

学術データベースとしての生物形態画像データベース

月井, 雄二 / 木原, 章

(出版者 / Publisher)

法政大学教養部

(雑誌名 / Journal or Publication Title)

法政大学教養部紀要. 自然科学編 / 法政大学教養部紀要. 自然科学編

(巻 / Volume)

118

(開始ページ / Start Page)

31

(終了ページ / End Page)

45

(発行年 / Year)

2001-02

(URL)

<https://doi.org/10.15002/00004664>

学術データベースとしての 生物形態画像データベース

木 原 章
月 井 雄 二

要 約

生物系研究者が所蔵するさまざまな学術情報をデータベース化し、知的基盤として活用するための手法を研究した。特に、形態情報を画像或いは動画データベースとして蓄積・公開するための手法の開発を行い、「原生生物情報サーバー」としてその事例を公開してきた。また、並行して他の形態画像データベース（昆虫、植物標本、花のカラー画像等）のデータベース構築のコンサルティングを行う過程で、その手法を或る程度ルーチン化することに成功した。今日までにルーチン化できた手法は、①35 mm カラーライド画像のデジタル化とその公開法。②動画と静止画を同一サンプルから記録し、同一インターフェイスで公開するための手法。③動画的表现を用いた、タイムラプス画像やマルチフォーカス画像のオーサリング・ブラウジング手法。④画像データと分類記載データのリンク手法。⑤電子顕微鏡ネガ画像のデジタル化。⑥古い文献データを画像データとして公開するための手法、である。これらの手法は、他の生物系学術データに応用可能な部分であり、具体的なマニュアルサイトを立ち上げることで、研究者がデータベースを構築する際のサポートシステムとして活用される予定である。

導 入

研究資材データの生産者は研究者である。研究者は、日々の研究の過程で、写真や計測データなど様々な資材情報を作成している。しかし、それら膨大な

量の資料情報の中で、論文等の形で公にされるのはごく一部にすぎない。従来はその多くが研究者の手許に非公開のまま残されていた。これらの資料情報をデータベース化しネットワーク上で公開するシステムが確立すれば、学術研究に役立つ重要な知的基盤となるはずである（月井ら 1997, Ogata ら 1999）。

DNA データベースはまさにそのような資料情報のデータベース化、ネットワーク公開の先がけといえるだろう。しかし、学術研究、とくに生物系の研究分野で作成される資料データは画像や数値データなど多種多様であり、それらのデータベース化・ネットワーク公開を、DNA 情報と同様に公的機関に頼ることは事実上不可能といえる。したがって、生物系研究資料データの多くは、研究者が自らの手でデータベース化しネットワーク上で公開する以外にない。本研究では、そのような研究者自身によるファクトデータベースの構築を促進するために、応用範囲の広い基本手法の簡素化及びルーチン化を行い、その手法を公開することで、研究者のデータベース作製の負担の軽減できるようなサポートシステムの構築を目指している。

方 法

本研究は、以下に示す具体的な DB 構築作業・あるいはそのコンサルティングにおいて、具体的なサポートシステムの在り方を研究・開発した（図 1）。

- 1) 研究者が自らの手で構築するデータベースの一例として「原生生物情報サーバー」（原虫 DB, 参考サイト 1）を構築し、その作業の過程で、画像データのデジタル化・蓄積手法、動画データの撮影・編集・蓄積手法、それらのデータを Web ページを通じて公開する手法の適正化を行った。
- 2) 米国ベックマン研究所で蓄積された電子顕微鏡ネガ画像をテクニシャンを使って画像データベースに取り込むためのルーチンを構築した。
- 3) 「日本産アリ類カラー画像データベース」（アリ DB, 参考サイト 2）構築のコンサルティングを行い、その中で分類記載データ、分類検索キーワード、標準画像データのデータ構造を決定し、リンクを辿りながら検索・同定するシステム設計を行った。
- 4) 「牧野標本館タイプ標本データベース」（牧野 DB, 参考サイト 3）構築のコンサルティングを行い、その中で植物標本画像特有の画像閲覧手法を開発した。



