

## 蝶モニタリングのためのデータ管理システム

安川 雅紀<sup>†</sup> 前角 達彦<sup>‡</sup> 須田 真一<sup>‡</sup> 中原 美理<sup>‡</sup> 鷲谷 いづみ<sup>‡</sup> 喜連川 優<sup>§</sup>

<sup>†</sup> 東京大学地球観測データ統融合連携研究機構 〒153-8505 東京都目黒区駒場 4-6-1

<sup>‡</sup> 東京大学大学院農学生命科学研究科 〒113-8657 東京都文京区弥生 1-1-1

<sup>§</sup> 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: <sup>†</sup> yasukawa@iis.u-tokyo.ac.jp,

<sup>‡</sup> corneavant02@yahoo.co.jp, s-suda@jcom.home.ne.jp, {amirin, awashi}@mail.ecc.u-tokyo.ac.jp,

<sup>§</sup> kitsure@tkl.iis.u-tokyo.ac.jp

**あらまし** 本研究では、市民参加型モニタリングによって得られる東京の蝶に関する大量のデータを収集する機能、そのデータを修正・品質チェックする機能、データ公開の機能等を備えたデータ管理システムを開発した。また、市民参加型モニタリングの新しい試みとして、デジタルカメラで撮影された画像も取り扱った。2年間で約1万件のデータを収集し、品質の高いデータを公開した。このようなシステムは、生物多様性条約第10回締約国会議(COPI0)で採択された愛知ターゲットに資するものである。

**キーワード** 生物多様性, 市民参加型モニタリング, データアップロード, 品質管理, データ解析

## A Data Management System for Butterfly Monitoring

Masaki YASUKAWA<sup>†</sup> Tatsuhiko MAEZUMI<sup>‡</sup> Shin-ichi SUDA<sup>‡</sup> Miri NAKAHARA<sup>‡</sup>  
Izumi WASHITANI<sup>‡</sup> and Masaru KITSUREGAWA<sup>§</sup>

<sup>†</sup> Earth Observation Data Integration and Fusion Research Initiative, the University of Tokyo  
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

<sup>‡</sup> Graduate School of Agricultural and Life Sciences, the University of Tokyo  
1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-8657 Japan

<sup>§</sup> Institute of Industrial Science, the University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

E-mail: <sup>†</sup> yasukawa@iis.u-tokyo.ac.jp,

<sup>‡</sup> corneavant02@yahoo.co.jp, s-suda@jcom.home.ne.jp, {amirin, awashi}@mail.ecc.u-tokyo.ac.jp,

<sup>§</sup> kitsure@tkl.iis.u-tokyo.ac.jp

### 1. はじめに

1992年にブラジルのリオ・デ・ジャネイロで開かれた地球サミットで「生物多様性条約」が採択されて以降、健全な生態系を維持し、持続可能な社会を築くためのキーワードとして「生物多様性」という言葉が多く使われてきた。2010年に名古屋で開催された生物多様性条約第10回締約国会議(COPI0)では「愛知ターゲット」が採択され、生物多様性のデータや情報の収集、解析やそれによって得られる新しい知識の共有の強化についても記載された[1]。日本では2009年に制定された生物多様性基本法の基本原則に沿って、今後は様々な政策が実施されるであろう。

上記のターゲットや政策に貢献するには、まず、生物多様性の現況や傾向を知る必要があり、モニタリングによる大量のデータ取得が不可欠である。モニタリ

ングする手法としては様々な方法があるが、市民参加型のモニタリングは、大量のデータが得られ、時間的・空間的に高密度の広域モニタリングが可能である[2]。一般の調査員が調査マニュアルに従って定期的に生き物を調査して調査用紙に記入し、事務局に提出する。しかし、一般の調査員は生き物調査に関する知識が少なく、調査に参加した初期の頃は種の同定での誤判別や調査項目の記入ミスが多い。データが紙媒体であったり、データに誤判別・記入ミスが含まれていたりするため、このデータを生態学の解析に用いるにはデータのデジタル化やクレンジングが必要であり、データ利用者にかかる負担は大きかった[3]。

また、デジタルデータが大量に収集された際には、そこから新しい知見が容易に得られる様な高価値の解析ツールが不可欠である。従来、生態学の研究者は、表計算ソフトやGISソフトなどを用いて半手動で手間

