

## 個別タスクの実施計画及び成果イメージ(案)

### 3 . 人材の育成 4 . データベース等ポータル構築

情報・システム研究機構 事務局

### 3. 人材の育成

課題	内容
キュレーター、 アノテーター、 DBマネージャー育成 のための教材作成	<b>キュレーションの実際の調査</b> 代表的なキュレーションDBである、KEGG、PMD、Mutation View等について、調査
	<b>アノテータ教育</b> ・アノテーション教育システムの構築 ・アノテータ養成のための教科書作成
	<b>DBマネージャー教育</b> ・教育資料編纂・閲覧システムの構築 ・DB管理者養成のための実習書作成

### 4. データベース(DB)等ポータル構築

課題	内容
<b>データベースポータル構築</b>	<b>DBポータル基本部分の連携システム開発</b> ； DB内容や規模を俯瞰可能な用語辞書に基づいたDB分類と、JST構築のDBポータル(WING)との間の連携システムを構築 <b>データベースポータル構築</b> ； 利用者からの書き込み追加が可能な、DBポータルサイトのプロトタイプシステムの開発
<b>WEBリソースポータル構築</b>	メソッドオントロジーのインターフェースとJSTのツールリンク集とをつなぐプロトタイプシステムの開発
<b>ポータルサイト連携</b>	「Jabion」, 「ライフサイエンスの広場」等の連携方策のアイデアを提示

# 「キュレーションの実際の調査」

- 目的

- 現在実際にキュレーションが行われている機関での作業内容や流れを調査する。

- 実施事項(図1参照)

1. 資料(文献等)および、当該施設がWebに公開している情報をもとに、キュレーション作業の実際を調査した。
2. 調査をもとに、教育教材作成のための内部レポートを作成した。

- 成果

1. 言葉の定義の明確化(人材育成の教材作成、職業としての確立にむけて)
  - 著名な生物学データベースの領域でのキュレーションとキュレーターの作業の実際が明らかになった。
2. 著名キュレーション型DBの比較(表1参照)
  - 各々のDBの人員的な規模や、キュレーション、キュレーターがDB全体のなかでどのように関与しているの  
が明らかにした。