図1 Webで公開された生物画像データベース

原生生物(左上), 日本産アリ類(右上), 牧野標本館(左下), アサガオ(右下)のホームページ

- 5) 著作権の切れた, 或いは著作者の許可を得た入手困難な古い文献情報を, 簡易画像データベースとして配信するためのルーチンを構築した。
- 6) 「日本産アリ類カラー画像データベース」, 「アサガオ画像データベース」(アサガオDB, 参考サイト4) 構築のコンサルティングを行い, その中で既に絶版となった図鑑を著者の指示の元に Web 化する手法の適正化を行った。

結 果

(1) 35 mm スライド画像のデジタル化とそのデータの蓄積手法

35 mm カラースライドは, 多くの領域で色再現性の高いカラー画像データ

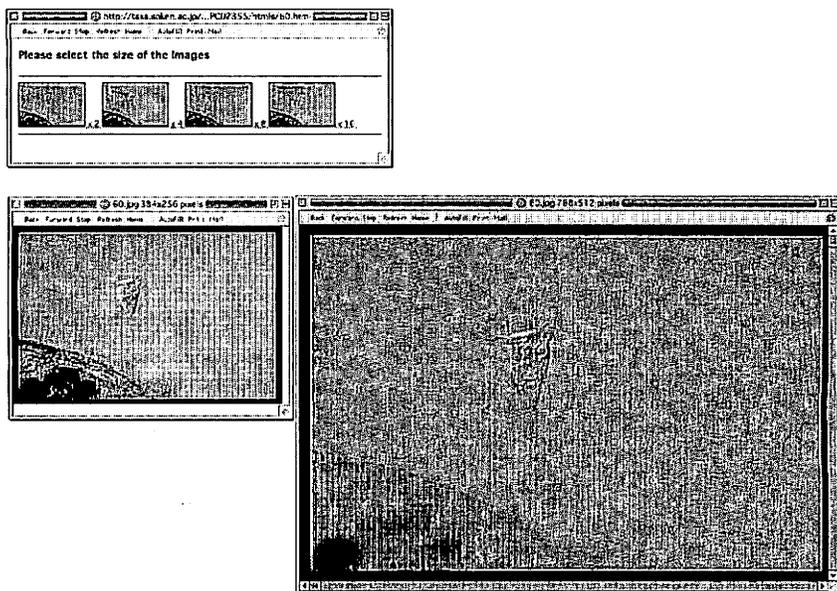


図 2 5 種類の画像サイズの設定

原生物情報サーバーのツリガネムシの画像。必要なサイズを選んで表示することができる。

のメディアとして普及している。この画像データを効率良く、かつ色再現性を失うことなくデータベース化する手法として、Kodak 社の Photo CD システムを利用した。Photo CD には、解像度と色再現性の高さのみならず、取り込んだ媒体にユニークなシリアル番号が添付される利点がある。その番号はそのままデータベース化におけるユニークなインディックス番号として利用することができた。また、Photo CD から取りだされる 5 段階の解像度設定の画像は、データベースにおける、アイコン、サムネール、プレビュー、拡大画像等にそのまま利用することが可能である。

Photo CD の色信号 (YCC 方式) をコンピュータの色信号 (RGB) に変換するための手法と、5 段階の解像度の画像をオリジナルインディックス番号と組み合わせてファイル化する手法をマニュアル化した (参考サイト 5)。

現在、本手法を用いて約 1 万 5 千枚の画像をデジタル化、データベース化し「原生物情報サーバー」に組み込んでいる (図 2)。また、同様の手法が他のデータベースでも実際に利用されはじめている。

(2) 動画データベースの構築手法

特に原生生物のように生きた状態でのみ形態的な特徴を示す材料では、静止画だけではなかなか特徴がつかみにくい。そこで、生きた細胞の動態を動画として記録し、形態情報データに追加することを試みた(図3)。しかし、その過程で同一場面を静止画として記録することで、より細部の情報を提示する必要性も明らかになった。

