

MusethequeV3

GBIF用データ生成作業の実際

—2つのデータベースをつなぐもの—

ユーザID

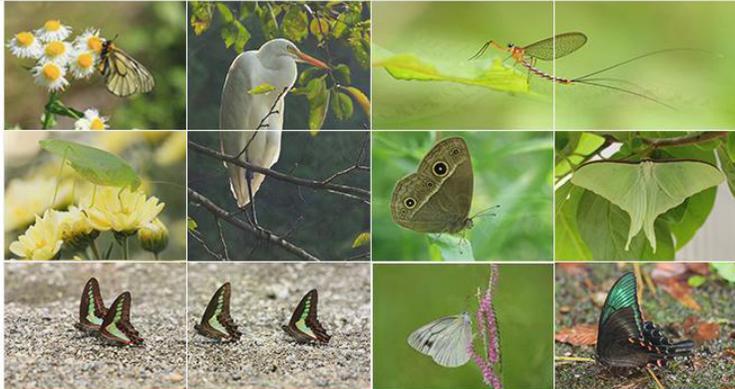
パスワード

ALL RIGHTS RESERVED, COPYRIGHT (C) F

ホーム お問い合わせ サイトマップ English twitter



地球規模生物多様性
情報機構日本ノード
Japan Node of Global Biodiversity Information Facility



世界中の生物多様性に関するデータの検索



和名による検索の詳細はこちらをご参照ください。



- GBIFデータの利用 →
- GBIFへのデータ登録 →
- GBIFニュースレター(日本語版)  →



群馬県立自然史博物館の概要

所在地: 群馬県富岡市
上黒岩1674-1

開館: 1997年10月

年間入館者: 15~18万人
学芸係: 9名(次長含む)

次長 1名

生物系 6名

地学系 2名



Welcome to Gunma Museum of Natural History

群馬県立自然史博物館の収蔵システム

1997 第一次システム



2004 第二次システム



2007 第三次システム

館独自のシステム

ビジュアルな収蔵紹介に主眼
「備品管理機能」にもウエイト
学名情報については弱点があった

富士通Musetheque Ver.2をカスタマイズ

合わせて入力項目の見直し
学名辞書の搭載可能化
エラーチェック機能の搭載

富士通Musetheque Ver.3をカスタマイズ

2次システムを踏襲するが
入力・修正の効率化と画面の大幅修正

群馬県立自然史博物館の収蔵システム

収蔵システムの入力フィールドと配置

2次システムでの収蔵システム項目決定の方針

動物と植物・菌類で項目を分ける

- ・命名規約自体が異なる
- ・菌類や着生・寄生植物の「基質・寄主」、動物の「性」など

採集記録フィールドは極力細分化

→「エクセルはセルを結合するより分割する方が面倒」

結果的に収蔵システムの項目はDarwin coreにほぼ対応していた

しかし、当館収蔵システムとS-Net/GBIFの項目の配列や、形式には微妙な差異がある。

(例)

- ・英文表記と和文表記の順序
- ・「採集日」のはじめ・おわり、「海拔・水深」の最小・最大

群馬自然史博でのデータ変換作業の実際

S-Net/GBIF公開データ選択の選択
(収蔵システムからのダウンロード)