暇かけて解析結果の図を作っていた。しかし、取得されるデータが多くなればなるほどデータの整理やハンドリングに要する時間が増え、研究者が本来行うべき解析の時間が減る傾向があった。また、データの種類が増えると既存のソフトでは解析が困難になる。そのため、データ管理ツールとデータ解析ツールが同一プラットフォーム上にあり、データ解析ツールが高機能であることが必要である。

現在、我々は生態学の研究者および市民生協の一つであるパルシステム東京との協働で、東京の蝶のモニタリング活動を行っている。蝶は幅広い環境に生息し一般市民に馴染み深く種を識別しやすいことから、モニタリングのターゲットとして採用している。IT側のミッションとして、データ管理ツール、データ公開ツール、データ解析ツール等を有する蝶のモニタリングシステムを構築中である。

本論文では、市民参加型モニタリングによって得られる東京における蝶のデータを調査員自身がデータベースにアップロードする機能、データをクレンジングして品質管理する機能、データ公開の機能等を備えたデータ管理システムを開発したので報告する。なお、市民参加型モニタリングの世界的に新しい試みとして、調査データはデジタルカメラで撮影された蝶の画像も添付するルールを取り入れ、管理出来る様にした。これはデータのクレンジングや品質管理で非常に役立つためである。このシステムにより、生物多様性研究における質の高い蝶のデータの共有を目指す。2年間で約1万件のデータを収集し、品質の高いデータを公開した。また、データ解析ツールの開発も積極的に進めている。このようなシステムやデータセットは、生物多様性条約第10回締約国会議(COP10)で採択された愛知ターゲットに資するものである。

## 2. 関連研究

生物多様性に関するデータベースは、分類学的ライブラリや知識情報のライブラリや、モニタリングのデータを蓄積したデータベースが存在する。

ESABII Database は様々な種について学名やDNA情報、標本情報と標本画像、分布情報を収録したデータベースで、地球規模生物多様性情報機構(GBIF)の日本ノードを担当している。百科事典的要素が強く、モニタリングデータ管理・品質管理・データ公開を有する本研究のようなシステムとは異なる[4]。

NaGISA (Natural Geography In Shore Areas : なぎさ) Database は海洋生物センサスの野外研究プロジェクト(Census of Marine Life: CoML)で、海辺を調査してCoMLのデータベースOBISに登録するもので、データベース上で品質管理はされておらず、品質はデータ

提供者次第である。データ解析ツールは考慮されていない[5]。

英国は市民参加型の生物モニタリング活動が非常に活発な国として知られている。1995年から英国チョウ類保全協会では、国土全体を10km区画に分けてボランティアがチョウの種類や数を数える調査を行っており、データ量は2009年までの15年間で500万件以上をアーカイブサイトで公開している。画像アップロードやデータ品質管理の機能は有していないため、データの信頼性が低く可用性に欠点がある[6]。

## 3. 蝶データ管理システムの構築

### 3.1. システム設計

本節では、システムの設計について検討する。

近年のインターネット、PCおよび携帯の普及により、研究者だけでなく一般市民もそれらを手軽に使える時代になった。そこで、調査データのアップロード機能、クレンジングおよび品質管理機能、データ公開機能はWeb上で行うことができるようにする。つまり、調査員サイドからデータはデジタルで取り扱われることになる。このことにより、従来では、生態学の研究者が調査用紙を回収してそれをデジタル化していたが、その手間が省略される。また、調査員の手元にはデジタルデータとして保存され、データの整理・加工で取り扱いが手軽になる。

調査データのアップロード機能では、PCに不慣れな調査員でも容易に使うことが出来るようにする必要がある。そこで、データ入力は表計算ソフト用フォーマットを使用せず、Webページのフォーム上に直接入力する方式とする。また、画像は送信フォームに添付できるようにする。各項目の入力は、タイピングミスを防ぐために、可能な限り選択方式を採用する。例えば、日時、天気、風、蝶の種名、その性別、行動等がある。また、同じ日時・場所で調査することが多いため、cookieを利用して、調査員名、調査員番号、日時、場所等の項目の内容をプリセットしておく。誤送信を防ぐために、データアップロードの前に送信確認画面を用意する。画像は、アップロードされた原画像およびサムネイル画像をサーバ上に保存する。

