

F-6 アジアオセアニア地域における生物多様性の減少解決のための世界分類学イニシアティブに関する研究

(2) GTI地域プログラム実施における生物多様性情報共有化と利用に関する研究

② 生物多様性分布解析のための情報学的研究

筑波大学

学術情報メディアセンター 佐藤 聡

平成14～16年度合計予算額 13,109千円

(うち、平成16年度予算額 ※7,700千円)

「※上記の予算額には、間接経費1,777千円を含む」

[要旨] 本研究では、多様な生物種の標本やライブカルチャのバウチャ情報の電子データを基にした生物多様性分布解析を容易に行うために必要となる情報学的な技術開発を目標とした。これら生物多様性情報を容易に交換できる仕組みが分布解析には必要不可欠である。多様な生物種が生息するアジアオセアニア地域にて管理されている標本やライブカルチャの情報を対象とし、それら電子データ交換のための問題点を明確化し、情報技術に基づいた新たな手法やシステムを用い、こられの問題点を解決することを本研究の目的とする。

2年目である平成15年度は、世界規模のデータ交換の現状の取組みについて調査し、サービス実現に向けての技術的課題、運用継続性に関する課題等調査し、解決方法についての議論を行った。その結果、何らかの地域または組織単位で集中的にデータを管理し提供する方法がコスト的に有効であると判断した。その結果を踏まえ、アジアオセアニア地域を対象としたサーバ、データの整備を行い、データ提供サービスを実現した。現在、GBIF(Global Biodiversity Information Facility)へのデータ提供を行っている。また、OBIS(Ocean Biogeographic Information System)との連携に関する技術的検討を行った。また、集中的にデータを管理する方法について、運用上の問題点の検討を行った。その結果に基づき、1) データ提供者が集中管理サーバへのデータ提供作業が容易に行えるためのツール、2) データ提供者が提供したデータの利用法を制御するための検索インタフェース開発ツールの開発を行った。

また、10月にJoint International Forum on Biodiversity Informationにおいて、情報科学分野の研究者および分類学・生態学分野の研究者を集め、IT/CS community meetingを開催し、生物多様性情報を用いた研究課題についての意見交換を行い、今後の研究動向についての指針を得た。

3年目である平成16年度は、世界規模のデータ交換の現場からデータを採取し、利用するための方法について調査研究を行った。具体的には分布図への加工を行った。そのためにも、データの採取を行いデータの傾向分析を行い、地名から緯度経度情報へ変換する仕組みとの連携を行った。

[キーワード] 生物多様性情報、データベース、情報統合、分散問い合わせ

## 1. はじめに

急速な工業化、地球温暖化の影響により、人類を取り巻く生物圏の変動が過去になく急速に発生している。この状況は、極めて小規模な限られたモニタリングによって示唆されている。しかしながら、地球規模の現象である気候変動などの生物圏全体に及ぼす影響に対して、これら小規模な結果を外挿することは、生物生息種が地域によって著しくことなるため、極めて困難である。これらを克服するためには、地域間や既存の分類体系を越えた俯瞰的なデータの再解析が必要となる。例えば、生息生物種の分布は、数世紀にわたる膨大な採集標本のバウチャに蓄積されている。地球規模での生物種に関するモニタリングを実際実施するには、人的、時間的、資金的に困難であるため、過去の採集保存生物種情報を用いて環境変動や移入種（人為的に地域に生存しなかった生物種を導入すること）の影響をシュミレーションすることが重要である。そのために、様々な生物種に関するデータベースの統合が必要となる。

様々な生物種に関するデータベースの統合により、地球横断的な分析が可能となる。分析の結果、さらに調査が必要になる地域や生物種が発見され、それらを対象とした調査が行われることにより、新たな生物種情報に関するデータベース構築が進む。

