

平成 13 年度
生物多様性情報データベース構築
フィジビリティ・スタディ (FS) 報告書

報告日：2002年3月4日

グループ名：無脊椎動物の生理機能の多様性に関する
データベースの開発
第二グループ 無脊椎動物の神経細胞データベース

グループ代表者氏名：桑澤 清明

1. 構築検討したデータベースについて

データベースの概要及び特徴	<p>【データベースの概要】</p> <p>第二ワーキンググループでは、無脊椎動物の生理機能のうち「神経細胞」を対象を限定しデータベースを構築する。また生物群として最も原始的な神経細胞を有するヒドラ類から、現在最も動物種記載数の多い「昆虫」までを対象とし、できるだけ広範囲の生物群の記載を多様性データベースとしてまとめるシステムの構築を模索している。</p> <p>FSでは、1. 神経細胞データベースのプロトタイプを作製し、2. 各動物群の担当者が、実際にデータ入力をおこないその過程で今後必要とされる作業時間や人件費等の見積りを行う事を主目的とした。</p> <p>【データの特徴】</p> <p>データ形態は、テキストと画像で、それぞれ原著論文からの引用になる。テキストは、入力者がレビュー的に要約し入力者の著作権とする。画像は、モニター上で判別できる程度の解像度を用い、印刷出力を前提とした高解像度画像は扱わない。また、必要に応じて著作権処理を行う。</p> <p>【データベースの特徴】</p> <p>本データベースでは、無脊椎動物において神経細胞として同定された細胞のリストと、それぞれの細胞において記載された形質のリストを作製する。</p> <p>神経細胞は、形態的・機能的に同定されるが、過去に同定された神経細胞と新たに同定された神経細胞が全くの同一細胞であるか、或いは異なる細胞であるかという判定は、その後行われる場合が多い。従って、本データベースでは分類データベースで使われている「タクソソコンセプト」と同様の概念として「神経コンセプト」を導入し、「神経名称テーブル」と独立させることで、将来的な神経細胞名の分離・統合に対応させた。</p> <p>一方、形質データは「解剖学、受容体の発現状況、伝達物質の種類、神経回路としての入・出力、行動との関連、感覚入力、運動出力」以上8カテゴリーに分類してデータ入力を行っている。また、各形質について、その実験的な根拠も同時に入力することで、データの信頼性の判断基準を示した。</p> <p>それぞれのデータは、原記載である引用文献と対応し、文献リストに関連付けしている。</p> <p>【他の類似のデータベースとの相違点】</p> <p>複数の動物群を対象とし、特定の形質に着目して作られたデータベースには、DNAデータベースやたんぱく質データベースが挙げられる。これらは何れも遺伝子型に基づいた発現分子レベルでの記載であり、その対象種は一部の実験生物に限られている。本データベースが目指しているより高次の「形質」は、表現型に依存し、上記のデータベースに比べより多くの種において記載が行われている。また、網羅的</p>
---------------	--

	<p>な種の記載を主目的とした分類データベースに比べて、対象種数では劣るものの、記載形質の情報量は圧倒的に勝る。特に、種間の比較を行う場合、共通の形質に着目することでより広範囲の生物群を対象とした比較検討が可能になる。</p> <p>神経細胞に関する記載のデータベース化は、実験材料を中心として既にいくつかのプロジェクトが進行中である。情報科学的に見ると、米国ではヒトの脳神経データベースが基幹として走っていて、そこから派生した形でアメフラシ神経系のデータベース化が行われている。一方、デジタルアーカイブ・データを共有する試みとしては、ショウジョウバエの神経の形態画像データベースが有る（FryBrain）。このFryBrainは、様々な突然変異体の形質を記載するうえで役立つシステムを採用している。また、線虫では全ての体細胞が、その発生の履歴から同定され全神経細胞に名前が付けられている。これらの各細胞ごとに、判明している形質が記載される形でデータベース化が進められている。いずれも、単一生物材料に特化したデータベースであり、本データベースが目指す多様性の記載と性格を異にしている。</p>
<p>類似のデータベース</p>	<p>神経細胞の特徴を記載するという点で、下記のデータベースが類似しているが、多様性に着目して複数の生物群に渡った比較を可能としているデータベースは類を見ない。</p> <p>【国外のみ】</p> <p>ショウジョウバエ脳神経系の解剖 (http://www.flybrain.org/)</p> <p>センチュウ全神経細胞の記載 (http://elegans.swmed.edu/)</p> <p>アメフラシ全神経 DB プロジェクト (http://mollusc.med.cornell.edu/)</p> <p>ヒト全神経 DB プロジェクト (http://www.nimh.nih.gov/neuroinformatics/index.cfm)</p>
<p>生物分野</p>	<p>生理学（特に比較生理学）、形態学、行動学、生態学、発生生物学、分類学</p>