クレンジングおよび品質管理機能では、生態学の研究者が調査員によってアップロードされたデータの間違いをWeb上で修正してデータ品質のランクを設定できるようにする。データが品質管理されていなければ公開時のデータの信頼性や利用性が低下するため、重要な機能である。機能のバックヤードでは、データ修正の内容と履歴を保存し、修正前の状態にいつでも復元できるようにする。住所から緯度・経度に変換して保存する。調査員の種名同定の学習のために、調査

員の同定結果と蝶の画像を使った研究者の同定結果の両方を保存する。上記と同様の理由で、調査員への連絡事項も保存する。

データ公開機能では、Web上で調査データを閲覧できるようにする。一般に公開してモニタリング活動のアウトリーチを行うとともに、クレンジングされたデータを調査員へ提供して調査員自身のモチベーション維持を図る。公開データは、個人情報あるいはそれを特定できる情報が表示されないよう工夫する。期間、種名、調査員、地図、画像等、様々な角度からのデータ検索が行えるようにする。調査データの詳細表示ページでは、個人が特定できないよう考慮しつつ、調査項目、地図上での調査地点、蝶の画像を表示する。

上記の各機能を連結し、1つのサーバ上で実現する。このことにより、調査員や生態学の研究者が利用し易い環境を構築し、また、システムのメンテナンス性や機能拡張性を保持する。

### 3.2. システム構成

本システム概念図は図1のようになる。流れは以下の通りである。Webを通して、蝶の調査員がデータアップロードページにアクセスし、調査項目入力および画像の添付を行い、データをサーバにアップロードする。サーバは取得したデータを自動的にデータベースに登録し、画像をストレージに保存する。生態学の研究者は、データ編集ページにアクセスし、アップロードされたデータの修正および品質チェック・フラグ付けを行う。データ公開ページでは、クライアントの検索条件に沿ったクエリを生成してデータをデータベースから引き出してデータ表示ページを自動生成してクライアントに結果を返す。データは、品質が高いとフラグ付けされたデータのみを公開する。

本システムは、蝶のデータと気候データとの統合など様々な機能の拡張を考慮して、データ統合・情報融合コアシステムのプラットフォーム上で稼働させる。データ統合・情報融合コアシステムは、ペタバイト級のストレージに地球環境に関する様々なデータをアーカイブして統合・解析することにより、科学的・社会的に有用な情報に変換することができる知の創出がある[7]。

表1に本システムのハードウェアとソフトウェアの構成を示す。データ統合・情報融合コアシステムのサーバは、大規模なデータ処理用サーバとデータ公開用サーバが存在するが、本研究ではデータ公開用サーバを使用し、表1にはその仕様のみ掲載した。本システムを利用するにはWebブラウザの他にJava Runtime EnvironmentとAdobe Flash Playerが必要である。

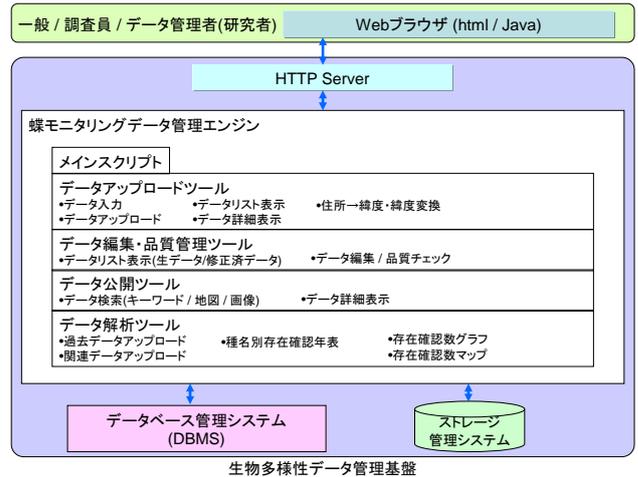


図1 本システムの概念図

表1 本システムの仕様

Hardware	Hitachi HA8500 (Intel Itanium 1.66GHz 4 processors 8 cores, 128GB Mem., 1.6PB Disk Array)
Software	Red Hat Enterprise Linux AS release 4 Apache 2.2.16, PHP 5.2.16, MySQL 5.5.6