地球には、多様な分類群、または多様な地域が存在する。それぞれ分類群、地域には、特性が存在するため、特定の地域、特定の分類群を対象としたデータベースには、それらの特性を生かすことのできる柔軟なデータスキーマが必要となる。また、生物種に関する情報は大変膨大であるために、地理的に分離された場所で独自にデータが管理されなければならない。データの管理が異なるために、それぞれの管理に適したデータスキーマが必要となる。

## 2. 研究目的

本研究では、生物多様性分布解析を容易に行うために必要となる情報学的な技術開発を目標とする。ここでは分布解析に用いるデータとして多様な生物種の標本やライブカルチャのバウチャ情報の電子データを対象とする。これらのデータを容易に交換できる仕組みが分布解析には必要不可欠である。多様な生物種が生息するアジアオセアニア地域にて管理されている標本やライブカルチャの情報を対象とし、それらの情報の電子データ交換のための問題点を明確化し、情報技術に基づいた新たな手法を用い、学名データベースや地名データベース等との連携し、分散管理されている生物多様性情報に対するインテリジェンスな検索、かつ、検索結果の高度な編集・加工を可能とする総合的なシステムの実現が最終目標である。

初年度である平成14年度は、関連する分野の研究者から問題を聞き取り調査することからはじめ、標本やライブカルチャを管理する側が容易に情報提供できるような枠組みおよび発信された情報の柔軟に活用する枠組みの実現を目指した。

2年目である平成15年度は、世界規模のデータ交換の現状の取組みについて調査し、サービス実現に向けての技術的課題、運用継続性に関する課題等調査し、実際にアジアオセアニア地域を対象とした対象としたサーバ、データの整備を行い、データ提供サービスの実現を目指した。また、集中的にデータを管理する方法について、運用上の問題点の検討を行い、それらの解決のために必要となるツール群の設計、実現を目指した。

### 3. 研究方法とその結果

平成14年度は以下の項目について研究を行った。

#### (1) XML形式による情報発信手法

標本・ライブラリカルチャのバウチャ情報のデータベース化は、標本やライブラリカルチャ自身の管理の手間を軽減化するために行われている。このような場合、バウチャに記載されている情報をそのまま電子化されている場合がほとんどであり、それを統一的な記載に変更することは難しい。たとえば、学名においても、バウチャに記載されている学名がデータベースに格納されているため、シノニムが格納される可能性は低い。これ以外にも、採取地や採取時刻に関する情報も統一的な記載方法で電子化を望むことは非常に難しい。そこで本研究では、それら記載方法の違いを解消するシステムを外部に取り込む方式について開発した。そして、分散管理されているバウチャ情報に関して、基本的なアクセス手法を設計することにより、容易に元のデータにアクセス可能とした。具体的には、Microsoft Accessにより管理されている標本情報を対象として、学名、または標本のID情報のいずれかにより標本情報を検索でき、その結果をXML形式にて出力するシステムを実装し、検索実験を行った。リレーショナルベースシステムにおいて、公開したいテーブル名とそのテーブルの中で公開したいフィールド名を指定するだけで、バウチャ情報を公開できることを確認した。これにより、バウチャ情報を取り扱う応用分野（たとえば、バウチャ情報を用いて分布解析を行う）のアプリケーション作成が容易に行えるようになる。