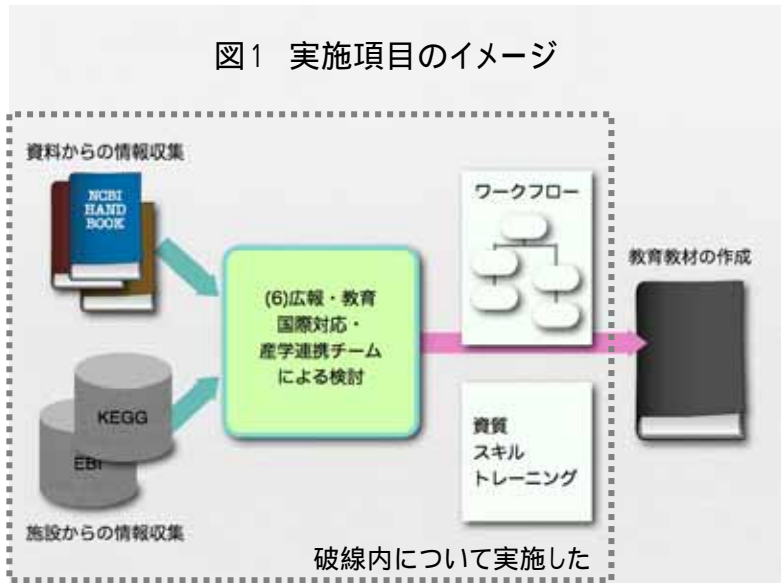


表1 代表的なキュレーション型DBのスタッフ規模

DB	組織	キュレーター	コーディネーター	ソフトウェア技術者、プログラマー	参考
RefSeq	NCBI	30	3	38	<a href="http://www.ncbi.nlm.nih.gov/RefSeq/staffcredits.html">http://www.ncbi.nlm.nih.gov/RefSeq/staffcredits.html</a>
	EBI	26	5	36	<a href="http://www.ebi.ac.uk/Information/Staff/viewgallery_seqdb.php?cid=4">http://www.ebi.ac.uk/Information/Staff/viewgallery_seqdb.php?cid=4</a>
UniProtKB	PIR	12	2	3	<a href="http://pir.georgetown.edu/pirwww/about/staff.html">http://pir.georgetown.edu/pirwww/about/staff.html</a>
	SIB	52 <sup>1)</sup>	5	13	<a href="http://au.expsy.org/people/swissproc.html">http://au.expsy.org/people/swissproc.html</a>
PATHWAY MAP BRITE KD	KEGG	21 <sup>2)</sup>	0 <sup>3)</sup>	4	<a href="http://kanehisa.kuicr.kyoto-u.ac.jp/people.html">http://kanehisa.kuicr.kyoto-u.ac.jp/people.html</a>

<sup>1)</sup> キュレーターという肩書きがなくアドバイザーとなっていたが、その方がいわゆるキュレーション作業もすると判断した。  
<sup>2)</sup> KEGGは基本的な生物情報だけでなく、化学物質から疾患情報、薬剤情報など非常に広範囲にわたるDBを作成している。21名という数字はそれら多数にわたるDBの作成にあたるスタッフの総数であり、実際に生物分野を担当しているスタッフは数名であると考えられる。  
<sup>3)</sup> いわゆるコーディネーターという肩書きはない。前出の21名スタッフが自分の担当分野において状況に応じてコーディネーター的役割も果たしてると考えられる。

# 実施項目（成果）

---

## 1. キュレーションの実際の調査

1. 「curation, curator」という言葉の英語圏での使われ方、職業としての位置を調査する  
教科書の執筆(19年度)

2. 現在実際にキュレーションが行われている機関での実際の作業の流れを調査する  
内部レポートの作成(18年度) 資料、Webからの情報収集を行い、  
教科書作成のための内部レポートを作成した

教科書の執筆(19年度)

# 「アノテーター教育」

- **目的**

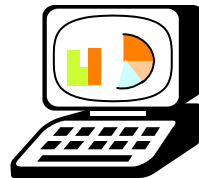
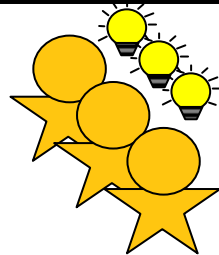
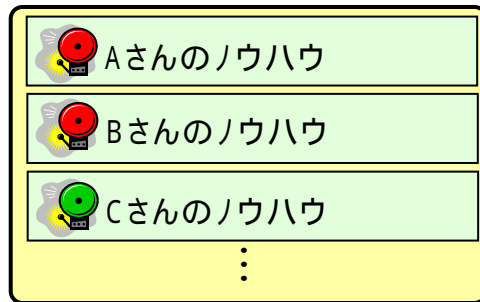
- ライフサイエンスデータ統合の基盤となる「ゲノムアノテーション」の実務に携わるアノテーターを支援する教育用システムを作成し、同時に実践的なアノテーター教育テキストを作成する。

- **実施事項**

1. ゲノムアノテーションのノウハウを抽出。
2. アノテーターによる手動アノテーションの手順を模倣・再現するプログラムを作成。
3. 「アノテーション教育テキスト」をWiki上に作成した。

- **成果**

1. 大量ゲノムアノテーションの現場で高度な技術を持つ実務者からのノウハウ抽出を行った。
2. DBやツールの利用方法やその結果の解釈を含み、実行したアノテーションの根拠を明示することでアノテーション初学者の理解を助け、学習を促進するシステムを作成
3. アノテーションにかかわる知識や注意点を文書化しゲノム解析型統合DB構築に役立つ教科書とした。



