

生物種情報データベースのためのデータ交換方式

佐藤 聡[†] 伊藤 希[†] 小野 哲^{††}
舵原 宏^{†††} 志村 純子^{†††}

地球規模の現象が生物圏全体に及ぼす影響を調べるために、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情報を統合する枠組みが必要となる。これら生物種情報への検索には学名が用いられる。この学名は新種発見などにより常に変動している。即ち、求められる枠組みは変化するドメインをキーとする検索機能がなければならない。本論文では、標本のバウチャを管理しているサイトの特性を生かし、かつ、それらの情報を学名を用いて検索し統合するためのシステムについて述べる。

Data Exchange Method for Biological Diversity Information Databases

AKIRA SATO,[†] NOZOMI YTOW,[†] SATOSHI ONO,^{††}
HIROSHI KAJIHARA^{†††} and JUNKO SHIMURA^{†††}

For an investigation of the animate nature influenced from the phenomenon of the earth scale, it is important to integrate the biological information which was accumulated between the some recent centuries. The biological information is accessed by scientific name. The scientific name is dynamically changing because new species is discovered. We must develop a access method using the key which domain is dynamically change. This paper discuss a system which allow the individuality of each database of biological information, integrate voucher information and search biological information by scientific name.

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在しなかった生物種を導入すること）の影響をシミュ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¹⁾。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。

これらの状況より、地球規模のデータ統合を行うためには、すべてに共通なスキーマを与えるのではなく、データ交換の枠組みが必要となる。

同一地域内での統合を行う際には、さらに、地域内での特性に応じたデータ交換を可能とするために、地名辞典の標準化、言語・慣用語の用いる使用言語コードの地域内の標準化を支援するシステムも必要となる。

一方、人類は、生物を分別し、その分類ごとに名前

[†] 筑波大学

University of Tsukuba

^{††} (株) ランス

LANCE INC.

^{†††} 国立環境研究所

National Institute for Environmental Studies

(学名)をつけ、自然界を理解してきた。人類の長い歴史において、様々な分類が行われ、また、新しい生物の発見に伴い、分類の見直しが行われ、その結果、学名は常に変化してきた。今後も地球規模のデータ解析の結果、新しい生物の発見などにより分類の見直しが行われ、学名も変化し続ける。しかし、人類は、その学名により生物を認識しており、様々な生物種情報はその学名を用いて管理されているといえる。即ち、生物種情報において、学名のドメインは常に変化し続けることとなり、その変化するドメインをキーとした情報検索の枠組みが必要になる。

我々は、世界規模の生物分類情報へのアクセス方法を実現するべく研究を行っている。これを実現するためには、2つの大きな課題がある。

(1) 既存のデータの統合

現在、様々な機関、または研究者個人がデータスキーマを独自に定めている。対象とする生物種毎に特化されたデータスキーマが必要とされている。様々なデータスキーマにしたがってすでに大量の情報がデータベースシステムに格納されている。

(2) これから生成されるデータの統合

現在、新種の発見に向けて、様々な調査が行われている。また、未調査の領域が世界中にはまだあり、今後調査される予定である。したがって、今後生成される情報を電子化するためには、生物分類学者にとって負荷の少ない入力支援システムが必要となる。

我々は、今後生成されるデータは既存のデータの一部として考えることができると考えている。したがって、この両者は独立して研究を進めることができる。本論文では、前者の課題を対象としている。特に、生物種情報として標本のバウチャ情報を対象として、既存の標本情報に関するデータを統合し、様々な分析を可能とするための基盤となる枠組みについて議論する。

2. 問題点の解析

現在、生物分類学者が保有している情報としては、バウチャ(voucher)またはカタログ(catalog)の情報がある。ここで、バウチャとは、標本のラベルに記載されている情報や標本台帳に記載されている情報などを指す。また、カタログとは、ライプカルチャに関して記載されている情報を指す。一般に、生物分類学では、研究対象を標本にするか、ライプカルチャとして生きたまま保存するかのいずれかの方法を用いて、研究対象の生物を保存している。いずれの方法をとるかは、対象となる生物種によって定まる。以後、本論文では、バウチャ情報とよぶ場合には、カタログ情報も含むものとし、標本とよぶ場合には、ライプカルチャを含むものとする。