## 4. システムの利用

### 4.1. 調査データ

調査員は、調査マニュアルに従って蝶の調査を行う。その際、図2のような項目について記録し、デジタルカメラで蝶を撮影する。調査データは調査項目と撮影画像で構成される。よって、複数の蝶の存在を確認した場合は、その個体数分の調査データを作成する。なお、天候、蝶の性別、蝶の行動等は選択肢の中から該当するものを選ぶ。各調査項目を記録するためのデバイスは、調査員自身の調査スタイルを考慮して自由としている。また、デジタルカメラでの撮影に関してルールは決めず、解像度は自由で、携帯電話での撮影も可としている。



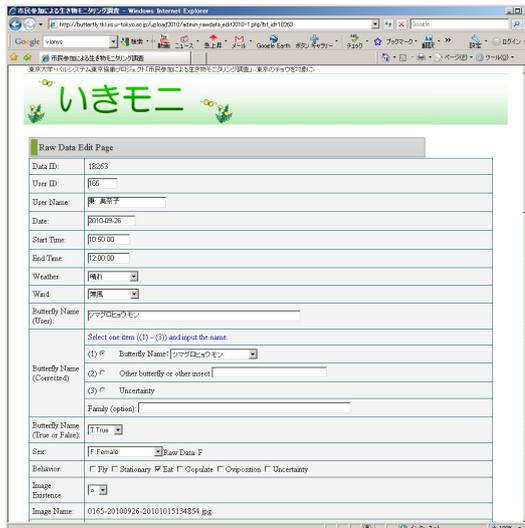
図2 調査データ

## 4.2. データアップロード

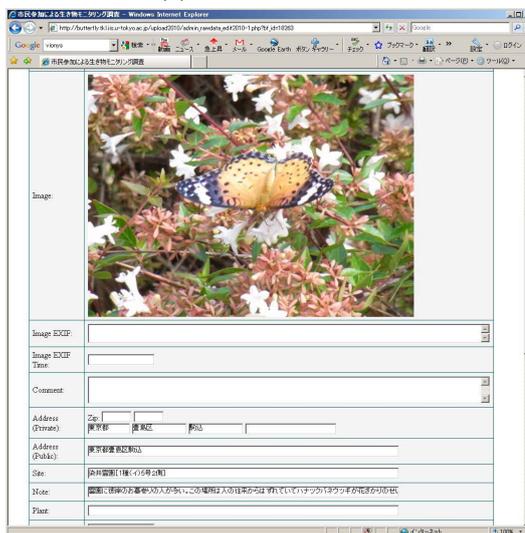
調査員が記録したデータは、調査員自身で図3のデータアップロード用のWebページを使って項目の入力、入力確認、サブミットを行ってアップロードすると、データベースに自動的に登録される。調査員の誤入力を防ぐために、調査日、調査時刻、天候、種名、蝶の性別、蝶の行動等、半数以上の項目を選択式の入力フォームとした。入力された住所は、自動的に緯度・経度に変換してデータベースに保存される。また、調査員が同じ日時・場所で複数の記録をすることが多いため、前回アップロードした時のWebのcookieを利用して、調査員名、調査員番号、調査日、調査時刻、調査地等の項目欄をプリセットする機能を備えた。

## 4.3. データクレンジングと品質管理

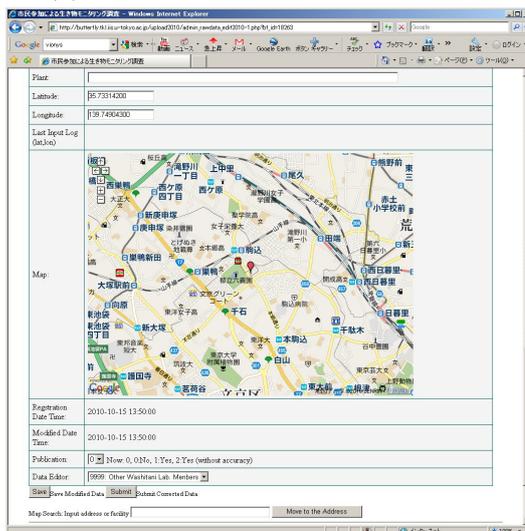
データベースに登録したデータ(以後、生データ)は、データ管理者ページにおいてリスト形式で自動的に生成し表示する。生データは、誤入力や蝶の種名の誤同定などが含まれていることがあるため、修正してデータの品質を保つ必要がある。そこで、データ管理者である生態学の研究者は、生データのリストページから1つのデータを選択し、図4のWebベースの生データ編集ページを用いて生データを修正し、データ品質のフラグを付与する。具体的には、テキスト入力による修正あるいは選択式による修正、図4(b)の画像を用いて研究者が同定した蝶の種名の入力、調査員によって同定した種名の正解・不正解のフラグ付け、Google Mapsを用いた調査地の緯度・経度の修正、研究者によるコメント記入、データ品質フラグ付け、データ修正者名記入、編集時刻入力等が可能である。種名同定の正解・不正解フラグは、調査員の種名同定学習に有用である。データ品質フラグは、データを公開あるいは解析する際に非常に重要な要素である。



