

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・
ユニットリーダー
榎屋 啓志

フェノタイプは生命科学の最重要情報

定義: **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、遺伝子発現...

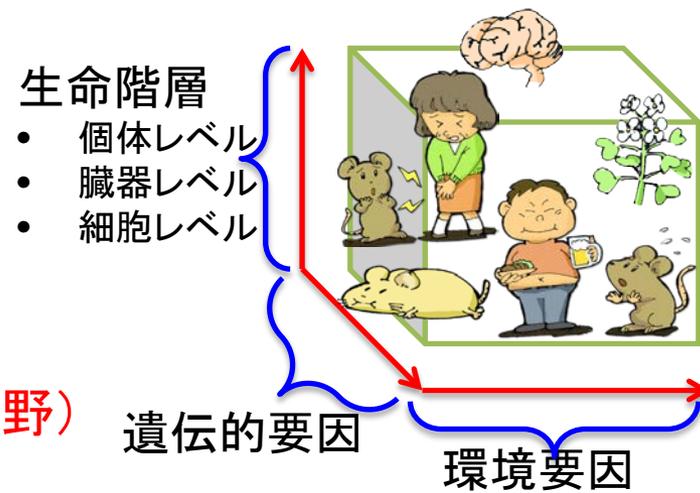
フェノタイプ相互の関連性は、あらゆるレベルで重要

- 個体: 分子パスウェイの共有を通して互いに関係
- 生物種間: 進化的つながり。モデル生物。
- 生態系: 生物間相互作用、環境応答

生命のフェノタイプの(横断的な)総体=フェノーム

- 「フェノーム」を、遺伝因子、環境因子、分子メカニズムの違いで理解する事は、生命原理の理解と応用を加速するための重要課題

- 「横断的統合」が必須(形質、生物種、研究分野)



我が国のフェノタイプ情報の現状

- 系統、個体や細胞のリソースのデータベース、解析の成果公開データベースにおいて、個体の特性、検査等の測定結果、遺伝子発現などとして公開されている。
- フェノタイプの記述方法、データの形式は、文章記述や、器官組織の名称のみを示すなど、個別でばらばらであり、**データを集積しまとめて解析する事はできない。**
- 結果的に、データは各生物種のコミュニティで限定的にしか用いられておらず、**今後、生命データを用いたイノベーションの障壁になると懸念される。**