生物分類学者は、これら保存された標本を観察して分類を定めている。たとえば、採取された生物種の同定作業は、保存している標本との比較が必要となる。その同定作業により、既存の生物種にないと判断された場合には新たな生物種として記載論文を記載する。また、同定の作業により分類が困難であると判断されると、すでにある生物種に関して再分類が行われる。この様にして学名が定められる²⁾。

この学名は、生物科学分野だけでなく利用されるものではなく、他の科学分野、さらには科学以外の一般的な分野においても利用される。例えば、これらの学名を検索キーとするようなデータベースとしては、動物植物図鑑などが良い例である。この学名を利用して、異分野間での概念の共有化が行われるため、学名を定めることは大変重要である。しかし、現状では、まだすべての種が発見されておらず、すべての生物について学名がつけられているわけではない。したがって、学名をつける作業を支援するシステムは大変重要である。

生物の分類作業において、保存されている標本を観察する前に必要となる情報としてバウチャ情報は大変重要である。このバウチャ情報を有効的に利用するために、すでにいくつかの機関、研究者個人がバウチャ情報のためのデータスキーマを定め、バウチャ情報の電子化作業を進めている。これらの情報が電子化される理由は、研究者自身の分類学研究を行う際の標本へのアクセスの効率化のためである。したがって、これらのデータスキーマは、研究者対象の生物種の特性を活かした固有のデータスキーマである。

また、保存される標本の数は膨大となるため、標本自身が地域毎に分散されて管理、保存されている。したがって、電子化されたバウチャ情報も分散的に管理され、それぞれの地域の特性を活かした管理方法が行われている。

電子化されたバウチャ情報は、標本の管理識別子(以後、標本IDとよぶ)、または、標本に同定されている学名にてアクセスされる。前述のとおり、学名は常に変化しているものである。さらに、分類体系は様々ある。電子化されたバウチャ情報には、研究者にとってもっとも含意できる体系に基づいた学名が格納されている。また、標本IDは、各組織や研究者個人ごとに一意になるようにつけられていると考えられるが、それらが世界中で一意に定まるものではない。

電子化されたバウチャ情報には、上記の標本ID、学名の属性以外の属性としては、例えば標本が採取された場所や、採取された日時などが含まれていると考えられる。しかし、どのような属性が含まれているかは、生物種毎に異なる。また、同じ意味を持つ属性が含まれていたとしても、そのデータ型が生物種毎に異なっている。

したがって、世界規模において、バウチャ情報のた

めのグローバルデータスキーマを作ることは大変困難である。また、利用者が入力した値とデータベースに格納されている値は同じ概念を指し示しているにもかかわらず、異なっている可能性がある。これは特に学名に関しては大きな問題となるが、学名以外でも、生物が採取された場所や時刻などに関しても同様な問題が存在する。また、バウチャ情報の電子化の方法毎で、検索方法が異なっているといった問題も存在する。

以上の議論よりすでに存在する生物種情報のデータを統合する際の問題点をまとめる。

- 対象とする情報を表す値（コンテキスト）が標準化されていない。これは属性値のドメインが定まっていないことを表す。例えば、ドメインが検索対象者、データ格納者間で異なっている。特に、学名に関しては、分類学の研究に伴い常に変化しているものと考えられる。
- 対象とする情報が様々な機関または研究者により管理されているため、検索方法に一様性がない。

3. 提案するシステム

本節では、前節で述べた問題点を解決し、すでに電子化されているバウチャ情報を統合して様々な分析に用いることができるデータを提供するためのシステムの設計の概要について述べる。

図1に提案するシステムの概要を示す。本システムは、既存の電子化されたバウチャ情報群から利用者の要求に応じて必要な情報を収集することを支援するシステムと位置つけることができる。本システムでは、既存の電子化されたバウチャ情報としては、各機関または研究者が定めたデータスキーマに従って格納されている情報を対象としている。ただし、ここでいうデータスキーマとは、広い意味のスキーマと考えており、データベース管理システムを用いて管理している情報だけでなく、ファイルを用いて管理している情報も対象とする。また、これらの情報は分散的に管理されているものとする。以下、バウチャ情報を提供する母体のことをデータ提供サイトとよぶ。