アノテーション学習者

推定結果の提示

推定根拠の明示

アノテーションの  
ノウハウ抽出と  
プログラム化



アノテーター教育用  
コンテンツの作成

# 「アノテーター教育」実施項目（成果）

## 1. ゲノムアノテーションのノウハウ抽出

- かずさDNA研究所で実際に大量ゲノム解析に携わる高度専門技術を有するアノテーターからの聴き取りならびに実務調査を行った  
実際に用いているツールやDBの利用手順や、解析結果の解釈法などのノウハウを抽出した

## 2. 抽出した情報の解析を元に、アノテーターによる手動アノテーションを模倣するプログラムを作成し、アノテーション教育システムの基盤部分とする

DBやツールの利用方法やその結果の解釈を含み、実行したアノテーションの根拠を明示することで初心者の学習を支援するプログラムを作成

## 3. ゲノムアノテーションの実践的な知識を文書化し、アノテーター教育を目的とした「アノテーション教育テキスト」を作成

上記項目で抽出されたノウハウをふくめ、具体例をあわせた実践的「アノテーター教育テキスト」コンテンツを作成

# 「DBマネージャー教育」

- **目的**

- ライフサイエンス関係のデータベースを構築・維持管理の実際を行う、DBマネージャーのための教育資料編纂・閲覧システムと実際のコンテンツ構築を行う

- **実施事項**

1. qwikwebを用いて、wikiと電子メールによって知識の蓄積が可能な教育資料編纂・閲覧システムを作成した。
2. 作成したシステムを用いてDBマネージャー養成のための実習書の「維持管理編」を作成した

- **成果**

1. wikiと電子メールを介した知識蓄積ができる教育資料編纂・閲覧システムの構築
  - Wikiだけではなく、電子メールからの入力も可能とすることでより多くの有識者から情報を集めることが可能に
2. DBマネージャー養成のための実習書(維持管理編)作成
  - 構築した教育資料編纂・閲覧システムを使って、DBの日々の管理、アップデート、バックアップを中心に作成
  - 個別の関連する事項に関して具体例も併せて解説

## 教育資料編纂・閲覧システム



## DBマネージャー養成のための実習書



# 「DBマネージャー教育」実施項目

## 1. 教育資料編纂・閲覧システムの構築

- 1) 散在する専門家の共同で知識の集積編纂による教材を作成するためのツールの調査、検討

既存ツールの利用による調査、検討

wikiと電子メールによって知識の蓄積が可能なqwikwebを採用

- 2) 一般ユーザーとの情報共有および知識編纂閲覧の整備

qwikwebをベースに教育資料編纂・閲覧に特化したシステムを作成

## 2. DBマネージャー養成のための実習書作成

- 1) 維持管理編の作成

DBの日々の管理、アップデート、バックアップを中心に個別の関連する事項に関して具体例も併せてコンテンツ作成

- 2) リファレンス編の作成

個別に有用な知識はリファレンス編として独立してコンテンツ作成



# 「データベースポータル構築」

## ・目的

- 利用者を目的にかなったDBやリソースに誘導するポータルサイトを構築する。

## ・実施事項

- DBの俯瞰や戦略立案の目的で作成したDBディレクトリと分類による一覧表示。
- 利用者向け日本語解説を付与し、今後、利用者が書き込み可能なシステムを構築。

## ・成果

- DB構築者や利用者が書き込み可能なDBポータルを構築。

## 情報の閲覧と書き込みシステム

PEPTIDOME

ペプチドを主成分に含有する形で取り出し、簡略化された方法により体系的に分類、同定し、分類過程で得られる特性を基に、化学構造、生物活性、存在量、立降構造、受容体などの情報と共に系統的に収録したデータベース。

目次(詳細)

- 1 利用方法
  - 1.1 検索のしくみ
  - 1.2 入力例
  - 1.3 データ取得方法
  - 1.4 シックシート
  - 1.5 関連データベース
  - 1.6 関連ソフトウェア
- 2 Top
- 3 その他

利用方法 [編集]

検索インターフェース [編集]