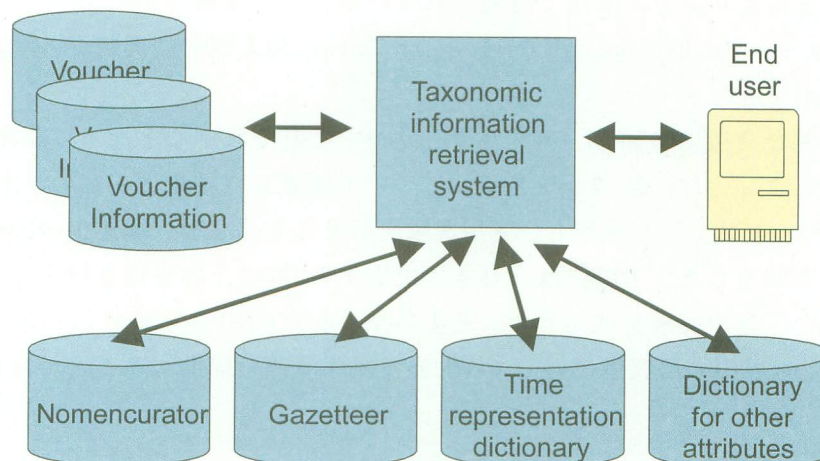


図 1 全体システムの概要図

## (2) XML形式のデータ交換のためのフレームワーク

図1に示したように、提案するシステムは様々なデータベースを接続することにより実現できる。利用者が要求するデータを問い合わせして入力し、問い合わせに用いて単語を必要に応じて変換し、バウチャ情報を発信する側が理解する問い合わせの形式に変換し、問い合わせを発行する。そして、検索結果を収集、統合し、必要に応じて記載されている項目について統一的な記述に変換する作業を施したり、または、利用者が望む形式に変換する。

これら一連の作業をXML形式の言語で記述する方法の提案を行い、さらにその言語で記述されたスクリプトを解釈実行するシステムの試作を行った。

これにより、検索した結果を一般的なWebブラウザで表示可能なHTML形式に変換して閲覧したり、採取地点情報（緯度経度情報）を基に地図上にプロットしたグラフィックス形式（SVGなど）に変換して閲覧することが可能となる。

平成15年度は以下の項目について研究を行った。

### (1) 世界規模のデータ交換手法に関する調査

世界規模のデータ交換の現状の取組みについて調査するために、次の2つの会議に参加した。

- ① 平成14年10月につくばで開催されたGBIF(Global Biodiversity Information Facility) GB7のGBIF Training Course : Becoming a GBIF Data Provider (<http://www.gbif.org/Events/train-2003-10-tsukuba>)。この会議に参加することにより、主にアメリカで開発されているデータ交換フォーマットであるDarwinCore、そのフォーマットを用いたデータ交換プロトコル、DiGIR(Distributed Generic Information Retrieval) およびそのプロトコルの実装システムについて調査することができた。
- ② 平成14年10月にポルトガルにて開催されたTDWG(Taxonomic Database Working Group) Annual Meeting、特にその中のBioCASE Provider training workshop。この会議に参加することにより、主に欧州で開発されているデータ交換フォーマット・プロトコル ABCD(Access to Biological Collection Data) およびその実装システムについて調査することができた。

2つの国際的な会議に参加し、アメリカと欧州がそれぞれ中心となって開発を行っているフォーマットおよびそれに基づくシステムについて調査した結果、現在の日本においては、DiGIRに基づくシステムを実現し、運用実験を行うことにより、運用上の問題点等の調査研究が必要との判断をした。実際には、日本に由来する2群の生物多様性情報をシステムに投入する作業を行い、運用実験を開始した。その結果、現在GBIFからそれらのデータは検索可能となっており、世界中の研究者に対して、生物多様性情報を提供することが可能となっている。

### (2) データ提供サーバの運用コストを低くするためのツール群の設計

生物多様性情報の解析を行うためには大量の生物多様性情報がネットワーク上で利用可能とならなければならない。そのためには、電子化される生物多様性情報の量を多くする必要がある。情報の量を増やす際に大きな障害となることとして、データ提供サーバの維持