本システムの特徴は、データ提供サイトと利用者間のドメインの違いを変換する機能をサブシステムとして分離し、これらのシステムを自由に組み合わせることを可能としている点にある。以下、ドメインの違いを変換するサブシステムのことを、翻訳システムとよぶ。例えば、異なる学名記述間の関係を解明するシステム（Nomenvurator）や地名と緯経度間関係を解明するシステム（Gazetteer）、異なる時刻表記間の関係を解明するシステムなどを翻訳システムとして本システム組み合わせることができる。

図2に本システムを用いた検索の流れを示す。利用者は、本システムのインタフェースに対し、必要となる情報の入力を行う。入力するキーワードは、標準 ID

表1 本システムにおいて用いる必須のデータスキーマ

属性名	意味	ドメイン
id	標準 ID	テキスト（文字列）
name	学名	テキスト（文字列）

または、学名としている。本システムは利用者の入力したキーワードから、必要としている情報が、本システム自身がつメタ情報からどのサイトに存在するかを判断する。そして、そのサイトに問合せを発行する前に、翻訳システムを用いてキーワードをデータ提供サイトのドメインに適応するように変換する。各サイトからの問合せ結果の内容に関して、同様に利用者のドメインに適用するように変換する。そして、検索結果を統合し、利用者の要求に応じた形式に変換して利用者に返す。

3.1 本システムでの基本概念

本システムの目的は、様々な機関または研究者個人が管理しているバウチャ情報を統合し様々な分析が可能となることである。より多くのデータ提供サイトの情報を統合し、分析可能とするために、我々は本システムでの設計における基本概念を以下のように定めた。

- データ提供サイトへのアクセスに用いる際に用いるスキーマはいかなるバウチャ情報に必ず含まれる属性のみを定義する。それ以外の属性に関しては、各データ提供サイトが自由に定義した属性を組み込むことを可能とする。また、そちら自由に定義した属性の属性値を閲覧できる API を定義する。
- データ提供サイトとの API は各データ提供サイトが実現可能な最低限のものだけを定義する。
- 翻訳システムが提供する API は特定のドメインから特定のドメインへの変換を行うものを定義する。

本システムでは最低限必要となるデータスキーマのみを定めた。それらを必須のスキーマとよぶ。必須のデータスキーマを表1に示す。電子化されているバウチャ情報には学名と標準 ID が含まれており、それらはテキスト（文字列型）で記載されていると考えられる。したがって、必須のデータスキーマは、学名、標準 ID のみとし、それらのドメインは文字列とする。

各データ提供サイトが実装すべき最小の API を定めた。それらを必須の API とよぶ。これらの API は、必須のデータスキーマを用いてデータ交換を行う。必須のデータスキーマは API の引数、および返値に用いられる。必須の API 表2に示す。なお、表中での item とは、必須のデータスキーマである name と id を含むバウチャ情報を表す属性値の組を表す。

SearchIdsByName バウチャ情報には何らかの分類体系に従った学名が記載されている。したがって、この API の実引数の値である文字列と記載されている学名との照合（例えば、単純なパター