本研究の動画データベース構築では、まず原生生物を対象とした形態画像の記録手法として、顕微鏡専用静止画・動画同時撮影装置を構築した。この装置では、静止画データを従来の35mmスライドデータとして記録する一方で、動画データはビデオカメラを通じてDV規格のテープに収録する。DV規格は、コンピュータ動画処理に親和性が高く、またデジタル化されているので劣化の無い複製が可能な点で、データベース化の一次データの記録方式としては最適である。

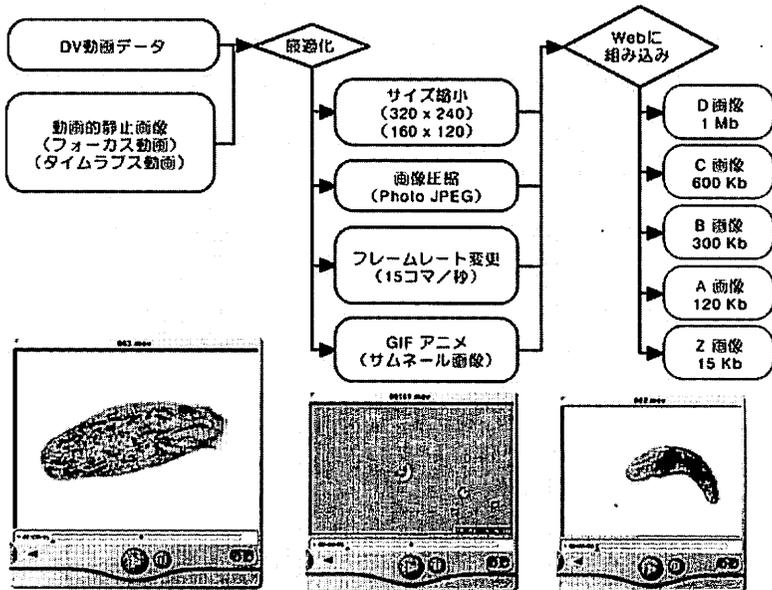


図3 動画データベース作製の流れ図

顕微鏡から得られた動画を圧縮・適正化して配信する。

DV テープに記録された動画データは、媒体の安定性や収納性に優れているが、内容の確認、編集と言ったランダムなアクセスには向かない。そこで、DVD-RAM ディスク上に DV 規格デジタルビデオ画像ファイルとして収集し、そこからクリップを切り取ることで、動画圧縮の元データ（二次データ）を複製した。

最終的にデータベースに登録する動画ファイルは約 10 秒程度のビデオクリップ（約30 Mb）として切り出し、単一クリップあたり 4 段階のファイルサイズに圧縮し（約 1 Mb, 500 Kb, 200 Kb, 100 Kb）、更に Web ページにアイコンサイズでインライン表示するための GIF アニメーション画像（約 20 Kb）も作成した。圧縮の手順は、画像の縮小あるいはトリミングによる画像サイズの縮小、Photo JPEG によるフレーム内画像圧縮、秒あたりのフレーム数（オリジナルは 30 フレーム）の削減（約 15～8 フレーム/秒）、以上三者を組み合わせで行った。

圧縮したビデオクリップは、静止画とほぼ同様のフォルダ構造でデータベースに組み込んだ。ただし、この場合ファイルサイズが大きいため将来的なデータベースのポータビリティが悪くなると予想される。この問題を解決するために、今後新たなデータの分散管理手法を考案する必要性が提示された。

(3) デジタルビデオカメラを用いたマルチフォーカス画像・ タイムラプス動画作製手法

高解像度デジタルビデオカメラを顕微鏡に接続し、顕微鏡画像を直接デジタル画像ファイルとして記録するシステムを構築した。このシステムでは、フレーム間で座標軸を完全に一致させる事が可能なので、スタック化したマルチフォーカス画像（図 4）の記録や長時間の間欠画像記録（タイムラプス記録）が可能