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

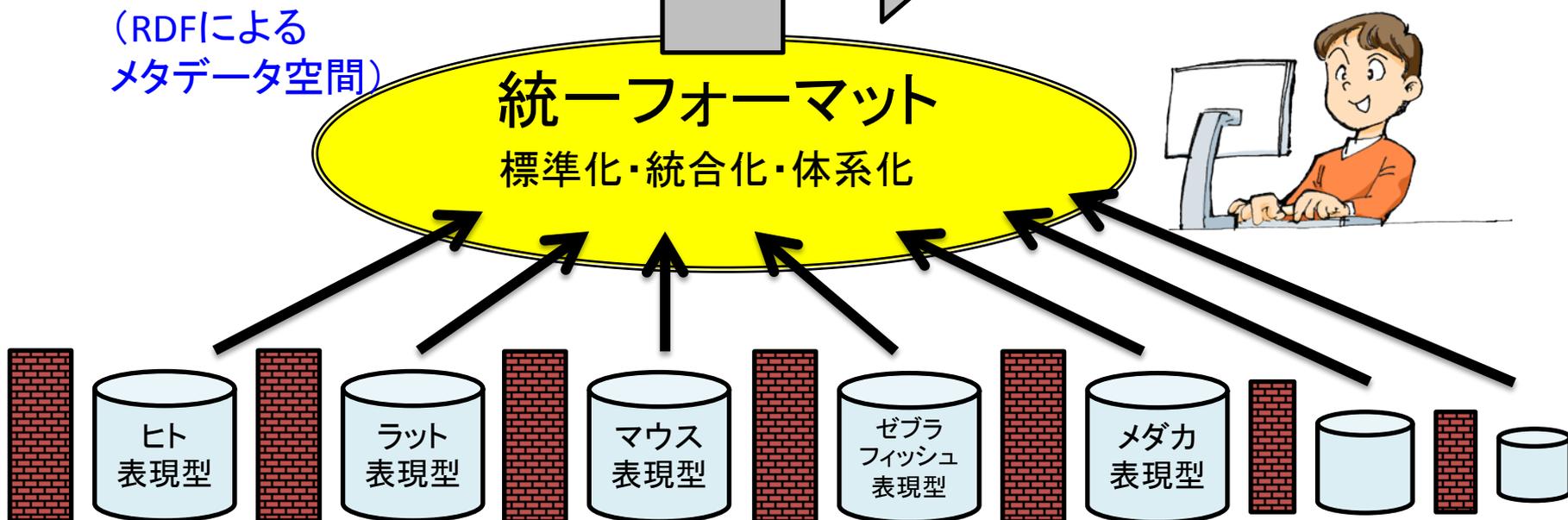
これまで利用者側で結び付けにくかった異種データ間の新たな結合を可能に。

(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

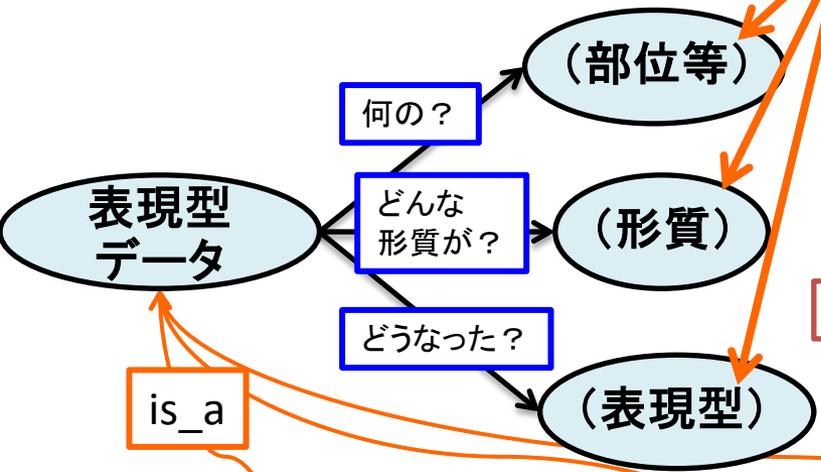


RDFによるフェノタイプの統一フォーマットの開発

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)

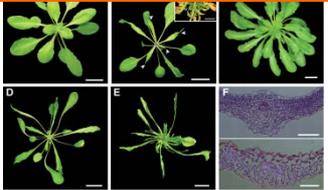
生物種に対応した
オントロジーで代入

何の?
どんな
形質が?
どうなった?

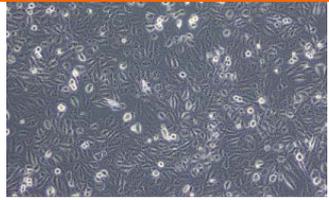


表現型アノテーション	系統	mammalian phenotype ontology	測定部位	測定対象の形質	測定結果	何の?/形質
phenotype of M100919	M100919	赤血球細胞数の増加	-	has number of	has extra parts of type	erythrocyte
phenotype of M100451	M100451	短足症	foot	size	decreased size	-
phenotype of M101156	M101156	光受容体外節の短縮	photoreceptor outer segment	length	decreased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

