

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： 植物多様性情報データベース

グループ代表者氏名： 伊藤 元己  
(東京大学大学院総合文化研究科)

## 1. 構築検討したデータベースについて

<p>データベースの概要及び特徴</p>	<p>本データベースは植物多様性に関する情報の中で基本となる2つの情報、植物種名データベースと植物標本データベースより構成される（GBIF 計画の2つの柱）。さらに両者を緊密に働かせることにより植物多様性情報のより有効な活用を目指すものである。</p> <p>植物種名データベースは日本植物分類学会で作成された Flora of Japan データベースを基に拡張を加えたもので、日本産植物の正名・異名を網羅したものである。本データベースの学名データは IOPI, Species2000 経由で GBIF の Catalog of Life に貢献する。</p> <p>植物標本データベースは日本の主要植物標本庫所蔵の植物標本をデータベース化するものである。他の類似データベースと比べ、標本の同定履歴、異名のサポート、詳細標本画像の添付などの点で先進性を有し、データのクオリティをあげるのにも貢献する。また、地理情報を緯度経度に変換するシステムを付属し、分布図などの作成が容易になる予定である。</p>
<p>類似のデータベース</p>	<p>国内：種名データベースに関しては本データベースの基になった Flora of Japan Database（日本分類学会作成）がある。本データベースはこの Flora of Japan Database のデータを基に拡張を加え、正名・異名とも同時に検索可能にした。他の種名データベースとして日本植物分類学会の絶滅植物専門委員会で作成中の植物グリーンリストがあるが、こちらは保全のための種名統一を目指したものであり、本データベースとは目的が異なり、また、単に種名を羅列したものである。</p> <p>標本データベースは東京大学、京都大学、都立大学でタイプ標本のデータベース化が行われているが、都立大学以外はまだ完成度が低く非公開である。本データベースのようにタイプ標本以外の標本を網羅した大規模なものは現存しない。現在、地域博物館の植物標本の現状について調査中であるが、その過程でいくつかの地域博物館では標本管理の目的でデータベース化が行われていることが明らかになった。これらは標本管理が目的のために非公開であるが、本データベースにデータを提供あるいは本データベースから検索が可能になるように交渉中である。</p> <p>国外：種名データベースとしては ITIS や Species2000 がある。しかしこれらは植物の種名に関しては北中米やヨーロッパ産の種が中心であり、アジア産の種名はほとんど含まれていない。現在作成中の種名データベースはこれらの既存種名データベースを相補するものになる。</p> <p>植物標本データベースはアメリカ、ヨーロッパ（イギリス、オランダ）、オーストラリアで進んでおり、大規模な植物標本館（庫）ではコレクションのデータベース化が終了あるいは進行中である。アジアは日本を含めデータベース化は遅れているが、台湾ではすでに中国科学院の主導で台湾内の標本庫の大規模なデータベース化を開始していることが判明した。これらのデータベースはおもに分布情報と標本管理を目的としたものであるため、タイプ標本を除き標本画像を持っていない。</p>
<p>生物分野</p>	<p>対象分野：分類学，生態学，遺伝学，資源生物学</p> <p>対象生物群：種子植物（被子植物・裸子植物），シダ植物 コケ植物，藻類（検討の結果，他の植物群と同一のフォーマットでデータベース化可能な海藻類のみを扱う）</p>

## 2. データ源およびデータの現在の状況について

(FSを踏まえて以下の項目について記述ください。)