管理のコストがかかるという点がある。電子化された生物多様性情報を所有している研究者からの聞き取り調査の結果、それらデータ提供サーバの維持管理のためのコストを少しでも軽減できるツール群が必要であるとの結論に至った。データ提供サーバの運用コストを低くするために必要となるツールの設計を行った。設計に際して、実際のデータ提供者のデータ開発状況の調査を行った。実際のデータ提供者はサーバ管理には不慣れではあるが、一部の提供者はパーソナルコンピュータの扱いには慣れていることが調査により判明した。そのことから、集中的に管理しているデータへはパーソナルコンピュータの多くで利用されているマイクロソフト社の表計算ソフトであるExcelの形式のファイルとして投入する方法を採用し、Webブラウザからそれらのファイルをアップロードすることで、提供者が入力したデータを一括してデータベースへ格納する仕組みを開発した。また、実際のデータ提供者は、データがどのような形で利用されているか、特にどのように検索されているかについて、データ提供者が非常に興味を持っていることがわかった。特に、データの検索方法については、非常に興味があることがわかった。しかしながら、サーバのためのプログラム作成を行うのは非常に難しいと思われるため、容易に検索インタフェースをカスタマイズできる仕組みを開発した。

### (3) 生物多様性情報の解析手法に関する調査

10月にJoint International Forum on Biodiversity Informationにおいて、情報科学分野の研究者および分類学・生態学分野の研究者を集め、IT/CS community meetingを開催し、生物多様性情報を用いた研究課題についての意見交換を行い、今後の研究動向についての指針を得た。また、生物多様性情報から抽出できる情報として生物分類体系があるが、抽出された複数の分類体系の相違点を判別する仕組みについて研究を行い、その試作システムを作成した。大量のデータを利用しての解析に関しては、今後、情報学の分野、生物分類学の分野、生態学の分野の知識を融合しなければできないことがわかった。それらのひとつとして、生物多様性情報から抽出できる生物分類体系を対象として、抽出された複数の分類体系の相違点を判別する仕組みについて研究を行い、これはラフ集合の概念を分類体系に適用し、ラフ集合の推論機構を用いて、分類体系のあいまい性を判別することを可能とした。その試作システムを作成し、IEEE Symposium on Information Visualization において開催されたInfoVis 2003 Contest - Visualization and Pair Wise Comparison of Trees - に出展し、2<sup>nd</sup> placeを獲得した。

平成16年度については以下の研究を行った。

#### (1) 分布図作成システムの開発

平成15年度の研究結果により、生物多様性情報のデータ発信手法について指針が得られたため、平成16年には、DiGIRによる検索結果より分布図を作成する手法についての研究開発を行った。2005年1月25日現在、GBIFに登録されているデータプロバイダ中、DiGIRにより取得可能な生物多様性情報から地理属性としてJapanと記載されている生物多様性データを収集したところ、8560件であった。この8560件のデータについて、地理情報に関連する四項目のデータの有無を調査した結果を、以下に示す。

パターン	Longitude	Latitude	StateProvince	Locality	件数
1	データ無し	データ無し	データ無し	データ無し	645件
2	データ無し	データ無し	データ無し	データ有り	1240件
3	データ無し	データ無し	データ有り	データ無し	102件
4	データ無し	データ無し	データ有り	データ有り	5885件
5	データ有り	データ有り	データ無し	データ無し	0件
6	データ有り	データ有り	データ無し	データ有り	96件
7	データ有り	データ有り	データ有り	データ無し	0件
8	データ有り	データ有り	データ有り	データ有り	592件

この結果より、地名等の地理属性から、緯度経度に変換する手法が大変重要であることがわかる。ただし、地名等の地理属性の記載にはあいまい性が多く、緯度経度属性の精度との対比が問題となる。そこで、本研究では地名情報のあいまい性の度合いという意味のレベルという概念を持ち込み、どのレベルまでの精度で緯度経度情報に変換できるかの仕組みを導入しその結果に応じて地図に反映させる方法のプロトタイプを実装した。このプロトタイプでは精度の低いものは円の半径を大きくしかつ色の透明度を高くした。また、精度の高いものは、円の半径を小さくしかつ色の透明度を低くした。変換した分布図を以下に示す。

