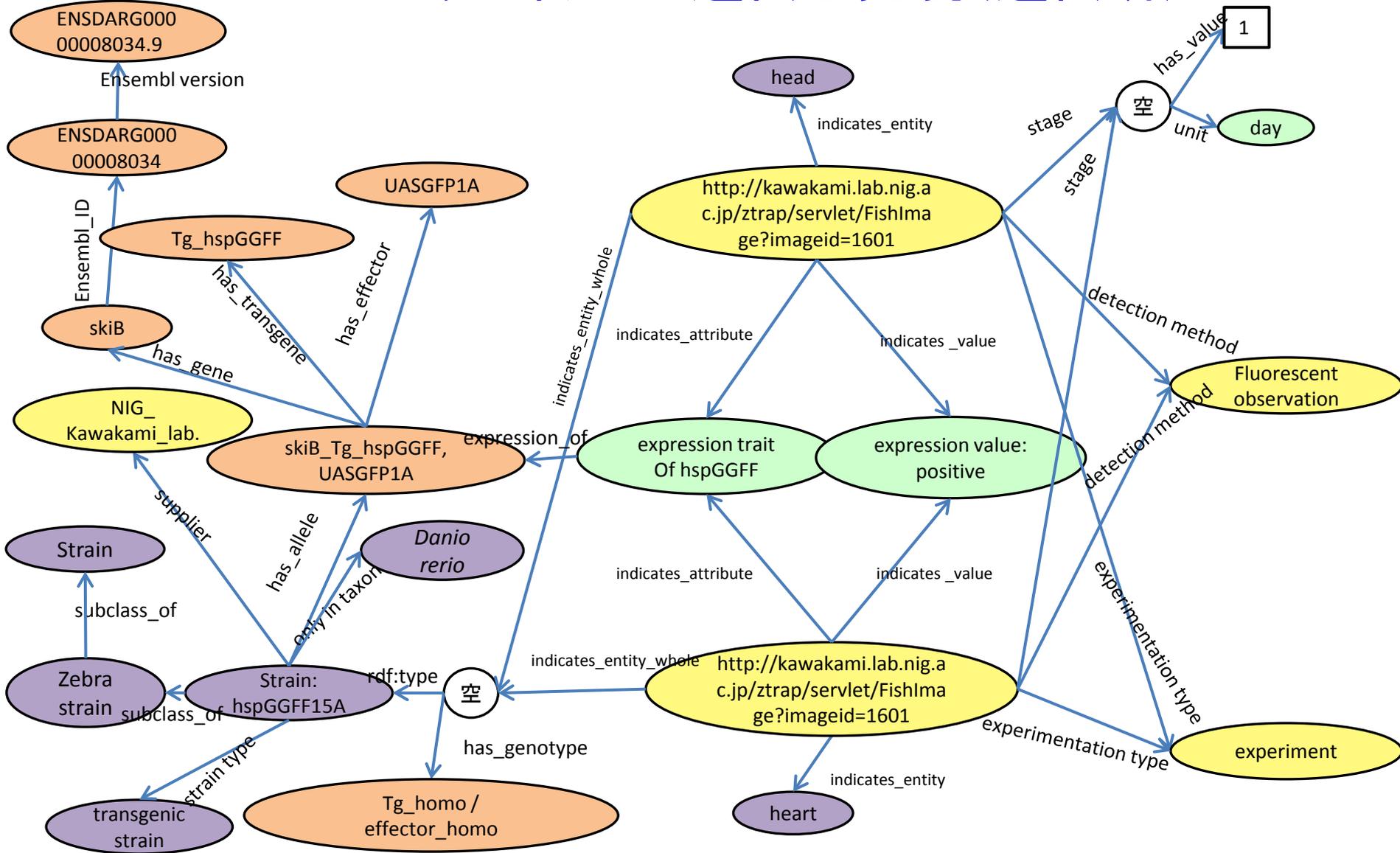
# 生物種の拡張と、RDFスキーマの見直し

## ラット・網羅的表現型解析(京大・NBRP)



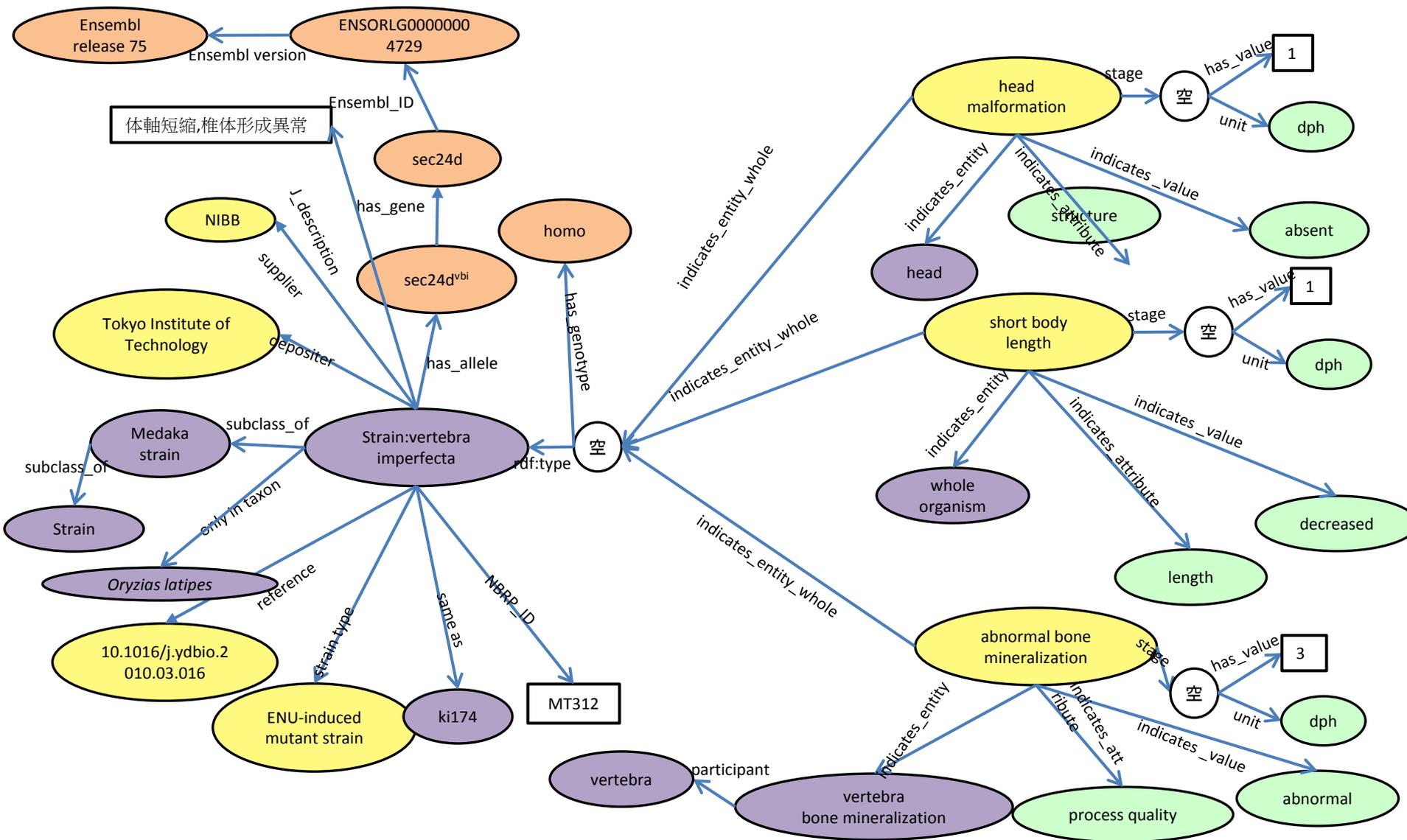
# 生物種の拡張と、RDFスキーマの見直し

## ゼブラフィッシュ遺伝子発現(遺伝研)



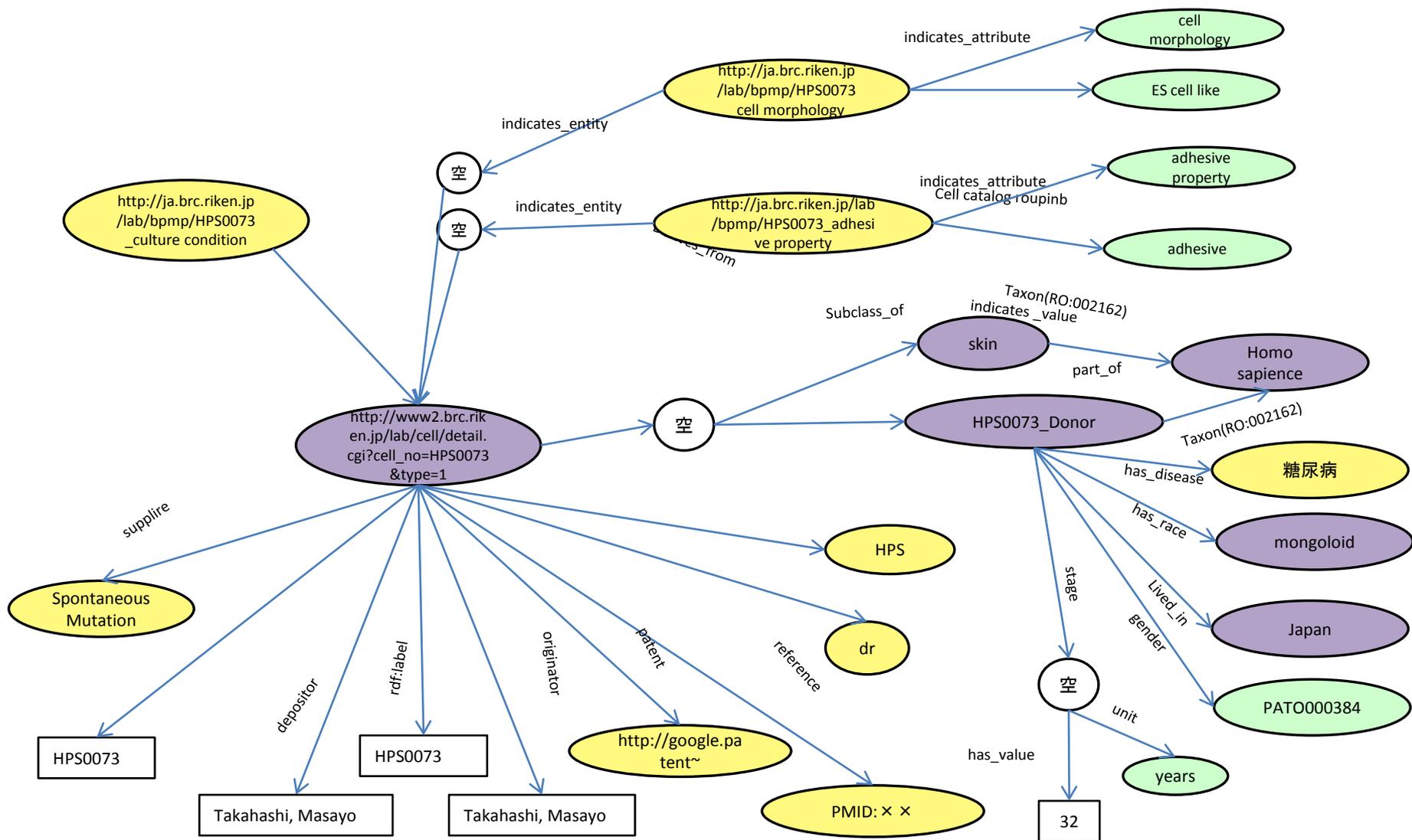
# 生物種の拡張と、RDFスキーマの見直し

## メダカ・表現型データ(基生研)



# 生物種の拡張と、RDFスキーマの見直し

## iPS細胞特性(理研BRC)



# 統合化推進プログラム(TPP)連携

統合に向けたRDF化支援、データの相互利用へ向けた調整と連携

各プロジェクトの『守備範囲』のヒートマップ

カテゴリ	データの種類	菅野グループ	春田グループ	成松グループ	梶屋グループ	田畑グループ	黒川グループ	DBCLSグループ	NBDC
Biological Entity	系統、株、家系								
	バイオサンプル								
	ウイルス・バクテリア								
	部位								
Chemical Entity	化合物(代謝物)								