(a) 基本調査項目



(b) 蝶の画像およびコメント記入欄



(c) 調査地マップおよびデータ品質フラグ付与欄

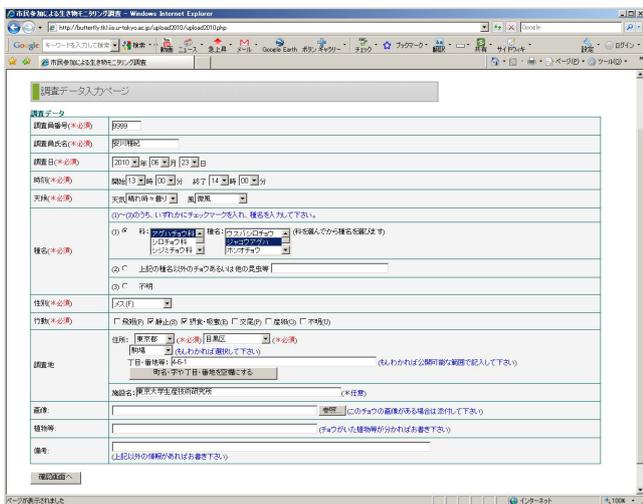


図3 データアップロードページ

図4 データ編集ページ

#### 4.4. データ公開

蓄積され品質管理された調査データを調査員・一般・研究者に公開するため、データ公開サイトを構築した。プロジェクト2年間で、調査員数はのべ約500名、データ数は約1万件である。サイトは2010年3月に公開した。

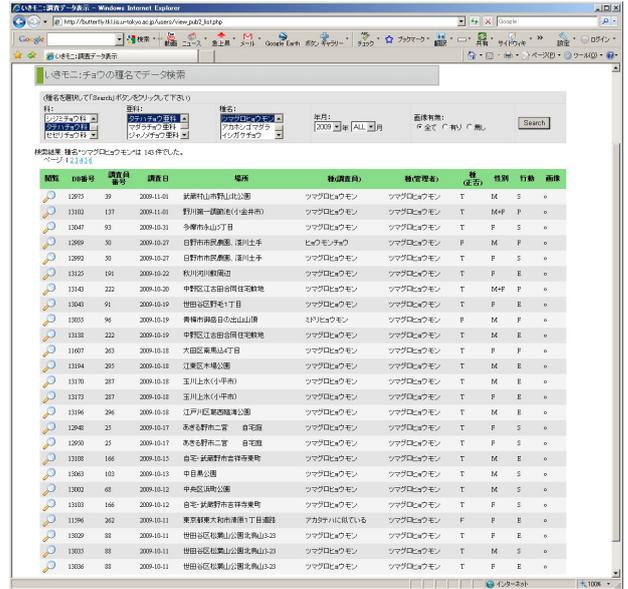
図5に公開サイトのトップページを示す。このページはパルシステムと相互リンクを張っており、一般の方がアクセスすることができる。公開サイトに対しては、種名からのデータ検索、地図からのデータ検索、画像からのデータ検索、調査員番号からのデータ検索、データ詳細表示等の各機能を開発した。なお、調査員番号からのデータ検索のみ、利用するには調査員または研究者のユーザ認証が必要である。他の機能については自由に利用することができる。