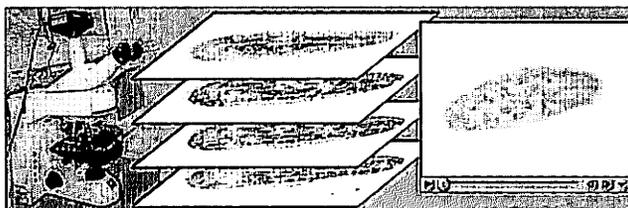


図 4 マルチフォーカス・アニメーション

焦点をずらして撮影した複数の画像をスタック化し、アニメーションとして表示する。

である。このシステムで撮影した画像を動画化してデータベースに組み込むことで、従来デジタル配信データとして提供が難しかった、3次元的な形態や、長時間の生命現象の記録をデータベース化することが可能となった。

(4) 分類データベースの構造化について

Web上に展開されるファクトデータベースでは、最初からHTML文としてデータを提供するケースが多い(例:原虫DB)。この方式は、内容の更新がさほど頻繁でない場合は作製期間を短くできるという点で効率的で、外部サーチエンジンのページ検索機能とブラウザのページ内文字検索機能の組み合わせで、ある程度のDB機能を発揮することが可能である。しかし、一方で大きな分類項目等の改定を行う場合は、そのカスケード下に有るページを全面的に改定する必要性が発生する。

形態分類DBの一例である「アリDB」の全面改定を行うに当たって、データの構造化を試みた(図5, Kihara等1997)。「アリDB」では、分類指標あるいは分類記載と標本画像の関連付けがDBの性能を決める大きな要因であった。基本データは、分類名(亜科名, 属名, 亜属名, 種名)をインデックスとし、命名者やシノニムをまとめた「名前DB」、現行種名をインデックスとし分類学的記載を列挙した「記載DB」、画像番号をインデックスとした「画像DB」、更に「文献DB」「分類キーDB」、以上5つのフラット・データベースにまとめた。これらのDBをリレーションしHTML文を自動生成するプログラムを作成した。その結果、内容とHTML出力形式の更新を、それぞれ独立で行うことが可能になった(Ugawaら1999)。

この方式を取り入れることで、「標本記載データ(分類データ)」と「標本画像データ」を元に「牧野DB」のHTML化ルーティンの作製も可能となった。

(5) 電子顕微鏡画像データのデジタルアーカイブシステム構築

電子顕微鏡画像は特殊サイズのネガに記録される。大量のネガ画像を元データとして保持する場合には、その閲覧が難しく、その状態から共通のリソースになるような画像データの提供は困難である。まず、研究者個人レベルでネガデータの閲覧が簡単に行うことができるシステムとして、ネガ画像のデジタルアーカイブ作製システムを構築した(木原・池田1999)。