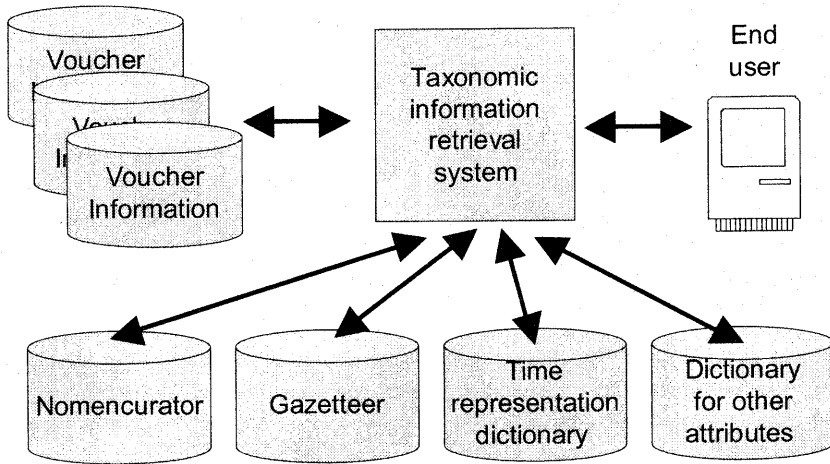


図1 本システムの概念図

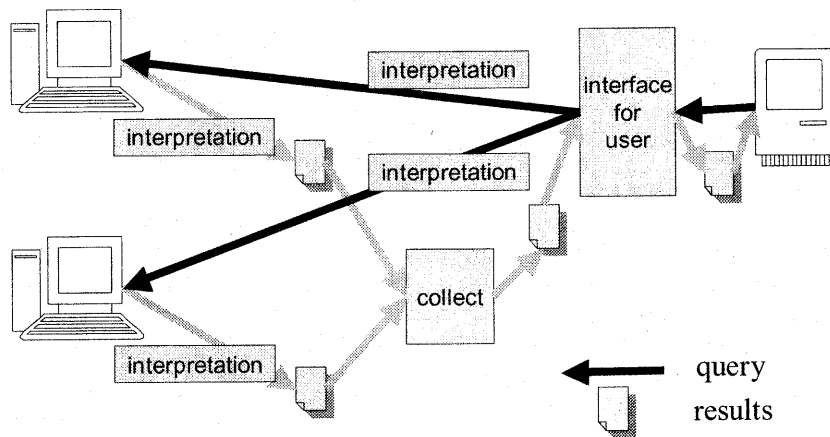


図2 本システムの概念図

ンマッチなど)は容易に実装できる。

SearchItemById バウチャ情報にはその情報の一意性を保障するために、なんらかのIDがつけられている。したがって、このAPIの実引数の値を用いて、利用者が要求しているバウチャ情報を一意に定めることは容易に実装できる(インタフェースが異なるが同様の機能がすでに実装されている場合が多い)。

SearchItemsByName

これは **SearchIdsByName** と **SearchItemById** を組み合わせることにより実現可能である。

生物分類学的には、標本の採取位置、採取時刻(日時)の情報は大変重要な情報であり、多くのバウチャにはこれの情報が記載されている。従って、これらの情報を交換する枠組みも重要と考える。しかし、これ

表2 本システムにおいて用いる必須のAPI

API名	意味
SearchIdsByName	学名(name)を引数とし、その学名に一致する標本の標本ID(id)のリスト(ids)を返す。
SearchItemById	標本ID(id)を引数とし、その標本IDが指し示すバウチャ情報(item)のリストを返す。
SearchItemsByName	学名(name)を引数とし、その学名に一致する標本のバウチャ情報(item)のリストを返す。

らの情報はすべてのバウチャ情報に含まれる属性ではない。また、位置、時刻の表現方法は様々な方法が存在する。現在、位置の表現方法としてはISO6709という標準が存在し、また、時刻の表現方法としては、ISO8601という標準が存在する。本システムの内部で

表 3 本システムにおいて用いる推奨のデータスキーマ

属性名	意味	ドメイン
location	採取位置	テキスト (文字列): ISO6709
date	採取時刻	テキスト (文字列): ISO8601

表 4 本システムにおいて用いる必須の API

API 名	意味
SearchSelectedItemsByName	学名 (name) を引数とし、その学名に一致する標本のバウチャ情報のうち、標本 ID (id) と学名 (name)、採取位置 (location)、採取時刻 (date) から構成される組 (selectedItem) のリストを返す。

の表現および、検索結果の表現としては、これらの標準を用いる。データ提供サイトは、これらの標準を含む様々な表現方法群中から、何らかの規則に従っている表現方法を選択し、位置・時刻の情報を表現する。規則が明記されていれば、標準に変換する機能を実現することは可能であると考えられる。従って、データ提供サイトは位置・時刻に関する情報を提供可能な場合は、その表現の形式をメタ情報として公開する。表 3 に内部で用いる位置・時刻のデータスキーマを定める。データ提供サイトと本システム間のデータ交換については、これと同じデータスキーマが用いられるが、厳密にはドメインが異なる点に注意する。これら位置・時刻のデータスキーマを本システムでは、推奨のデータスキーマとよぶ。また、この推奨スキーマを用いた問合せの API を表 4 に示す。ここで、selectedItem とは、id、name を必ず含み、バウチャ情報に採取位置または採取時刻が含まれればそれを含むバウチャ情報の属性値の組のことを表す。