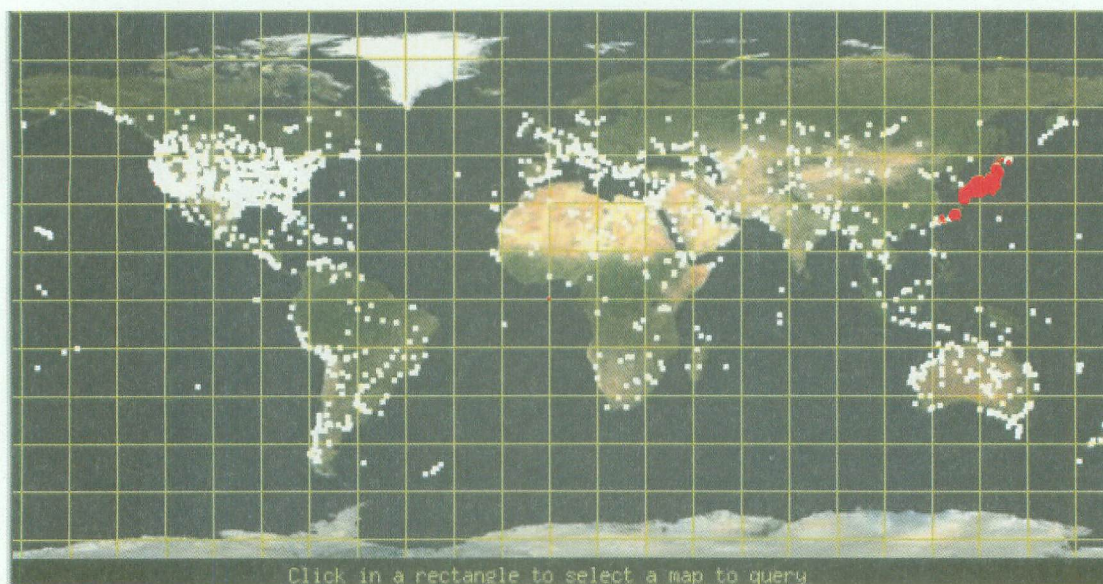


図2 あいまいな地理情報からの生物種分布表示例

#### 4. 考察

今回のプロジェクトを通して、生物多様性情報データの発信についてはさまざまな課題があることがわかった。例えば、様々な粒度の組織単位でのデータ提供サーバを構築することが可能ではあるが、分散することにより、小さい組織単位で提供できるデータの精度の制御や情報の修正頻度を高めることができる反面、各組織毎にサーバの維持運用を行う必要がある。昨今では、コンピュータウィルスの拡散によりサーバが運用停止になるなどの被害が頻繁におこっており、それらに対応するためには、さらにコストが必要になる。それらを考慮すると、国家レベルでの実装において、データ提供サーバの管理単位を細かく分離すること

により総コストが高ってしまう。現状では、ある程度の単位で集中的に管理する方式がコストの面から考えて実現可能性が高いといえる。その際に問題点となる更新頻度の高める方法としては、データ提供者自身が遠隔地よりデータをサーバに投入できる仕組みを用意することで解決できると考えられる。

また、アメリカ、欧州中心のデータフォーマットでは、非アルファベット言語の支援がなく、アジア・オセアニア地域ですでに入力されているデータにおいては対応不可能な問題も発生してくる。具体的には、採取地名や採取者氏名などは非アルファベットである現地の言語で記載されており、現在、それらを英語に翻訳できないなどの問題点もある。したがって、非アルファベット言語をそのまま提供できる枠組みをアジア・オセアニア地域から提案する必要がある。

さらには実際に生物多様性情報を活用して研究するシーンが曖昧である点にある。分布図の利用を想定し、研究開発してきたが、実際の科学研究のデータとして使える分布図を作成するには、もとのデータに記載している情報だけでは不十分であり、また推論するためには、相当数の背景知識を投入しなければならない。それは今後の課題としたい。