電子顕微鏡写真の特徴は、高コントラスト・低ラティチュードである点であ

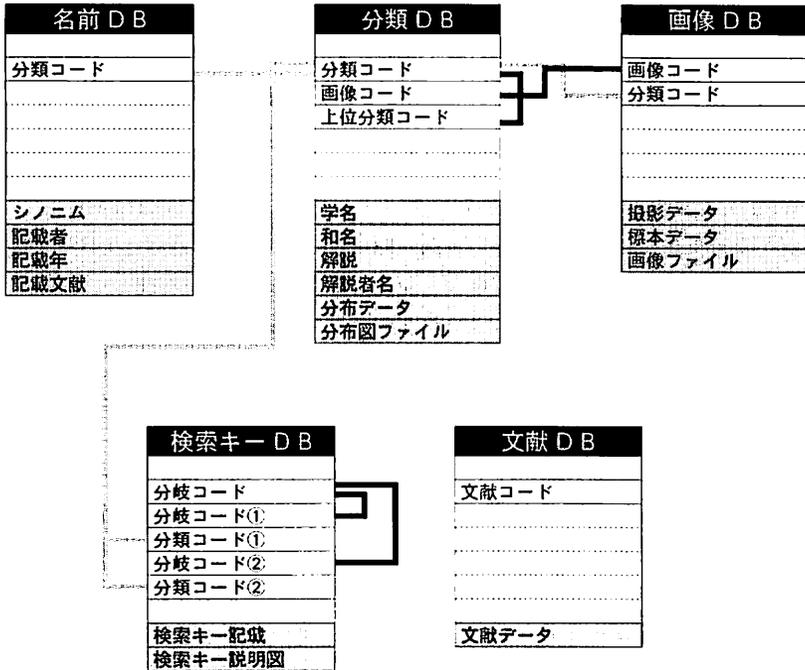


図5 リレーショナルデータベースの構造図

日本産アリ類データベースでは、上記の5種類のフラットデータベースをリレーションすることとて、適正な構造化が可能になった。

る。従って、ネガフィルムのごく狭い幅の濃度差をできるだけ豊かな階調幅でデジタル化する必要がある。そのために、できるだけ階調レンジの広い(3.2 D) イメージスキャナーを選定し、かつフレーム毎に自動露光補正を行いながら最適化するデジタル化ルーティンを作製した。デジタル化した電顕画像データは、粒子性が高くJPEGのように高頻度の濃淡変化を削除するタイプの圧縮法には適していない。従って、一次画像データはTIFファイル形式で記録した。電顕撮影時にネガに写し込まれるネガのシリアル番号をファイル名として利用し保存した。

本システムを実際に利用した結果、テクニシャンクラスの補助要員を約2時間講習することで、一時間あたり20枚程度の入力スピードを確保することができた。現在、本システムにより個人閲覧目的のデータアーカイブが構築され

つつある。

(6) 文献データのデジタル化

学術刊行物には既に絶版となり入手不可能な文献が多い。そのうち、著作権が切れた文献、あるいは著作者・出版社の許可が得られた文献については、ネットワーク上で閲覧できるデジタルリソースとして公開することが可能である。そこで、古い文献のデジタル化の実例として、既に著作権が切れた文献のデジタル化と、著作者と出版社の了解を得た上での絶版本のデジタル化を行い、その手続きの有効性を調べた（参考サイト7）。

著作権が切れた過去の文献は、全てのページをイメージスキャナーで取り込み、中解像度（325×552ピクセル）、高解像度（540×917ピクセル）のGIF画像ファイルとして保存した。各ページの画像には、JavaScriptを用いてページ順、指定のページ、索引にジャンプできるようなインターフェイスを作製した（図6、参考サイト8）。

また、著作権者からテキストデータを得ることができた2冊については、原

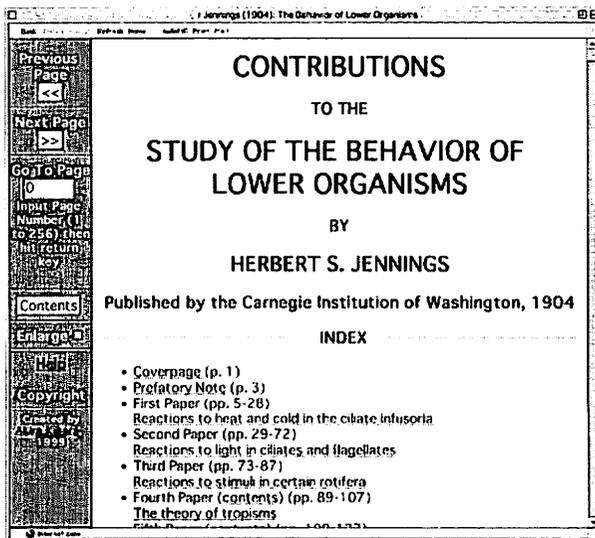


図6 古い文献画像のデータベース化

左の画面のボタンを利用することで、ページをめくったり、目的のページを表示することができる。

書中の写真・イラストをスキャナーで取り込み画像ファイル化し、テキスト部分をHTML化することで、できるだけ原著に忠実に画像をインライン表示させるように工夫した。

いずれも、研究者がボランティアとしてできる程度の作業量で、かなり高い品質のデジタル画像本を複製することができた。専門書は元々発行部数が少なく、かつ絶版に至る期間が早い。従ってデジタル化した場合においても商業化できるほどの頻繁なアクセスを期待できるものではない。このような学術刊行物を、多大な労力をかけることなくデジタル化し、末長くネット上で閲覧できるようにする試みは、今後のネットワークリソースの多様化のために必要であると考えられた。