<1> MS Peaks/Peptides Information :

http://peptidome.gi.f.u-tokyo.ac.jp/cgi-bin/pepmain/search.asp  
\* Source Name

別称	ペプチドーム
提供機関	厚生労働省国立遺伝学研究所
提供国	日本
提供機関の所在地	
URL	http://www.peptidome.org/index.html
カテゴリ	生体内ペプチドデータベース
取り扱い生物種	マウス、ブタ
提供種別	Webアプリケーションによる検索および結果の閲覧
利用条件	Sequence of Probe 検索に関しては、ユーザー登録が必要
データ量	
更新頻度	
更新記録	

WZNG

WZNGデータベースポータルサイトへようこそ。ライフサイエンス分野のデータベースについての情報が収集、整理、登録されており、45649件以上の記事が存在します。ディレクトリからは、データベースの種類による分類と生物種および対象による分類でデータベース一覧できます。当サイトの「記事の目録」登録済みの分野の関連性により行いますが、今後、新規データベース、既存データベースの特長利用方法の案内など、多くの研究者の参加が期待されるよう整備します。また行は本データベースのライフサイエンス分野の関連データベース整備事業「平成19年度」より実施され、この一環として運営されています。

ディレクトリ

- ・ 種類別分類
- ・ 物質概念による分類
- ・ 生物種による分類
- ・ 対象物
- ・ 構造概念による分類
- ・ タンパク質
- ・ 複製概念による分類
- ・ 複製データによる分類
- ・ 既知データベース

関連の記事

AsIndex

URL: <http://www.genome.jp/asindex/>

提供機関: 京都大学化学研究所

提供形式: Webアプリケーションによる検索および結果の閲覧  
FTPによるダウンロード

概要: AsIndexは、タンパク質を構成する要素としてのアミノ酸の様々な特性、群を表現したデータを収集したデータベース。AsIndex1、AsIndex2、AsIndex3の3つのデータベースから構成されている。AsIndex1は、アミノ酸の物理化学的性質や生物学的特性を表現した、21の群別セットのデータベース。AsIndex2は、アミノ酸置換マトリックスのデータベース。AsIndex3は、アミノ酸のアミノ酸置換マトリックスのデータベースである。

生物種

他の関連記事: アーカイブ | 関連記事の通知

DB構築者

ユーザ

管理者

# 「データベースポータル構築」実施項目(成果)

1. 利用者およびDB構築者の助言を反映する機能の開発
  - 1) Wikipediaのような仕組みを前提にDBの日本語解説やDBの有効性や利用法を記述する機能を開発 **利用者が書き込み可能なwikiを構築。371DBの詳細説明有り。**
  - 2) 特定の研究者やシステム管理者による管理人の機能は付与  
運用方法および書き込みのインセンティブ付与方法(今後の課題)
2. WEBリソースポータルとの連携
  - 1) 別途開発するWEBリソースポータルとの連携機能を開発
  - 2) DBの内容にソフトウェアやWEBリソースが記述されると自動的にWEBリソースポータル部分にリンク形成する。メモ記述の追加を可能とする。  
**WEBリソースポータルとの連携は不完全。**  
WEBリソースの個別解説記事もwikiに追加(今後の課題)
3. 統合検索機能の開発
  - ・受託研究課題での成果物(調査報告書を含む)についても、  
統合検索できる機能を開発 **googleアプライアンス利用 設定済み**
4. 利用状況等を含む各種管理機能の開発
  - wikiの機能利用 他に統計解析ツール設定予定**