	糖鎖								
	糖タンパク質								
Genomic Entity	ゲノムの中の位置情報								
	オルソログ								
	遺伝子・遺伝マーカー・QTL								
	アレル,バリエーション, SNP								
Molecular Phenotype	エピゲノム								
	マススペクトル								
	RNAseq								
	トランスクリプトーム								
Biological Phenotype	表現型								
	植物疾患、病害抵抗性								
	ヒト疾患								
Molecular role	薬理活性								
Environmental Factor	実験条件								
	環境								
	メタゲノム								
	培地								
Informational Entity	実験方法								
	論文								
	データベースリンク集								

 データを保持している / 作成する  
 外部データの参照を行なう

生物サンプル、化合物、文献、ゲノム情報等について、共通スキーマ策定中

# 進捗状況

収集対象	目標数	データ取得数	備考
マウスCreドライバー系統(熊本大・理研BRC)	50系統	130系統	RDFスキーマ作成済 テストデータ作成中
ゼブラフィッシュ ジェントラップ系統(遺伝研)	500系統	-	データスキーマ協議中 (テストデータ作成中)
マウス網羅的表現型データ(理研BRC)	100系統	約50コホート (21系統)	RDFスキーマ作成済 国際プロジェクト(IMPC)とのデータ連携を 協議中
コンソミックマウス網羅的表現型データ(遺伝研)		60コホート (30系統)	RDFスキーマ作成済
ラット網羅的表現型データ(京大)	1000系統	216コホート (172系統)	RDFスキーマ作成済 テストデータ作成中
ラット表現型データ(RGD)		約1000系統	RGDよりインポート
メダカ表現型(基生研)	500系統	15系統	RDFスキーマ作成済 テストデータ作成中 基生研と連携で、オントロジーを用いた表 現型アノテーションを開始
BRCマウス系統	(第1期課題)	5,722系統	第1期からのRDFスキーマの見直し テストデータ作成中
BRC細胞株	(第1期課題)	3,775株	第1期からのRDFスキーマの見直し テストデータ作成中
JCM微生物株	(第1期課題)	14,752株	第1期からのRDFスキーマの見直し 黒川Gとのデータ連携準備中 テストデータ作成中