考 察

本研究で一貫して問題としている課題は、第一線で活躍する研究者が個人で所蔵するデータをインターネット上に発信するために、どのようなサポートが必要であるかを明らかにすることである。我々は、既にいくつかのデータベースプロジェクトのコンサルティングを行う間に、既存の文献を通じては公開・入手できないカラー画像・動画等の公開・閲覧に対するの要望の強さを感じてきた。テキストデータについては、既存の文献データベース、また培養株の特性などを記したカタログ本がバイオリソースとして既に活用可能な状態にある。しかし、出版経費のかさむカラー画像については、分類情報の基盤となるデータであるにもかかわらず、なかなかネットワーク上に公開されていないのが本プロジェクト開始当時の現状であった。本プロジェクト期間中に、我々がコンサルティングを行ったり、またその結果できたサイトを参考にした画像データベースがいくつか出現したことは喜ばしいことであるが、未だ研究者におけるデータ公開が一般化していると言える状況に無い。研究者が時間を割いて何かを行うということは、当然そこにオリジナリティーを要求される。その意味においても、未だ基盤として整わない生物の形態画像データベースを主目的として、研究者のデータ発信をサポートするシステムを構築することは今後の生物系バイオリソースの充実を目指す上で意義の有ることと考えられる。

そこで、まず既存の標本がほとんど皆無である原生物の形態画像の或いは動画のアーカイブ手法から始まり、画像を中心としてデータベース作りにおい

て何が最も重要であるかを考察した。

(1) 研究者が自らデータを発信する為の環境作り

多くの研究者は、データベースを自ら作製した経験を持たない。また、そのような研究者に対して難しい仕組みをトレーニングさせる試みは、決して良い結果にはつながらない。事実、筆者らも「アリ DB」、「牧野 DB」をコンサルティングする過程で、必要以上に複雑なシステムを導入することは却って研究者のデータベースに対する興味をそぐ結果に繋がることを経験した。従って、できるだけ従来の研究スタイルを変えることなく、新しい高度情報化社会へのデータ提供を促進するようなデータベース構築方式を考察してきた。この点は、最先端の技術導入により企業イメージを確保しようとする商業レベルのデータベースと本研究で対象にしている学術データベースは初期の構築過程で方向性が全く異なることを示している。

画像データベースの為のデータ収集においては、従来の研究者が出版社に対して行ってきたことと同様に、35 mm スライドデータをデータ媒体とした方式を取り入れた。Photo CD システムを使った画像のデジタル化は多くの研究者の希望を満たすだけの品質を確保することが可能で、またその手続きも業者に発注できるという点で、十分簡素であった。また、そこから得られる5段階の解像度の画像はそれを表示する Web ページのレイアウトに柔軟性を持たせる事が可能であった。

一方では、研究機器のデジタル化に対応すべく、デジタルカメラを使った画像データ取得システムの研究も行ってきた。現時点で販売されているデジタル画像入力システムのほとんどは、顕微鏡撮影などの特殊な環境では必ずしも高い性能を発揮できるものではなく、高品位を目指す指向では今一つ実力にかけられる機器が多いことが判明した。しかし、コスト面や即時性という点で優れていることから、35 mm フィルム系との棲み分けがしばらく続くものと思われた。「原虫 DB」、「アリ DB」、「牧野 DB」で一貫して取り入れた、分類カスケードをサムネール画像のリストでたどる方式は、今後、他の分類群の形態画像データベースにおいても雛形となりうるものである。このような雛形とその解説をサポートシステムの部品として提示することで、データを発信しようとする研究者が具体的な完成イメージを持つことができる。その結果、より高いモチベーションで作業を行うことが可能になるものと考えられた。