<p>データ源</p>	<p>1) H13年度データ取得可能なデータ源(件数など):  植物名データ: 総計 9808 件 (Accepted: 3494 件, Synonym: 6314 件)  植物標本データ: 3,000 件 東京大学, 京都大学, 神戸大学, 広島大学, 服部植物研究所所蔵の植物標本.</p> <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):  植物標本データ: 国内の潜在的データ源は植物標本約 1,000 万点にのぼる. FSでの標本データのテスト入力を行った結果, 入力作業は 50 点/人・日と見積もられ, 入力人件費 160 万円当たり 10,000 点になる. 実際に可能な入力キャパシティは 20 万点/年と予想され(入力人件費 3200 万円), その場合の実行機関(入力点数)は東大(6 万点), 京大(6 万点), 神戸大(1 万点), 広島大(1 万点), 服部植物研究所(1 万点), 東京都立大学(2 万点), 東北大学(2 万点), 金沢大学(1 万点)を予定する.(入力人件費 480 万円の場合の予定は東大, 京大, 神戸大, 広島大, 服部植物研究所の所蔵標本 30,000 点)</p> <p>植物名データ: 総計 4,000 件(被子植物の未完成部および, コケ植物の学名データ)</p> <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):  植物標本データ: H15年度以降は H14年度の実行機関に加え, 各地の地域博物館の標本の入力も行う. 入力キャパシティは H14年度の 20 万点/年に加え, 地域博物館の 5 万点/年がプラスされる.</p> <p>植物名データ: 総計 8,000 件(コケ植物の学名データおよび海藻類の学名データ)</p>
<p>データの発生・収集場所</p>	<p>標本データ  自機関(東京大学): 被子植物 50,000 点, 裸子・シダ植物 10,000 点  他機関 京都大学: 被子植物 50,000 点, 裸子・シダ植物 10,000 点  神戸大学: 藻類(海藻) 10,000 点  広島大学: コケ植物 10,000 点  服部植物研究所: コケ植物 10,000 点</p> <p>植物学名データ  自機関(東京大学): 総計 2,000 件(被子植物の学名データ)  他機関(国立環境研究所): 総計 2,000 件(コケ植物の学名データ)</p>
<p>他機関の場合の連携協力について</p>	<p>他機関との連携  14年度に標本データを入力予定の他機関はすでに今年度のFSでテスト入力を開始している. また, すでに国内の主要植物標本庫の責任者とは打ち合わせ済みであり, 予算的裏付けがあれば, すぐに他の各標本庫においても標本データベース化ができる体制を整えている.</p> <p>学会等との連携  日本植物分類学会では植物データベース専門委員会(委員長: 伊藤元己)が作られており, 植物学名データベースはこの委員会が主体となって作成している. また, 標本データベースに関しても日本植物分類学会による全面的な協力体制が整っている. コケ植物に関しては日本蘚苔類学会の有志によりデータ収集が行われていて, データベース化の準備は整っている.</p>

<p>データフォーマット</p>	<p>植物種名データベース 文字情報</p> <p>植物標本データベース（詳細な構造は図を参照のこと） 文字情報：ラベル記載情報とその英訳（図の上の段） 種名同定履歴（図の左下） 画像情報：5M 画素標本写真（JPEG ファイル）</p>
<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） 植物名データ：総計 9808 件（Accepted:3494 件，Synonym: 6314 件，），保存媒体は CD-R と MO（約 10M）. 植物標本データ：3,000 件，東京大学，京都大学，神戸大学，広島大学，服部植物研究所所蔵の標本，保存媒体は CD-R と MO（約 1.5G）.</p> <p>2) 平成 14 年度の見込み 植物標本データ FS での標本データのテスト入力を行った結果，入力作業は 50 点 / 人・日と見積もられ，入力人件費 160 万円当たり 10,000 点に．実際に可能な入力キャパシティは 20 万点 / 年と予想され（入力人件費 3200 万円），その場合の実行機関（入力点数）は東大（6 万点），京大（6 万点），神戸大（1 万点），広島大（1 万点），服部植物研究所（1 万点），東京都立大学（2 万点），東北大学（2 万点），金沢大学（1 万点）を予定する．（入力人件費 480 万円の場合の予定は東大，京大，神戸大，広島大，服部植物研究所の所蔵標本 30,000 点）</p>

	<p>植物学名データ  自機関（東京大学）：総計 2,000 件（被子植物の学名データ）  他機関（服部植物研究所）：総計 2,000 件（コケ植物の学名データ）</p> <p>3）平成 15 年度以降の見込み  植物標本データ  H 14 年度の実行機関に加え，各地の地域博物館の標本の入力も行うため，入力キャパシティは H 14 年度の 20 万点 / 年に加え，地域博物館の 5 万点 / 年の合計 25 万点 / 年となる．</p> <p>植物名データ  総計 8,000 件（コケ植物の学名データおよび海藻類の学名データ）</p>
<p>データ・ベースの実現方式とデータのクオリティ</p>	<p>実現方法 実際運用するデータベースは LINUX サーバー上で PostgreSQL で構築し，PHP にて WEB 連携プログラムを作成，WEB サーバープログラム（Appach）にてインターネットに公開する．このシステムのプロトタイプはすでに作成済み．PostgreSQL からの出力は内部処理の段階で XML 化してあるので XML 文書での出力も可能になっている．</p> <p>データの入力 各標本庫でパーソナルコンピュータ（Windows, Mac）を使用して入力が容易に行えるように入力フォームをファイルメーカーで作成した．入力後のデータは Linux サーバーに移して運用する．</p> <p>クオリティチェック 日本植物分類学会，日本蘚苔類学会，日本藻類学会所属のそれぞれの植物群の専門家によるクオリティチェック体制を確立している．本データベースは各レコードに 500 万画素の詳細画像付加がデフォルトになっているので，実際の標本画像を見ながら同定が正しいかどうかのクオリティチェックが可能である．</p>