2. データ源およびデータの現在の状況について

<p>データ源</p>	<ul style="list-style-type: none"> ・原則として原著論文よりデータを収集する。古いデータについては二次的に引用されたレビューや本から記載する。 ・各担当者の所蔵する文献、或いは担当者が新たに収集する文献からデータを収集する。文献は、図書館保存、個人所蔵、リプリント等が想定される。 ・登録システムを並行して作製することで、新規データについては研究者の投稿により追記する。 <p>1) H13年度データ取得可能なデータ源(件数など):</p> <ul style="list-style-type: none"> 神経細胞: 60 形質: 610 文献: 61 <p>2) H14年度データ取得が可能なデータ源の見込み(件数など):</p> <ul style="list-style-type: none"> 神経細胞: 240 形質: 1850 文献: 300 <p>3) H15年度以降のデータ取得が可能なデータ源の見込み(件数など):</p> <ul style="list-style-type: none"> 神経細胞: 6,000 形質: 42,000 文献: 5000 対象種数: 約70種
<p>データの発生・収集場所</p>	<p>委員の主たる分担は下記の通りで、各データは委員の所属する機関で発生する</p> <ul style="list-style-type: none"> 昆虫: 蟻川 謙太郎(横浜市立大学)、神崎亮平(筑波大学) 甲殻類: 蔵本武照(筑波大学下田臨海実験所)、木原 章(法政大学) 軟体動物: 桑沢清明(東京都立大学)、黒川 信(東京都立大学) 腔腸動物: 小泉 修(福岡女子大学) <p>データ収集は、各担当者が自機関で行い、そこで入力補助者(大学院生を想定)を調達する。必要に応じて、図書館相互利用(国立大学図書館協会・公立大学図書館協会・私立大学図書館協会)により不特定多数の他機関より文献を調達する</p>
<p>他機関の場合の連携協力について</p>	<p>日本比較生理生化学会を対象として、公募型のデータを入力を行う予定である。現時点では、学会誌「比較生理生化学」を通じて当データベースの内容を紹介した。今後ワークショップなどを開催し、学会員の参加を募る予定である。</p>