初期のデータ収集過程では、その段階が逐一 Web を通じて公開できるような、HTML 形式でのデータ統合が有効である。次の段階として、データがある程度出そろった段階で、それらを構造化することでより高度なデータベース形態への発展が可能となる。これは、必ずしもデータを発信する研究者によって行われる必要は無い。何故ならば、一旦 Web で発信されたデータはそのサイトが有るかぎり、リンクとして他にも流用が可能だからである。しかし、一方では分類カスケードや分類キーによる絞り込みの手順はそれ自体が研究者のオリジナリティーを示す要素である。そのような、分類・検索手順を Web 上のリンクとして反映するためには、元データの構造化は是非とも行われなければならないステップで有る。「アリ DB」においては、初期の HTML 型のファクトデータベースを元に、データベースソフトを利用したリレーショナルデータベースへの構造化を図り、その時点でより高度なリンクを持つ Web を作り上げることができた (Kihara ら 1997)。その過程で行われた、「分類記載データ」「画像データ」「文献データ」「分類キーデータ」の独立化は、他の分類群においても利用可能な雛形の一つとして、今後サポートサイトにおいて提示することになる。

(2) デジタル画像データの可能性の提示

カラー画像を低コストで配信することができる点が、現在の Web の大きな特徴である。更には、既存のメディアでは配信することができなかったデータ形態を発信・閲覧することが可能となる。本研究では、新たな可能性の一例として動画データベースを元に、顕微鏡の 3 次元画像データ配信、デジタルタイムラプス動画の配信・閲覧手法について研究した。

原生生物の形態は動的で、活動的に動いている状態でのみ観察しうるケースが多い。例えば、ツリガネムシの柄のらせん形状は収縮した状態でのみ観察される形態で伸張時は消える (図 2 参照)。類似の形態の繊毛虫で柄が伸びた状態を保つ場合には別の種に分類される。したがって、この柄の伸縮性は分類学的にも重要なキーであるが、それがどのような状況下で、どの程度のスピードで収縮するかを知らなければ実際には認知することが難しい。このような動的形態の描写は、従来既に観察経験がある事を前提に文章で記載する手法しか無かったが、動画を共有データとして利用することで、未経験の相手をも対象に記載することができるようになる。

また、動画的な描写を用いることで顕微鏡観察時に焦点を変えながら立体的な形態を観察する手順と同様の動画画像を配信することも可能になった。更に、ゆっくりとした動きのために認識できなかった形態変化もタイムラプス動画を使って表現することができる。

このように従来メディアでは不可能であった画像（映像）データをマルチメディア型のデータベースに取り入れることは、今後の研究情報を飛躍的に豊かにするきっかけになり、その価値が認められればますますバイオリソースとしてのデータベース活用の促進につながるものと期待される。

(3) 学術データアーカイブのための共同作業

研究者の活動はたんに新しいデータを世に示すだけでは済まない。新しい知見は、同時代の研究者の手によって、レビューとして過去の業績とともに取りまとめられ知的基盤としての性格を深めていく。言わば、時代時代で、それぞれの視点に立った過去の業績のデータベース化が行われてきたといえる。自然科学的研究手法においては、新しい知見が常に過去の知見の上に成り立っていることが重要な意味を持つ。したがって、今後新しい情報システム上で知的基盤を形成するうえにおいて、同一システム上で過去の知見を並行して蓄積していくことが重要である。

本研究で行った過去の文献のデータベース化の試みは、研究者が保管している過去の貴重な文献を将来に渡って利用可能にする作業であり、従来の研究スタイルにおいてレビュー等を通じて印刷物の更新が行われる過程と共通した意味合いを持つ。特定分野の関連文献が現在でも入手が難しく、かつ商業ベースにおける復刻も困難であることから、今後研究者が必要に応じてそれらの文献をデジタル化する作業を行うことが現実的である。我々は、本方式に準じた活動を多数の研究者が行うことで、結果的にはインターネットを通じた共同作業となり、やがて巨大学術アーカイブが集積されることを提案している（図7）。

本研究の当面の目標は研究者が自らデータベースを構築・維持・管理するためのシステム開発である。しかし、いかに個々のシステムを開発しても、肝心の研究者にデータベースを構築・公開しようとする意欲がなければ役に立たない。そのような研究者の意欲を引き出すには、構築・公開されたデータベースが「研究業績」として認められ、公的機関により恒久的に保存される社会システム（知的基盤）の整備が必要である。