## 5. 本研究により得られた成果

これまで生物多様性情報の整備の必要性は主に情報を利用する側の要求にて議論が進められてきた。しかしながら、情報整備は、情報を提供する側の善意や義務感に頼って進められてきた。本研究では、生物多様性情報提供に関する世界動向を調査し、実際のシステムの実現を行い運用に関する調査研究を行った。管理対象の粒度を細かくしすぎる際の利点と欠点を把握することができ、アジアオセアニア地域においてデータ提供環境を整備するためには、管理対象の粒度をある程度大きくしたほうが実現が容易にできるという指針を得た。また、データの更新頻度を高めることを可能とするツール、および、データの検索インタフェースをカスタマイズ可能とするツールを開発することにより、集中管理する際のいくつかの問題点を解決する方法を提案することができた。

## 6. 引用文献

Distributed Generic Information Retrieval (DiGIR) <http://digir.sourceforge.net/>  
Access to Biological Collection Data <http://www.bgbm.org/tdwg/CODATA/Schema/default.htm>

## 7. 国際共同研究等の状況

The Natural History Museum のDavid McL. Roberts博士、The Open UniversityのDavid R. Morse博士、および筑波大学生物科学系の伊藤希博士と、分類概念を取り扱うためのデータ構造およびそのデータ構造を使った概念比較に関する共同研究を行っている。

## 8. 研究成果の発表状況

### (1) 誌上発表 (学術誌・書籍)

<学術誌 (査読あり)>

なし

<学術誌 (査読なし)>

- ①. 相良 毅、松浦啓一、佐藤 聡、志村純子：日本データベース学会Letters、1, 1, 39-42(2002)  
「曖昧な地名照合手法を用いた生物種標本の地図ブラウザ構築」

<書籍>

なし

<報告書類等>

なし

### (2) 口頭発表

- ①. 佐藤 聡、伊藤 希、小野 哲、柁原 宏、志村純子：情報処理学会研究報告  
2002-DBS-128,255-262(2002)「生物種情報データベースのためのデータ交換方式」
- ②. Sato A., Kajihara H., Sagara T., Ono S., Shimura, J. 1<sup>st</sup> GTI Regional Workshop in Asia,  
Putrajaya/Malaysia (2002), “Implementation Method for Data Exchange of Biological Diversity  
Information Database”
- ③. 佐藤義弘、佐藤 聡、板野肯三：第14回データ工学ワークショップ(DEWS2003)論文集  
<http://www.ieice.org/iss/de/DEWS/proc/2003/> ISSN 1347-4413(2002)「分散XML言語処理系の  
設計」
- ④. Nozomi Ytow, David R. Morse, Akira Sato, and David McL. Roberts: TaxoNote Comparator: An  
Application of Formal Concept Analysis with Rough Set Approximation to Multiple Hierarchies  
Visualization, Joint International Forum on Biodiversity Information Building Capacity in  
Biodiversity Information Sharring 2003.
- ⑤. 佐藤聡、伊藤希、David R. Morse, David McL Roberts 「ラフ集合が示す木構造と階層構造の  
違い」、日本知能情報ファジィ学会 ラフ集合研究部会 第2回ワークショップ (2003)

### (3) 出願特許

なし

### (4) 受賞等

- ①. IEEE Symposium on Information Visualization において開催されたInfoVis 2003 Contest -  
Visualization and Pair Wise Comparison of Trees - に以下のプログラムを出展し、2<sup>nd</sup> placeを  
獲得した (<http://www.cs.umd.edu/hcil/iv03contest/results.html>) 。
  - Comparison of Multiple Taxonomic Hierarchies Using TaxoNote: David R. Morse, Nozomi  
Ytow, David McL. Roberts, Akira Sato

### (5) 一般への公表・報道等

なし

## 9. 成果の政策的な寄与・貢献について

アジアオセアニア地域を対象としたサーバ、データの整備を行い、生物多様性データ提供サ



ービスを実現した。現在、GBIF(Global Biodiversity Information Facility)へのデータ提供を行っている。また、OBIS(Ocean Biogeographic Information System)との連携に関する技術的検討を行った。来年度はOBISへのデータ提供も開始する予定である。

また、作成したツール群については今後Web上で公開する予定である。