# 「WEBリソースポータル構築」



## Webリソースポータルサイト

リソースのキーワード検索  
リソースの分野、目的別提示

## メソッド

- ・カテゴリ
- ・キーワード
- ・解説(目的、特徴など)
- ・提供サイト
- ・参考文献
- ・引用文献

## ワークフロー

- ・メソッド構成
- ・カテゴリ
- ・キーワード
- ・解説(目的、構成など)
- ・提供サイト
- ・引用論文

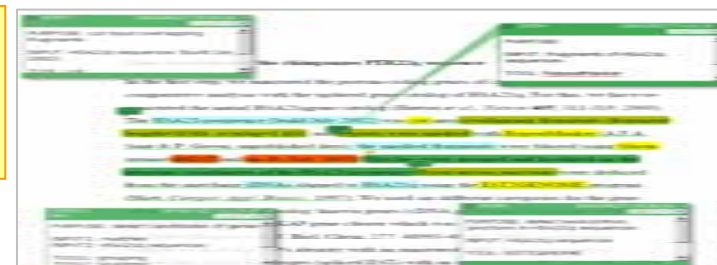
## 研究事例論文

- ・メソッドのハイライト、リンク
- ・ワークフローへのリンク



リソース詳細表示

リソース名	tRNAscarn-SE
別称	
解説	ゲノム配列からtRNAを予測するツール。tRNAの検出を速く、tRNAの共分散モデルを用いて判定する。
キーワード	処理要素別>同定/予測する>ゲノム配列構造を同定/予測する>非コードRNA領域を同定/予測する
カテゴリ	分野別ゲノム解析>アンテーション>非コードRNA領域予測>[ArabidopsisChr1_Flow0] tRNA領域を同定/予測する>Step14 分野別ゲノム解析>アンテーション>非コードRNA領域予測>[Aspergillus_Flow1] tRNA領域を同定/予測する>Step1
提供サイト	UCSC Pasteur Institute
参考文献	tRNAscarn-SE: a program for improved detection of transfer RNA genes in genomic sequence
引用文献	
入出力データ	



# 「WEBリソースポータル構築」の実施項目(成果)

## 1. WEBリソース情報のデータベース化

WEBリソース(ソフトウェア含む)に、ワークフロー(実験方法)のような複雑なデータ構造を収録できるように、RDBSを利用しデータベース化する。

**解析ツール456のカatalog化とある領域のワークフロー29の構造化を実施。**

## 2. 検索機能の開発 **下記を開発した。**

- 1) ユーザが適切なリソースを容易に探し出せるように、適切なインデックスを用意する。
- 2) 分野別、目的別、解析対象別などリソースを絞り込むための階層構造型のリソース提示機能を開発する。

## 3. 表示機能の開発

通常が表示に加え、ワークフローレベルでの目的やツール構成を表示する機能、実際にそのワークフローを利用している研究事例論文のハイライト表示機能を開発する。

**4論文について試験的に実行可能とした。**

## 4. データベースポータルとの連携 **連携は未完成(今後の課題)**

データベースを利用する個別解析ソフトウェアやワークフローの表示ページから、データベースポータルサイトへのリンクをもうける。

**1部の領域について、一通りのプロトタイプ構築を行った。**領域の拡充とそのためシステムの手法の開発(今後の課題)

# 「ポータルサイト連携のための調査」の実施項目(成果)

1. 「Jabion」日本語バイオポータルサイト (<http://www.bioportal.jp>) の調査
2. 「ライフサイエンスの広場」文部科学省ライフサイエンスポータルサイト (<http://www.lifescience-mext.jp/index.html>) の調査
3. 「UMIN」大学病院医療情報ネットワーク (<http://www.umin.ac.jp>) の調査
  - ・一般公開用ページの掲載内容の概要
4. 上記1, 2の調査結果を基に有用な連携の仕組みアイデア提示
  - ・掲載内容から相互利用に有効な事項を選び、組み合わせを提示。
  - ・利用者の目的からサイトマップを提示。
    - ・有効なアイデアの実装(今後の課題)
    - ・ポータル連携と個別ポータルへの支援枠組みの提示(今後の課題)

## 「文献情報との連携調査」の実施項目(成果)

JSTが運営している文献情報提供事業で提供されている文献情報を活用し、遺伝子情報に遺伝子機能を付加する調査を実施した。100程度の抄録から遺伝子関連の表現型や疾患名を抽出した。

GenLibi (<http://gene.jst.go.jp/>) としてテスト中。(統合DB予算はないが、提案書に記載) 遺伝子と関連していると判断する基準や実際の抽出作業(今後の課題)