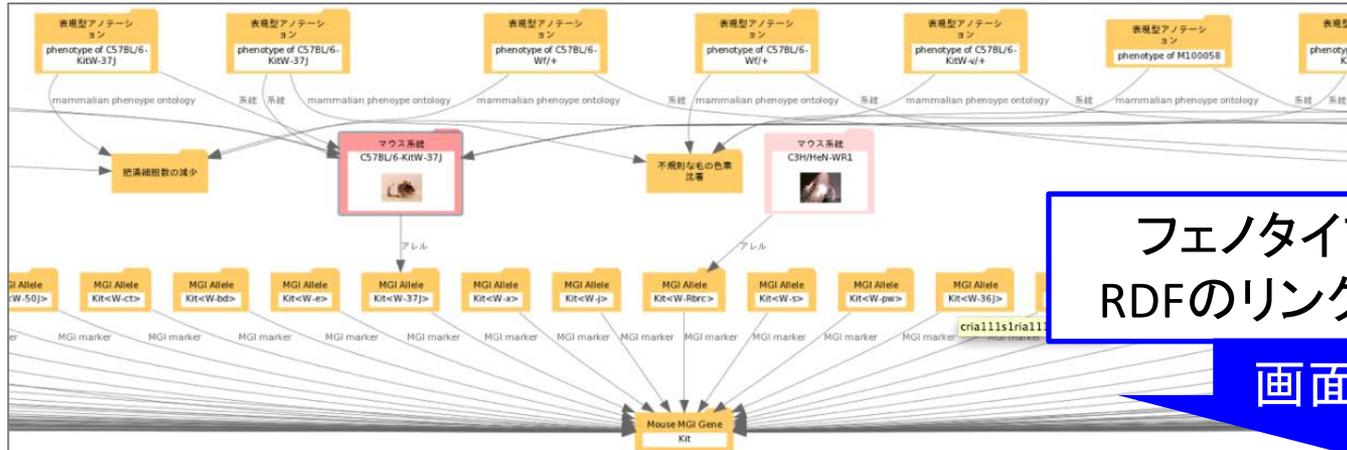


(マウス表現型の例)