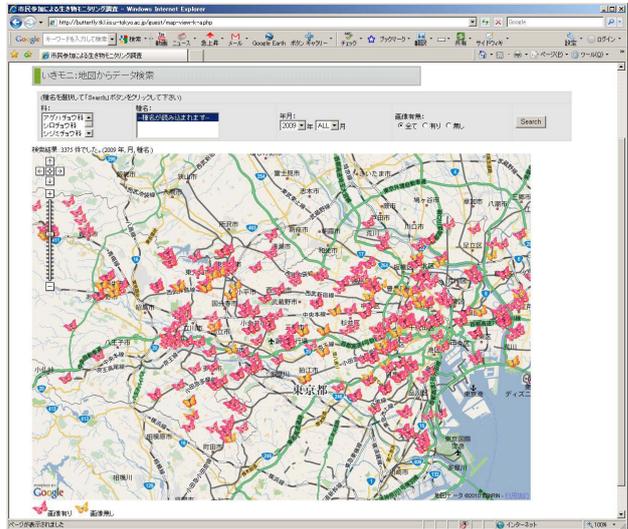
図5 データ公開サイトのトップページ

図6(a)に蝶の種名からのデータ検索の例を示す。これは、検索条件で蝶の種名、年月、蝶の画像有無を指定すると、WebサーバはSQL文を生成しデータベースから条件の合致したものを抽出し、データの概要をリスト形式で表示することができる。指定した種名の存在確認件数や時期、場所等を知るのに便利である。表の左には虫眼鏡のアイコンがあり、当該データの詳細ページにリンクが張られている(図7)。

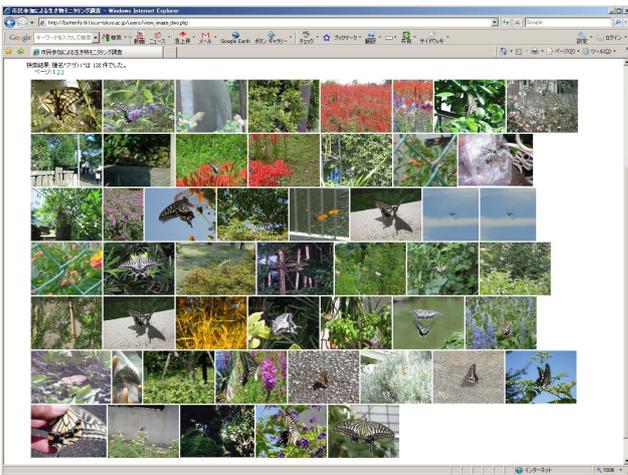
同様に調査員番号からのデータ検索も可能である。この検索は、調査員自身のデータ整理に有用であり、蝶の種名の同定結果を知ることができ、同定学習に非常に役立つ。蝶モニタリングに対するモチベーションの維持にも有効である。



(a) 種名・年月からのデータ検索



(b) 地図からのデータ検索



(c) 画像からのデータ検索

図6 データ検索

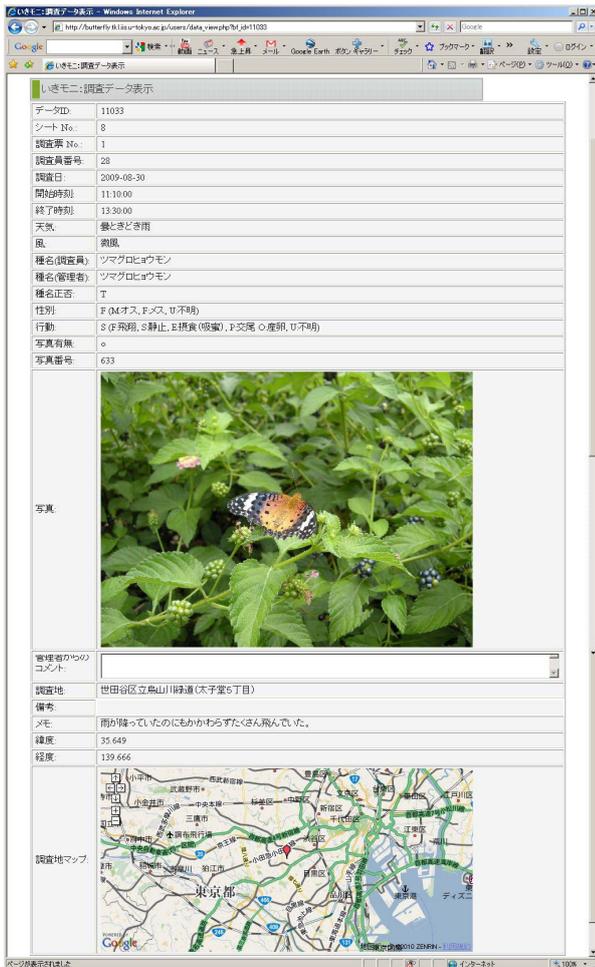


図7 データ詳細ページ

図6(b)は地図からのデータ検索の例である。検索条件に科名・種名, 年月, 画像有無を指定すると, データベースから条件にマッチしたものを抽出し, 調査地を地図上にプロットすることができる。指定した種の空間分布やその時系列変化を見るのに有用である。また, 地図上にプロットした蝶のアイコンにはリンクが張られており, アイコンをクリックすると図7と同様にデータの詳細を閲覧することができる。