この API は SearchItemsByName と同様の検索結果から必要な部分だけを射影することである。したがって、各データ提供サイトが格納しているバウチャ情報の属性に採取位置、採取時刻が含まれている場合は、容易に実現できる。

3.2 データ提供サイトに求める要件

データ提供サイトは API の実装を行う。必須の API のうちはじめの 2 つ API の実装に関しては以下の問題が存在する。

SearchIdsByName この API の実装では、与えられた学名と検索結果である標本 ID が指すバウチャ情報の学名が意味的に同一であるかどうかを保障する必要はない。利用者はこの API を用いた検索の結果により得られた標本 ID を用いて、その標本 ID が実際にほしい情報であるかどうかを

SearchItemById を用いて調べることが可能である。但し、標本 ID についてはデータ提供サイト毎に一意性が保障されることが必要となる。

SearchItemById 利用者が、データ提供サイトによって管理されているバウチャ情報を自由に閲覧可能とするために、この API による学名、標本 ID 以外の属性の利用者への返値の方法は属性の属性名とその属性の属性値とを列挙する方法とする。データ提供サイトは何らかのデータスキーマに従ってバウチャ情報を管理しているため、この方法は容易に実現できる。データ提供サイトの特性を許すために、他のデータ提供サイトからの検索結果との統合の際に、検索結果の属性を解釈する翻訳システムが必要となる。この翻訳システムを第三者が作成可能とすることは、本システムの実現において必須である。したがって、この翻訳システムを第三者が作成可能とするために、データ提供サイトは公開する属性の属性名、その属性の意味、そして、その属性のドメインの詳細 (例えばどのような値であるか、どのような表現規則を用いているか等) を公開する。

以上の議論より、データ提供サイトに求める要求を以下に示す。

- 標本 ID については、データ提供サイトが提供するデータの中では一意性を保障する。
- 標本 ID、学名以外のバウチャ情報の属性については、その属性名、その属性の意味、そして、ドメインの詳細についての情報を公開する。

3.3 本システムでの機能

- (1) 学名の翻訳システムへの問合せ機能
利用者が入力した学名とデータ提供サイトが管理しているバウチャ情報に記載されている学名とは同じ観念を表していても異なる可能性がある。その原因としては、学名のシノニムがあげられる。現在、学名のデータベースは様々な研究機関で作成が進められている。学名の歴史な変化および複数の分類体系を包含的に検索可能とするシステムの研究も行われている⁴⁾。それら学名データベースを用いて利用者が入力した学名からその学名とシノニムを持つと思われる学名群を検索する。本システムでは、この検索を行う機能を有する。なお、学名のシノニム集合を得るための API を定義し、様々な学名データベースとの接続可能性を提供する。
- (2) 標本 ID の本システム上での一意な ID への変換機能
データ提供サイトでは標本 ID はそのサイト内で一意になるように保障されている。これを本システム内で一意となるように変換する。
- (3) 各サイトへの問合せ機能
利用者が入力した学名は、学名の翻訳システム

によりシノニムの集合に変換される。そのシノニム集合の要素ごとに、データ提供サイトへ問合せを行う。また、本システムで用いている標本 ID からどのデータ提供サイトに問合せればよいかを判断し、そのデータ提供サイトのための標本 ID を用いて問合せを行う。