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

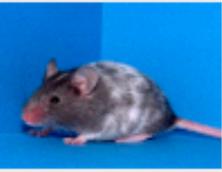


研究者に新たな気づきを与える

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

類似度高

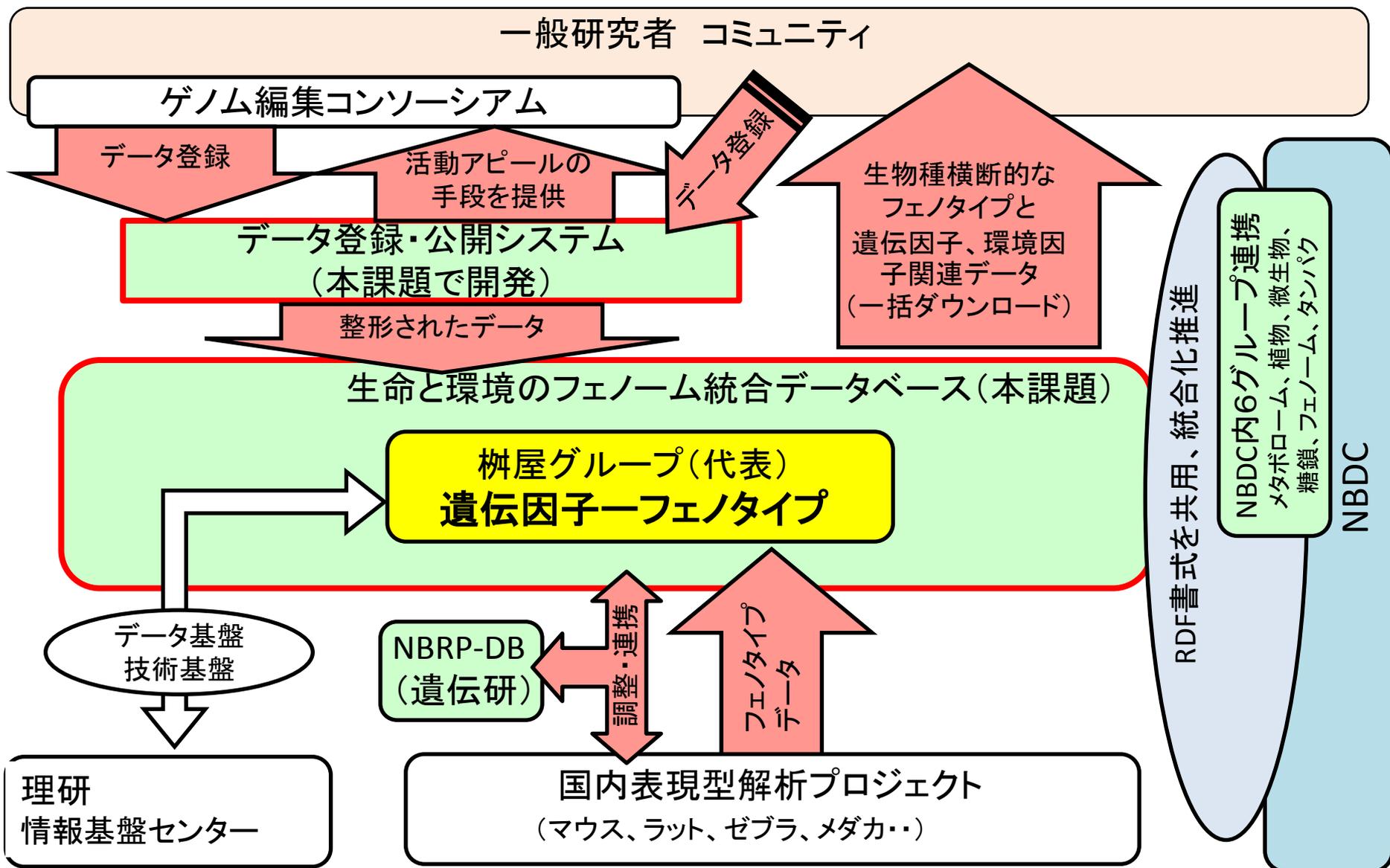
類似度低

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

第2期研究開発目標(第1期からの発展)

第1期(終了)		第2期
<p>主要な統合対象データ</p>	<p>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)</p> <p>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)</p>	<p>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)</p> <p>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)</p> <p>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)</p>
<p>データ収集方法</p>	<p>データベースより直接収集 (バイオリソースセンター等)</p>	<p>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</p>
<p>可視化</p>	<p>つながり検索、お勧めリソース</p>	<p>疾患研究者が最適な研究モデル系を探せるシステム</p>
<p>ユーザー</p>	<p>疾患研究者、バイオリソース利用者(マウス、細胞、微生物)</p>	<p>さらに広い、疾患研究者、バイオリソース利用者 データサイエンティスト</p>
<p>RDF技術連携</p>	<p>独自色の強いRDF統合 (パフォーマンス重視)</p>	<p>NBDC各グループと調整し、実際的な相互利用を実現 RDFレポジトリ SPARQLによるインターフェース開発</p>