# DB機能計画: 疾患研究のモデル生物を探索する

画面案: ヒト疾患名をキーとしてのモデル動物(の可能性)を生物種横断的に探索できる。  
最後はオリジナルのデータベースに案内する。

○ ヒト疾患との関連で探す  
○ 生物種から探す

表現型の関連で探す  
 遺伝子の関連で探す

疾患名を入力

検索

検索結果: [白内障](#) と関連する表現型表示しています。

生物種	表現型	検索条件との関連	部位	画像	リソース	オリジナルサイト
ラット	白内障	表現型一致	eye		0155	<a href="#">NBRP Rat</a>
マウス	白内障	表現型一致	lens		RBRC00376	<a href="#">RIKEN BRC</a>
	小眼球症	部位の一致	eye			
ゼブラフィッシュ	Marker for srebfs is expressed at day 5	部位の一致	eye		hapGGFF3A	<a href="#">zTRAP</a>
	Marker for srebfs is expressed at day 5	病因遺伝子一致 (srebfs)	-			
ゼブラフィッシュ	Marker for unknown gene is expressed at day1	部位の一致	eye		SAGFF(LF)2 26F	<a href="#">zTRAP</a>

# DB機能計画: 他の生物種におけるモデル系を探す

普段、マウスやラットなど他の生物を使用している研究者が、例えばゼブラフィッシュ系統など、別種での生命現象研究のモデル系を探索できるようにする。

ヒト疾患との関連で探す  
 生物種から探す

表現型の関連で探す  
 遺伝子の関連で探す

生物種

の部位名  
で探す

表現型の種類

遺伝子との関連

遺伝子名を入力

検索

検索結果: [limb](#) と関連する表現型表示しています。

生物種	表現型	検索条件との関連	部位	画像	リソース	オリジナル サイト
ラット	手足の短小化	部位の一致	四肢		NBRP Rat No: 0474	<a href="#">NBRP Rat</a>
ラット	手足の短小化	部位の一致	四肢		NBRP Rat No: 0651	<a href="#">NBRP Rat</a>

# まとめ：第2期の展望

- **データとしての『フェノーム』の実現**  
相互につながった「フェノタイプの総体」をデータとして実現し、データサイエンティストが自由かつ手軽に利用できる、生物横断的かつ標準化かつ体系的に統合化されたデータアーカイブとする。
- **分子情報と共に利用できるフェノタイプデータ**  
遺伝因子との関連性の記述、およびNBDC各データとの連携強化により、フェノタイプ情報を、分子情報と共に利用できる統合データの一部とする。これにより、生命科学イノベーションを後押しする。
- **国内のデータベース連携の強化と各DBレベルでの改良を推進**  
フェノタイプを扱う国内の各データベースや、プロジェクト、コンソーシアムとの具体的な連携体制を構築し、拡げていく。  
生物種横断的な統合が可能であることと、そのメリット（インセンティブ）を示すことで、NBDCのデータ統合に賛同するデータベースを増やし、各データベースが積極的に統合に参加し、標準化に取り組む土台を築く

# 謝辞

- 国立遺伝学研究所・哺乳動物遺伝研究室、初期発生研究部門 (マウス、ゼブラフィッシュ)
- 基礎生物学研究所・バイオリソース研究室 (メダカ)
- 熊本大学・生命資源研究・支援センター (マウス)
- 京都大学大学院医学研究科・附属動物実験施設 (ラット)
- 広島大学・理学研究科・分子遺伝学研究室 (ゲノム編集)
- 理化学研究所・バイオリソースセンター (マウス、細胞、微生物)
- 東京大学大学院医学系研究科医療情報経済学分野 (臨床医学オントロジー)
- 大阪大学・産業科学研究所・知識科学研究分野 (上位オントロジー)
- 北陸先端大学・サービスサイエンス研究センター (上位オントロジー)
- ケンブリッジ大 (フェノタイプオントロジー)
- ゲノム編集コンソーシアム (遺伝子改変技術・生物横断的)
- Cell Line Ontology Consortium (培養細胞データ)
- International Mouse Phenotyping Consortium (マウス大規模解析)
- NBDC・統合化推進プログラムの各プロジェクト (生命科学情報統合)