- (4) 学名以外の属性に関する翻訳システムへの問合せ機能

各データ提供サイトからの問合せ結果は、各データ提供サイトが格納している情報に近い形のデータとなっている。様々なデータ提供サイトからの情報を有効に利用するためには、利用者の要求に応じて、属性名の変換、および、属性値の翻訳が必要となる。高度な背景知識等を用いた属性値の高度な翻訳を実現するために、本システムではそれらを外部の機能として組み込む。すなわち、本システムでは、検索結果にある属性名から、その属性値がどのように記載されているかを判断し、利用者が要求する属性値の表現に適用するための外部の翻訳システムを特定する。そして、そのシステムを用いて属性値の翻訳を行い、利用者が要求している属性名と翻訳された属性値とに変換する。ここでは、変換の機能の API を定義することにより、様々な属性に関する翻訳機能との接続可能性を提供する。

- (5) 各サイトからの問合せ結果の統合
様々なデータ提供サイトへの問合せの結果群を一つの検索結果にまとめる機能を有する。その際には、利用者が要求する形式への変換を行う必要がある。

3.4 検索例

本システムは、様々なデータ提供サイトから必要な情報を収集し統合するための枠組みと捕らえることができる。本システムを用いて実現可能なデータ収集の一例を以下に示す。

- (1) ある学名が記載されているパウチャ情報へのリンクリストの作成
利用者は、パウチャ情報を必要とする学名を入力する。これは、種名だけではなく、学名の階層構造の上位の学名を入力することが可能となる。外部システムを利用して、種名の場合は、シノニムの集合に展開して検索を行う。階層構造の上位の学名の場合はその下位に属する種名の集合に展開して検索を行う。各データ提供サイトから標本 ID を得る。その標本 ID から利用者が参照可能なリンクに変換し、リンクリストを作成する。
- (2) ある学名と同定されている標本が採取位置の分布図
利用者が入力した学名からパウチャ情報を検索

```
<!ELEMENT id (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT location (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT otherinfo (#PCDATA)>
<!ELEMENT ids (id*)>
<!ELEMENT item
(id,name,location?,date?,otherinfo)>
<!ELEMENT items (item*)>
<!ELEMENT results (items|ids)>
```

図 3 必須および推奨のデータスキーマのための DTD

する。その結果により得られた属性に採取位置情報が記載されている場合は、その情報を緯経度情報に変換する。それらを地図と重ね合わせて表示することにより、分布図を作成することができる。また、地図上のプロットには、パウチャ情報へのリンクを埋め込むことにより、地図上のシンボルをクリックすることにより詳細な情報を表示できる。利用者が複数の学名を入力することにより、分布の違いが比較可能な分布図を作成することができる。

- (3) ある学名と同定されている標本が採取された時代変異の図
上記分布図と同様に検索を行い、検索結果より得られた属性に採取時刻情報が記載されている場合には、その情報を標準的な日時表示に変換する。これにより、採取時刻に関する処理（例えば統計処理）が可能となり、標本の採取時刻情報の分析結果等を表す図を作成することができる。

4. 実現方式

我々は、データ提供サイトの本システムへの接続可能性をよくするために、また、翻訳システムの本システムへの接続可能性をよくするために、それらとのデータ転送には XML⁵⁾ を用いた。その利点は以下の通りである。

- 人間でも読むことが可能である。
- 属性名や属性値が容易に変換できる。
- 自由にタグを定義できる。

本システムにおいては、データ提供サイトの特性を活かした属性を定めることを可能とする。そのためには、XML は非常に親和性が高いといえる。

我々は、表 1、表 3 に示したデータスキーマに対応する DTD を定義した。図 3 にその DTD を示す。ここで、id は各データ提供サイトが管理しているパウチャ情報の標本 ID を記述するのに用い、name は学

名を記述するのに用いる。location、date は、本システムが定める内部表現に従った場合の位置情報、および、時刻情報のために用いる。otherinfo には、各データ提供サイトが独自に定義しているバウチャ情報の属性の属性値を定義する。各データ提供サイトは、属性名をタグ名とし、その属性値をそのタグのノードの値として記載する。本システムの内部表現に従わない位置情報。時刻情報の表現の場合も、同様に記述する。これらのノードは、otherinfo の子ノードとして記述する。1つのバウチャ情報は、items に対応する。また、items、names はそれぞれ、item、name の集合を表すノードである。各データ提供サイトからの問合せ結果は result ノードとする。