図6(c)は画像からのデータ検索の例である。我々の蝶モニタリングプロジェクトでは調査の際に蝶の画像を撮影することになっているため, 非常に多くの画像がシステムにアーカイブされている。検索条件に種名を入れると, データベースから条件にマッチしたものを抽出し, 画像のサムネイルをタイル状に並べて表示することができる。種名同定のための学習や蝶の図鑑作成に有用である。また, 蝶は同じ種名・性別でも時期によって大きさや模様が異なることを利用して, 気候変化と蝶の関係を解析したり, 画像の背景には木や植物など周辺環境が写り込んでいるために蝶の種名毎の生態を解析したりすることも, システムを拡張して

いくことで可能となる。

図7に調査データの詳細ページを示す。詳細データ表示機能は, データベースから調査データを引き出し, Webブラウザ上で表示できるようにページを自動生成することができる。種の同定の項目では, 調査員自身の同定の正解・不正解を知ることができ, 調査員の種の同定学習に有用である。調査地を中心とした地図を表示することにより, その種がどの生息環境区分で発見されたかを知ることができる。個人情報保護の観点から, 緯度・経度の表示および地図の解像度は, 個人の建物等を特定できない程度に粗く表示するよう工夫した。

図8にデータ公開サイトのアクセス数を示す。サイトは2010年3月に公開したが, ほぼ順調にアクセス数が増加したことがわかった。

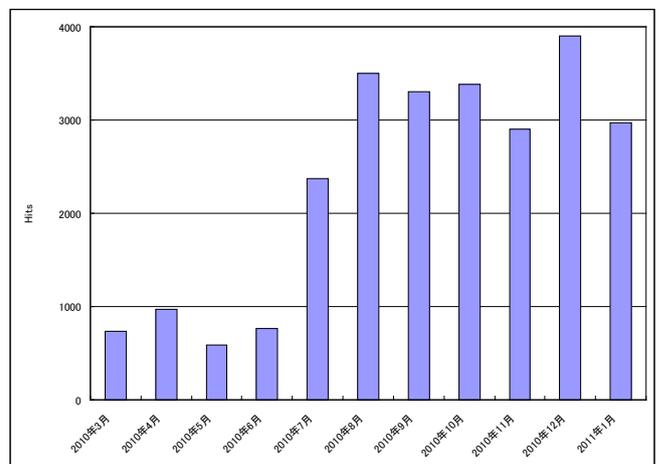


図8 データ公開サイトのアクセス数

#### 4.5. データ解析

本節では登録されたデータを用いた蝶の解析例について述べる。

図9は, 図6(b)を用いてアカボシゴマダラを検索して分布をマップで表示したものである。関東に生息するアカボシゴマダラは, 中国大陸亜種の外来種であり, 神奈川県鎌倉市で放蝶され以来, 生息域を広げ, 本来生息しているゴマダラチョウの数が減っているといわれており, 監視が必要な蝶である。図9のようにアカボシゴマダラが東京都の広い範囲で生息が確認され, 生息域を広げていることが確認できる。

図10に種名毎の存在確認件数を示す。行が種名, 列に年毎の存在確認件数であり, 調査で種が確認されれば「○」とその件数を表示し, 確認されなければ「×」を表示する。種によって生息できる環境が異なるため, 蝶と温暖化・都市化の関係を調査することが可能である。