<p>データフォーマット</p>	<p>本データベースは、神経コンセプト DB、データ DB、文献 DB、名前 DB の 4 テーブルから構成される。その内の主たるテーブル 2 つの主たる記入項目は以下の通りである。</p> <p>【神経コンセプト・テーブル】</p> <table border="0"> <tr><td>神経コンセプト ID</td><td>テキスト</td></tr> <tr><td>神経コンセプト名</td><td>テキスト</td></tr> <tr><td>対象種学名</td><td>テキスト</td></tr> <tr><td>対象種一般名</td><td>テキスト</td></tr> <tr><td>対象種画像</td><td>画像</td></tr> <tr><td>形態概要</td><td>テキスト</td></tr> <tr><td>機能概要</td><td>テキスト</td></tr> <tr><td>分類</td><td>テキスト</td></tr> <tr><td>引用文献 ID</td><td>テキスト</td></tr> <tr><td>入力者名</td><td>テキスト</td></tr> </table> <p>【神経データ・テーブル】</p> <table border="0"> <tr><td>データ ID</td><td>テキスト</td></tr> <tr><td>データ種</td><td>テキスト</td></tr> <tr><td>解剖学</td><td>テキスト</td></tr> <tr><td>解剖学</td><td>テキスト</td></tr> <tr><td>受容体の発現状況</td><td>テキスト</td></tr> <tr><td>伝達物質の種類</td><td>テキスト</td></tr> <tr><td>神経回路入力</td><td>テキスト</td></tr> <tr><td>神経回路出力</td><td>テキスト</td></tr> <tr><td>行動との関連</td><td>テキスト</td></tr> <tr><td>感覚入力</td><td>テキスト</td></tr> <tr><td>運動出力</td><td>テキスト</td></tr> <tr><td>その他</td><td>テキスト</td></tr> <tr><td>根拠</td><td>テキスト</td></tr> <tr><td>図 1 の解説</td><td>テキスト</td></tr> <tr><td>図 1</td><td>画像</td></tr> <tr><td>図 2 の解説</td><td>テキスト</td></tr> <tr><td>図 2</td><td>画像</td></tr> <tr><td>図 3 の解説</td><td>テキスト</td></tr> <tr><td>図 3</td><td>画像</td></tr> <tr><td>引用文献 ID</td><td>テキスト</td></tr> <tr><td>神経コンセプト ID</td><td>テキスト</td></tr> <tr><td>入力者名</td><td>テキスト</td></tr> </table>	神経コンセプト ID	テキスト	神経コンセプト名	テキスト	対象種学名	テキスト	対象種一般名	テキスト	対象種画像	画像	形態概要	テキスト	機能概要	テキスト	分類	テキスト	引用文献 ID	テキスト	入力者名	テキスト	データ ID	テキスト	データ種	テキスト	解剖学	テキスト	解剖学	テキスト	受容体の発現状況	テキスト	伝達物質の種類	テキスト	神経回路入力	テキスト	神経回路出力	テキスト	行動との関連	テキスト	感覚入力	テキスト	運動出力	テキスト	その他	テキスト	根拠	テキスト	図 1 の解説	テキスト	図 1	画像	図 2 の解説	テキスト	図 2	画像	図 3 の解説	テキスト	図 3	画像	引用文献 ID	テキスト	神経コンセプト ID	テキスト	入力者名	テキスト
神経コンセプト ID	テキスト																																																																
神経コンセプト名	テキスト																																																																
対象種学名	テキスト																																																																
対象種一般名	テキスト																																																																
対象種画像	画像																																																																
形態概要	テキスト																																																																
機能概要	テキスト																																																																
分類	テキスト																																																																
引用文献 ID	テキスト																																																																
入力者名	テキスト																																																																
データ ID	テキスト																																																																
データ種	テキスト																																																																
解剖学	テキスト																																																																
解剖学	テキスト																																																																
受容体の発現状況	テキスト																																																																
伝達物質の種類	テキスト																																																																
神経回路入力	テキスト																																																																
神経回路出力	テキスト																																																																
行動との関連	テキスト																																																																
感覚入力	テキスト																																																																
運動出力	テキスト																																																																
その他	テキスト																																																																
根拠	テキスト																																																																
図 1 の解説	テキスト																																																																
図 1	画像																																																																
図 2 の解説	テキスト																																																																
図 2	画像																																																																
図 3 の解説	テキスト																																																																
図 3	画像																																																																
引用文献 ID	テキスト																																																																
神経コンセプト ID	テキスト																																																																
入力者名	テキスト																																																																
<p>デジタル化されたデータについて</p>	<p>1) 現在保有するデータ総件数と保存媒体 (H13 年度末見込み)</p> <p>神経細胞 : 6 0 形質 : 6 1 0 文献 : 6 1 (サーバーのハードディスク及び C D ? R へのバックアップ)</p> <p>2) 平成 1 4 年度の見込み</p> <p>神経細胞 : 240 形質 : 1850 文献 : 300</p> <p>3) 平成 1 5 年度以降の見込み</p> <p>神経細胞 : 6,000 形質 : 42,000 文献 : 5000 対象種数 : 約 7 0 種</p>																																																																

<p>データ・ベースの実現方式とデータのクオリティ</p>	<ol style="list-style-type: none"> 1 .既に査読を受けて出版された文献に基づくデータのみを収録する。 2 . 入力者の氏名を明記し、その後の改定の責任者とする。 3 . 形質データについては、その判定の裏付けとなるデータ(根拠)も記載する
-------------------------------	--

3 . FS で得たデータベース化する際の知見について

<p>データ・フォーマットやデータの加工内容などについて</p>	<p>データの標準化</p> <p>多様性データベースとして形質の多様性を記載する際に、各分類群について共通の尺度を設定することが第一の課題として挙げられる。「神経細胞」は、その機能として共通性を有するが、形態、性質共に多様である。現在、それらの多様な形質を「解剖学、受容体の発現状況、伝達物質の種類、神経回路としての入・出力、行動との関連」以上6カテゴリーに分類してデータ入力を行っているが、その記載法についての標準化には至っていない。その為に、入力作業が遅れ気味である。つまり、入力時の判断に要する負担が大きな限定要因となっているものと考えられる。しかし、本データベースは未だ世界に類を見ないデータベースでありデータの標準化に向けて今後進むべき方向性が示されれば、それだけでも大きな成果だと考えている。</p> <p>一方、「神経コンセプト」は分類と類似して、今後複数のコンセプトが一つにマージしたり、逆に単一のコンセプトが複数にスプリットすることが予想される。この取り扱いについては、分類データベースで既に示されているデータモデルを適用する事で解決可能である。</p> <p>入力フォームの改良</p> <p>データ入力時の研究者の負担を軽減し、かつ誤入力を防ぐ意味で各形質に特化した入力フォームの開発が必要である。現在、特定の神経伝達物質に着目して特化した入力フォームのプロトタイプを試作中である。今後データが増えるにしたがって、それらを参考にした新たな入力フォームの作製が必要となる。</p> <p>入力フォームに基づいた柔軟なデータ構造</p> <p>特定機能に特化した専用入力フォームを構築した場合、データの更なる要素化が行われる。それらは、現在のデータモデルにおいてはフォーマット化されたテキスト形式で一次データとした保管されるが、様々な検索に効率的に対応するためにはオブジェクトDB</p>
----------------------------------	--