研究実施体制

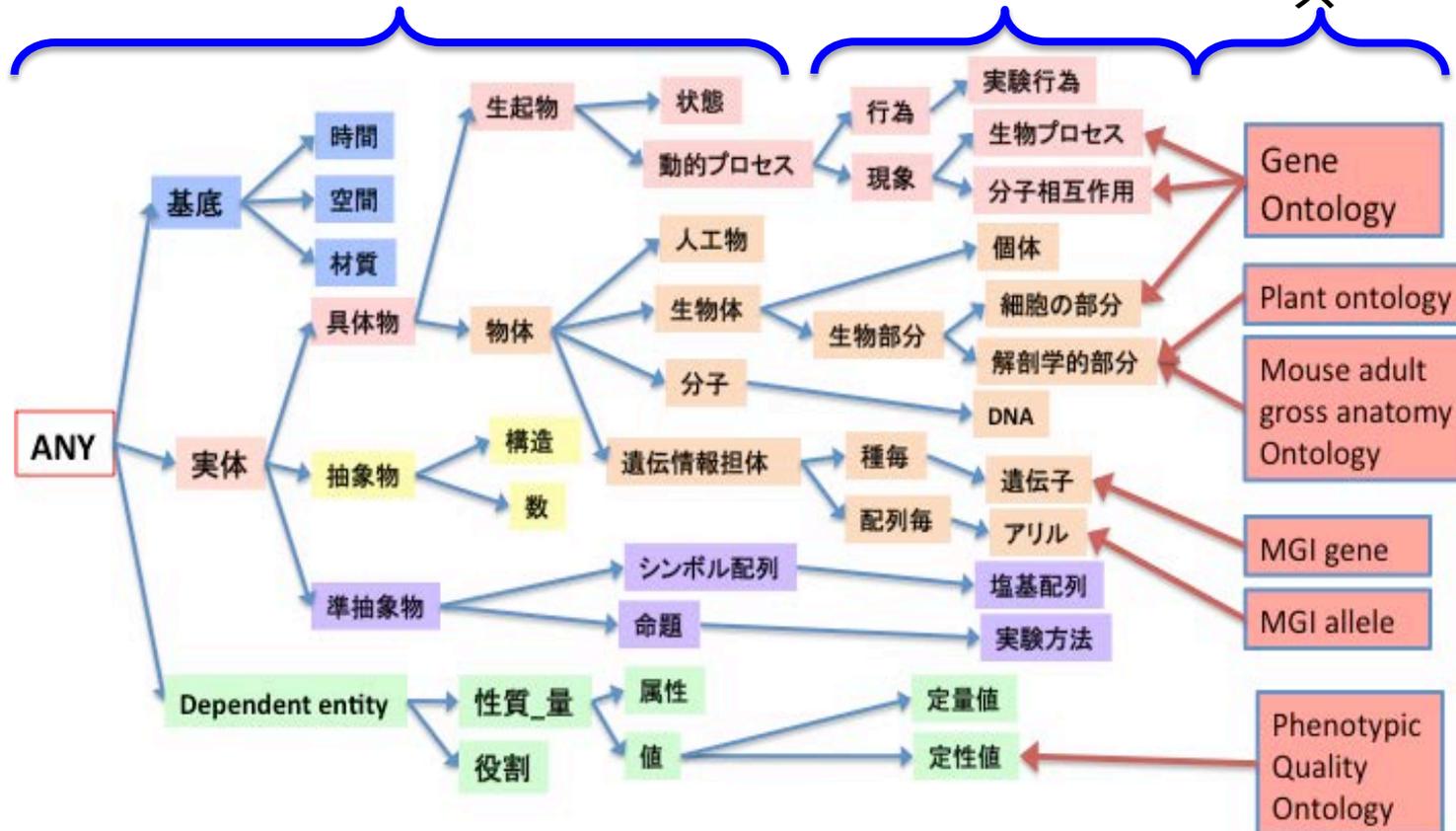


先行研究: 生命科学を記述する上位オントロジー

上位オントロジー: YAMATO
概念の普遍的な分類体系

生物学に登場する
概念: GXO

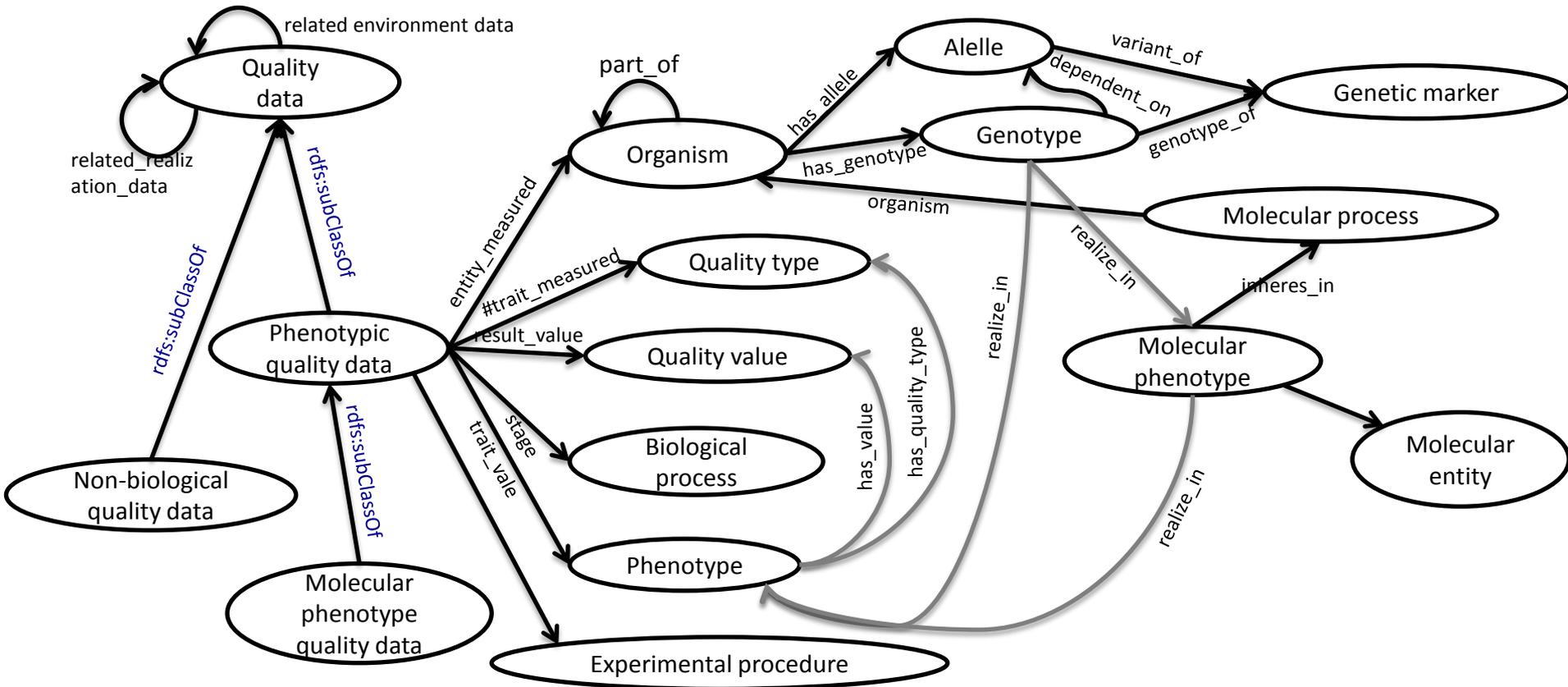
生物分野の
データベース



- 統合データベースを設計するための指標

NBDC内での連携: RDFスキーマや要素の共通化へ向けて

遺伝因子から、分子表現型を経て、環境と関連して表現型(データ)が測定されるスキーマ(概要)



NBDC各プロジェクトとオーバーラップ部分の確認を行い、データの相互利用、スキーマの標準化のための調整を行っていく(2ヶ月毎程度、SPARQLthonにて調整)

各コミュニティとの連携状況

- 国立遺伝学研究所(マウス、ゼブラフィッシュ)
- 基礎生物学研究所(メダカ)
- 熊本大学(マウス)
- 京都大学(ラット、パスウェイDB)
- 東京大学・大阪大学(臨床医学オントロジー)
- 北陸先端大学(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)
- DBCLS(微生物リソース、表現型、RDF)
- 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)
- Linked Open Data Challenge Japan (オープンデータ:産業・分野横断)

将来展望

- **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