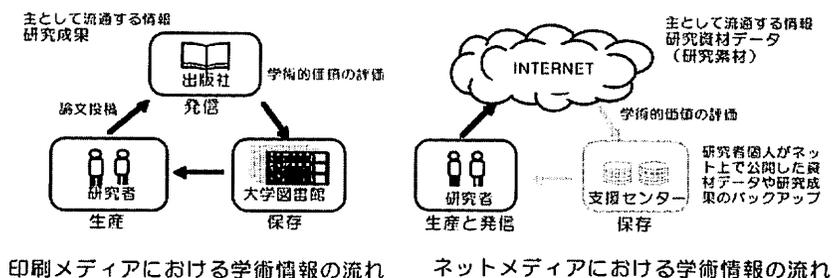


図7 公的サポートシステムの概略図

既存の印刷メディア（左）における「図書館」と同様の役割をするネットメディア（右）上のサポートシステムを提案している。

従来の科学では、研究者が作り出した情報は学術雑誌等の印刷メディアに記録され公開される。そして、公開されたものは大学図書館等で恒久的に保存される。この学術情報の恒久的な保存により、後になって論文等で引用することが可能となり、また、公開された情報が研究者の業績として認められるのである。しかし、現在、ネットワーク上でボランティア的に発信された情報については、印刷メディアにおける大学図書館のような恒久的な保存のための公的機関がない。このため、ネットワーク上にある情報は、論文等で引用することもできず、研究者の業績としても認められない。業績にならないならば、研究者に情報発信しようという意欲が生まれないのは当然といえる。

したがって、研究者によるネットワーク上での情報発信を活性化するには、その発信された情報を研究業績として認知する社会システムを確立しなければならない。そして、その基盤となるのが、発信された情報を評価し、学術的価値があると認められたものを恒久的に保存する公的機関の設置である。現在、そのような公的機関をどのように機能させればよいか、そして、そのための社会的なルール作り（引用文献の記述法、バックアップの方法、等）について検討中である。

参考サイト

1. 原生物情報サーバー
http://protist.i.hosei.ac.jp/Protist_menu.html
2. 日本産アリ類カラー画像データベース
<http://ant.i.hosei.ac.jp/Ant.WWW/INDEX.HTM>

3. 牧野標本館タイプ標本データベース
<http://ameba.i.hosei.ac.jp/BIDP/makino/>
4. アサガオ画像データベース
<http://protist.i.hosei.ac.jp/Asagao/Yoneda/menu.html>
5. Photo CD を Web に載せるまで
<http://taxa.soken.ac.jp/proceeding/OLWS/H2PhotoCD/index.html>
6. 古い文献の復活計画
<http://ant.i.hosei.ac.jp/WWW/Kihara/DataBase/DigitalBook/index.html>
7. CONTRIBUTIONS TO THE STUDY OF THE BEHAVIOR OF LOWER ORGANISMS
<http://ameba.i.hosei.ac.jp/DIB/Jenn04/>

《参考文献》

- 1) Ogata K, et al. (1999) Japanese ant color image database (1): an impact on taxonomic information, Proc. XIII Inter. Cong. IUSSI, p. 350.
- 2) Kihara A, Tsukii Y, Ugawa Y (1997) Making the regular structure of color image database of Japanese ants, Zool. Sci., 14: Suppl. 41.
- 3) 木原 章, 池田和夫 (1999) 電子顕微鏡画像データベースによるデータ処理の効率化, コンピュータサイエンス, 5, 101-109.
- 4) 月井雄二, 木原 章, 鶴川義弘 (1997) 原生物と日本産アリ類の広域画像データベース, コンピュータサイエンス, 4 (1): 73-74.
- 5) Ugawa Y., et al. (1999) Japanese ant color image database (2): structure and functions, Proc. XIII Inter. Cong. IUSSI, p. 482.

(細胞生理学/生物情報学・第一教養部専任講師)
(遺伝学/生物情報学・第一教養部教授)