	<p>として入力フォーム等のメソッドをデータ内に定義できるようにすることが望ましい。</p> <p>現在のパソコン(M a c)を用いたプラットフォームではその実現は難しいが、将来的に汎用プラットフォームに移行したうえで、オブジェクトDBへ移行することが望まれる。</p> <p>汎用プラットフォームへの移行</p> <p>現在のF Sでは、予算的に活動が絞られているために、データの標準化に最もエネルギーが注がれている。当初予定された予算が確保された場合は、米国に於ける汎用プラットフォームでのエンジン開発も考えていたが、現段階ではそれは難しい。今後もデータの標準化に時間がかかることが予想されるため、しばらくはパソコンを用いたプラットフォームでの運用を計るが、最終的には汎用プラットフォーム移行して、それを維持管理できる人材を得たいと考えている。</p>
--	---

4 . その他

<p>データベースを公開する上での問題点の解決について</p>	<p>データベースの公開とプライオリティーについて</p> <p>本プロジェクトは、他に類を見ないオリジナリティーが高い内容を持つ。従って、データベースを世界に向けて公開する時点では、それなりのプライオリティーの確保が重要であるものと考えられる。本F Sで得られた知見を元に、無脊椎動物神経データベースのデータ標準化について論文を公開し、それに続いてデータベースの公開を行うことが望ましいと考えている。</p> <p>著作権の問題について</p> <p>本データベースのデータソースは全て論文として出版されたデータである。文章については、入力者がまとめ直すことでReviewと同様の扱いになり、著作権は入力者に帰属する。入力者からは予めデータベースとしての使用承諾を得ることで、著作権問題はクリアされる。図版については、印刷に耐えられない程度の解像度に落とすことで、複製のメリットを無くす手段を用いる。しかし、その後出版社への利用依頼が必要に成るものと考えられる。但し、その許可を得るためには、データベースの概要が明らかにならないと、手続きを開始することが出来ない。今回のF Sに置いては、全て自ら著作権を持つ研究者によってデータ入力が行われるため、大きな問題には成らないが、今後データベースを公開した時点で、第三者の画像データ利用の許可を考慮したい。</p>
---------------------------------	--

その他	<p>本データベースは未だ世界に類を見ない、新しい試みである。多様性を、単なる種数のみで捕らえることなく、その形質の豊かさをデータベース化する事は、バイオリソースとして生物を捕らえるうえにおいて、必ず重要な役割を果すものと考えている。将来、日本発の独自の多様性データベースとして世界に貢献する意味で、ぜひこの様な試みを続けたいと考えている。</p>
-----	--

5 . 確立できた推進体制(具体的な参加メンバーをリストアップしてください)

<p>開発責任者</p>	<p>木原 章 (法政大学・第一教養部・講師)</p>
<p>研究協力者</p>	<p>桑沢 清明 (東京都立大学・理学研究科・教授)</p> <p>蟻川 謙太郎 (横浜市立大学大学院・総合理学研究科・教授)</p> <p>神崎亮平 (筑波大学・生物科学系・助教授)</p> <p>蔵本武照 (筑波大学・筑波大学下田臨海実験センター・教授)</p> <p>黒川 信 (東京都立大学・理学研究科・助教授)</p> <p>小泉 修 (福岡女子大学・人間環境学部・教授)</p> <p>Chauncey W. Bowers (米国ベックマン研究所・神経科学部門)</p> <p>Lisa Dahm (米国ベックマン研究所・Research Computer Support Core)</p>

6 . FS 後の推進スケジュール

	H 1 4	H 1 5	H 1 6	H 1 7
1.データベース基本設計	■			
2.データベース詳細設計		■		
3.情報機器の導入		■	■	■
4.データベースプログラミング			■	
5.データ作成・入力	■	■	■	■
6.運用試験				■
7.試験公開				■
8.公開				