データ提供サイトは、特性を活かした属性のために独自の DTD を定義する。検索結果にその DTD にて定義したタグを含める場合は、XML 名前空間⁶⁾を用いる。これにより、独自の表現方法であるかどうかの判断が可能となる。また、データ提供サイトはタグの意味とその値の表現方法の規則を別途定義する。

あるデータ提供サイト (foo.ac.jp) が採取位置を地名で管理しており、かつ採取時刻を「〇〇年〇月〇日」という表示形式を用いるものとする。このサイトが「茨城県つくば市」で「2002年6月1日」に「ID0001」という識別子にて管理されている「latipes」の標本を所有している場合の SearchItemByNames の検索結果例を図4に示す。検索例からわかるように、このサイトは独自の表現形式を用いていることがタグで判断できる。日本語の地名から緯経度を求める辞書³⁾を用いて採取位置等を変換することができる。その変換結果を図5に示す。複数のデータ提供サイトからの検索結果をそれぞれ図5のような形式に変換し、それら変換結果を統合することで、様々な分析に用いることが可能になる。例えば、統合した XML 文書を SVG⁷⁾ に変換し、採取位置を白地図上にプロットすることで、分布図を作成することもできる。

現在、これら設計に基づき、いくつかのデータ提供サイトを対象として、必須および推奨の API の実装を行っている。また、いくつかの翻訳システムの実装も行っている。今後はこれらの実装に基づいて検証実験を行う予定である。

5. 終わりに

本論文では、地球規模の生物種情報を統合するためのデータ交換を実現するシステムの設計について議論した。生物種情報の問合せに用いられる学名は、生物分類の見直しが行われる毎に変化するため、今後も変化する。この学名を用いた情報交換の枠組みに関する研究は、キーとなるドメインが常に変化するような情報に対する情報交換の枠組みに関する研究と捉えることができる。我々は最低限のデータスキーマを定義し、

```
<result xmlns:org=
"http://foo.ac.jp/schema">
<items>
<item>
<id>ID0001</id>
<name>latipes</name>
<otherinfo>
<org:location>茨城県つくば市
</org:location>
<org:date>2002年6月1日</org:date>
</otherinfo>
</item>
</items>
</result>
```

図4 検索結果の例

```
<result>
<items>
<item>
<id>ID0001</id>
<name>latipes</name>
<otherinfo>
<location> +36.10.40-140.04.01/
</location>
<date>2002-6-1</date>
</otherinfo>
</item>
</items>
</result>
```

図5 変換結果の例

そのスキーマに基づいて情報交換することを提案した。さらにはドメインの変化に対応するために、それぞれのドメインへの翻訳機能を自由に組み込むことができる方法を提案した。データ交換には、XML 技術を用いることで、柔軟なデータスキーマに対応する方法を提案した。

今後は、本論文での提案方式を実装し、実装に基づく検証実験を行う予定である。

謝 辞

本研究の一部は、地球環境研究総合推進費「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」の支援を受けている。

参 考 文 献

- 1) 北上始, 森康真, 有川正俊, 佐藤聡: 意味的な異種性を有する生物分類樹データベースの統合化方式, 電子情報通信学会論文誌 D-I, Vol. J82-D-1, No. 1, pp. 303-314(1999).
- 2) 馬渡峻輔, 動物分類学の理論 多様性を認識する方法, 東京大学出版会 (1994).
- 3) 相良毅, 有川正俊, 坂内正夫: ジオリファレンス情報を用いた空間情報抽出システム, 情報処理学会論文誌: データベース, Vol. 41, No. SIG6(TOD7), pp. 69-80(2000).
- 4) NOZOMI YTOW, DAVID R. MORSE, DAVID McL. ROBERTS: Nomenclator: a nomenclatural history to handle multiple taxonomic views, Biological Journal of the Linnean Society, No. 73, pp. 81-89(2001).
- 5) Extensible Markup Language(XML) 1.0, <http://www.w3.org/TR/REC-xml/>.
- 6) Namespaces in XML, <http://www.w3.org/TR/REC-xml-names/>.
- 7) Scalable Vector Graphics (SVG) 1.0 Specification, <http://www.w3.org/TR/SVG/>.