### 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>複数の同定意見，異名に対する対応          分類システムに非依存，同定意見の専門家間の相違に対応するため，Multi-nominal システムを採用した．具体的には標本の記述データと同定データを別ファイルに分け，リレーショナルデータベースとすることで1枚の標本を複数の名前で取り扱うことを可能にした．このようなシステムを実現している植物標本データベースはまだ少数であり，本システムの先進的な部分である．</p> <p>画像付加          モニター上で画像のラベルの情報を読むには500万画素以上の画像が必要であった．また，この解像度でなら，標本の細部の観察が可能であることが確認された．そのため，特別な理由がない限り，500万画素以上の画像を付加することにした．また，500万画素の画像のみではデータベースの作動が遅くなるので2段階に縮小した画像も合わせて付加する．この縮小画像は Graphic converter で自動作成する．現時点では解像度を選べるだけだが，将来的にはクリックابلマップを用いることにより，必要部分を拡大可能にする予定である．</p> <p>標本データ入力法          標本データを入力する際の効率的な方法を検討した．その結果，標本の画像をまとめて撮影し，モニター上で画像のラベルを確認しながらラベル上のデータを入力していく方法が最も効率的であった（50点 / 8時間）．また，標本画像撮影と入力の過程を分けることができるので標本庫外でのデータ入力作業が可能になる．手狭な標本庫外で入力作業ができるので実際の効率を上げることもつながる．</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>同定クオリティ          データベース化すると同定された名前が標本の実体を離れて流通してしまう．そのため同定クオリティは非常に重要である．本データベースでは500万画素の詳細画像を付加し，クオリティチェックを行うと共に，特に問題のある時にはユーザー自身が確認することを可能としている．</p> <p>絶滅危惧種のデータ公開          絶滅危惧種のデータは無制限に公開すると詳細な産地が知られることになり，植物の絶滅危惧種のデータ公開に関するガイドラインの設定（種の選択）を現在，日本植物分類学会の絶滅植物専門委員会に依頼しており，そのリストに従って，指定種については詳細な位置情報を非公開にするようにデータベースアクセスに制限を設ける．</p>
<p>その他</p>	<p>地理情報システム          本データベース利用者にとって標本の位置情報は重要であるが，既存の標本には地名の記述しかない．本格的な位置情報の利用には緯度経度情報の入力が必要不可欠であるが，1点1点入力するのは手間のかかる作業である．そこで地名辞書を作成し，地名から緯度経度情報を自動入力するシステムを考えた．また，そのための地名辞書フォーマットを標準化して，他のプロジェクトと共有して地名辞書作成の時間・コストを削減する．さらに緯度経度情報に信頼度を加えることにより，分布解析の信頼性を上げる仕組みを導入した．</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)

開発責任者	伊藤元己, 東京大学大学院総合文化研究科, 助教授
研究協力者	<p>データベース設計WG</p> <p>梶田 忠 東京大学大学院理学系研究科, 助手</p> <p>加藤英寿 東京都立大学理学研究科, 助手</p> <p>永益英敏 京都大学総合博物館, 助教授</p> <p>伊藤 希 筑波大学生物科学系, 講師</p> <p>清水英幸 国立環境研究所・国際共同研究官</p> <p>太田道人 富山市科学文化センター, 主任学芸員</p> <p>被子植物WG</p> <p>邑田 仁 東京大学大学院理学系研究科, 教授</p> <p>青木誠志郎 東京大学大学院総合文化研究科, 助手</p> <p>戸部 博 京都大学大学院理学研究科, 教授</p> <p>井上 健 信州大学理学部, 教授</p> <p>裸子植物・シダ植物WG</p> <p>鈴木三男 東北大学理学研究科, 教授</p> <p>村上哲明 京都大学大学院理学研究科, 助教授</p> <p>高宮正之 熊本大学理学部・助教授</p> <p>朝川毅守 千葉大学理学部・助手</p> <p>コケ植物WG</p> <p>岩月善之助, 財団法人服部植物研究所岡崎分室, 分室長</p> <p>長谷川二郎 南九州大学・教授</p> <p>山口富美夫 広島大学理学部・助教授</p> <p>樋口正信 国立科学博物館・主任研究官</p> <p>古木達郎, 千葉県立中央博物館・上席研究員</p>

藻類（海藻）WG

川井浩史 神戸大学内海域機能教育研究センター・教授

北山太樹 国立科学博物館植物研究部, 研究官

河地 正伸 国立環境研究所 生物圏環境部, 研究官

アドバイザー

岩槻邦男 放送大学, 教授

大場秀章 東京大学総合研究博物館, 教授

矢原徹一 九州大学理学研究院, 教授

若林三千男 東京都立大学理学研究科, 教授

加藤雅啓 東京大学大学院理学系研究科, 教授

出口博則 広島大学理学部, 教授

水谷正美 財団法人服部植物研究所, 元所長

千原光雄 千葉県立中央博物館, 館長



6 . FS 後の推進スケジュール（案）

（今後のスケジュール案と平成14年度見込みについて記述して下さい。）

（例）

	H 14	H 15	H 16	H 17
1.データベース基本設計	H13 終了			
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				