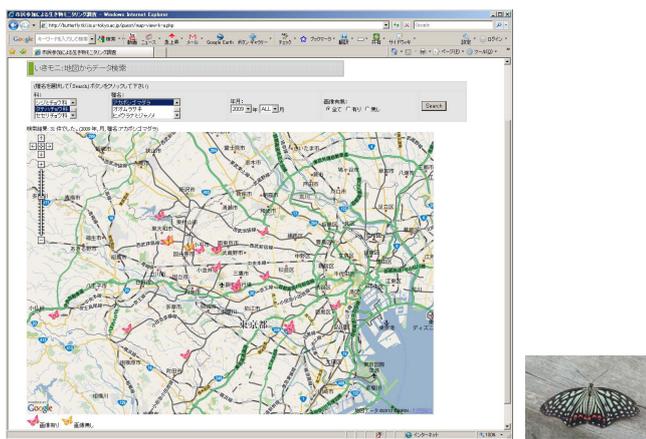


図9 アカボシゴマダラの分布(2009年度)

種名	Scientific Name	2009-2009	2010-2010	2009-2010
ヒメギフチョウ	Lusidictya puzosi	x	x	x
ギフチョウ	Lusidictya japonica	x	x	x
フスハキチョウ	Panaeetus evenemanni	x	x	x
ヒメスハシロチョウ	Panaeetus boeni	x	x	x
フスハシロチョウ	Panaeetus citrinatus	○(1)	○(1)	○(2)
ベニモンアゲハ	Pachloptis aristolochiae	x	x	x
シヨウアゲハ	Eyes alcinous	○(2)	○(1)	○(2)
キアゲハ	Papilio machaon	○(4)	○(4)	○(4)
アゲハ	Papilio xuthus	○(17)	○(17)	○(14)
オナガアゲハ	Papilio machletus	○(1)	○(1)	○(1)
シロホアゲハ	Papilio polytes	x	x	x
クロアゲハ	Papilio protenor	○(1)	○(1)	○(1)
ナガサキアゲハ	Papilio nermana	○(1)	○(1)	○(1)
モンキアゲハ	Papilio helenus	○(1)	○(1)	○(1)
カラスアゲハ	Papilio dehaani	○(1)	○(1)	○(1)
ミヤマカラスアゲハ	Papilio maackii	x	○(1)	○(1)
アオスジアゲハ	Graellsia japonica	○(1)	○(1)	○(1)
ミドリアゲハ	Graellsia japonica	x	x	x

図10 種名毎の存在確認件数表

## 5. まとめ

本研究では、東京の蝶のモニタリングデータ管理システムを開発した。具体的には、データアップロード・データベース自動登録機能の開発、データ編集機能の開発、各種データ検索機能およびデータ詳細表示機能を有するデータ公開サイトの開発・構築を行った。本システムは世界に類を見ない、画像収集を重要視した市民参加型の生物多様性モニタリングシステムである。既に実運用しており、2年間で約1万件のデータを登録し一般に公開した。また、生物多様性条約第10回締約国会議のサイドイベントである「生物多様性交流フェア」では、本システムのデモンストレーションを行い、概ね好評を得た。

今後は、解析ツールの拡張、過去の蝶のモニタリングデータへの適用とそのデータを用いた時系列解析ツ

ールの開発、蝶以外の生物モニタリングデータへの適用等を行っていく。また、システムの改良を重ねていくことで、市民参加型の生物多様性モニタリングデータに対するクラウド・ソーシングの効率的なフレームワークを具体化させる。

## 参考文献

- [1] Convention on Biological Diversity (CBD), “Report of the Tenth Meeting of the Conference of the Parties to the Convention on Biological Diversity”, <http://www.cbd.int/doc/meetings/cop/cop-10/official/cop-10-27-en.pdf>, pp. 118-120, 2010.
- [2] J. Silvertown, “A new dawn for citizen science”, *Trends in Ecology & Evolution*, Vol. 24, No. 9, pp. 467-471, 2009.
- [3] 藤原宣夫, 日置佳之, 須田真一, “-MBR 方式による- 住民参加の生きもの調査ガイドブック”, 国土技術政策総合研究所資料, No. 139, 2003.
- [4] East and Southeast Asia Biodiversity Information Initiative (ESABII), “ESABII Database”, <http://www.esabii.org/>, 2010.
- [5] The Natural Geography In Shore Areas Project (NaGISA), “NaGISA Database”, <http://www.esabii.org/>, 2011.
- [6] The UK Butterfly Monitoring Scheme (UKBMS), <http://www.ukbms.org/>, 2011.
- [7] Data Integration & Analysis System (DIAS), <http://www.editoria.u-tokyo.ac.jp/dias/>, 2011