各資料担当 ↔ GBIF担当

↓
フィールド選択・結合・削除

データ整理員(委託) ↔ 各資料担当

↓
入力項目のチェック・形式変換

データ整理員(委託) ↔ 統括:GBIF担当

↓
データ生成ツールによるチェック・
ファイル完成

GBIF担当

↓
入力情報・RDB情報の最終確認

GBIF担当 ↔ 各資料担当

それでは変換作業の流れと、学名情報・採集情報の変換作業の実際を順に説明します。

学名情報の
入力と注意事項

採用する分類体系と学名

収蔵システムで登録された分類体系と生物名
＝登録時点での分類体系と生物名

運用開始から16年・・・

→分類学的取扱の変更、命名規約の変更、・・・

しかし過去にさかのぼって分類情報のアップデートは難しい

入力ミスへの対応

→大半の分類群では辞書によるミスチェックとゆらぎ防止

しかし1次システムでは辞書はなく、入力ミスはそのまま

生物名・分類情報のミス・不統一の発見は
人材派遣会社から委託されたデータ変換作業員では実質無理。
: GBIF担当から各資料担当に照会し、修正している。

群馬自然史博でのデータ変換作業の実際

学名情報の 入力と注意事項

植物学名と変換ツールの悩み

「種以下の分類群」の煩わしさと命名者の問題？

- ・群馬自然史の場合、亜種・変種・品種のフィールドを分けているが・・・
2次システム以降のデータは収蔵システムの学名フィールドと1対1対応。

雑種式への不対応

- ・しかたがないので、たとえば *grypoceras* x *V. kusanoana* も種に対応するものと割り切って考えるしかないが・・・

学名著者の小・2世の f. に対する変換ツールの反応

- ・L. f. や Hook. f. などの f. を品種を示す f. と間違えて分割してしまう。
収蔵システムからのエクスポートファイルの f. をいったん fil. に置換し、
ツールにかけた後に f. に戻す方法もあるが・・・
→ 品種の f. まで置換するおそれがあり、1点ずつ確認が必要。

産地情報の
チェックと変換

地名入力にまつわる問題点

地名の変遷への対応

運用開始から16年・・・

→町村合併とそれに伴う町域変更の確認が大きなネック

英文表記の不統一への対応

資料担当内で統一が図られていればいいと割り切っているが・・・

次の場合は注意が必要

- ・外国人にわからない自然地名、施設名（そのままローマ字に直しただけのもの）
- ・Nativeでなければわからない地形表現
fen、mire、bogは（下手したらswampもmarshも）湿原
溪谷もvalley、stream、glen

国交省やJRの表記ほか、観光庁が指針で示した英語表記がベースラインになる？

群馬自然史博でのデータ変換作業の実際

Darwin coreの標高・水深・日付の表記

最小値・最大値両方を記入する形式である。

→しかし収録システムのこれらのフィールドは1項目

標高・水深に幅がある場合の収録システムの対処法:

- ①「その他住所」に 500- 600 m in alt.のように記入する。
- ②「採集に関する備考」に記入する。 **対処法は資料担当によってまちまちである。**

日付の幅に対する対処法

→2次システムは不明部分*でエクスポートされるようになった。

ところが！

**1次システムでは日付不明に対するルールがなく、00、99、空白が混在
これが今も変換ツールエラーのもとになっている。**

産地情報の
チェックと変換

収蔵システム緯度経度への対応

BA				BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ								
緯度経度				緯度変換式				経度変換式																			
				↓=MID(\$BA4,2,2)																							
+363220.280,+1384038.980				36	32	20	138	40	38	36	度	32	分	20	秒	138	度	40	分								
				↑=MID(\$BA4,4,2)																							
+362814.000,+1392318.000				36	28	14	139	23	18	36	度	28	分	14	秒	139	度	23	分								
				↑=MID(\$BA7,14,3)																							
BL	BM	BN	BO	BP	BQ	BR	BS	BT				BU				BV				BW							
								緯度合成式				経度合成式				緯度数値				経度数値							
								↓=BH4&BI4&BJ4&BK4&BL4&BM4																			
												↓=BN4&BO4&BP4&BQ4&BR4&BS4															
20	秒	138	度	40	分	38	秒	36度32分20秒				138度40分38秒				BT、BU列を値複写											
0	秒	0	度	0	分	0	秒	度分0秒				0度0分0秒				36度34分21秒				139度20分12秒							
14	秒	139	度	23	分	18	秒	36度28分14秒				139度23分18秒				36度32分20秒				139度23分18秒							
	秒		度	0	分	0	秒	分秒				度0分0秒				度分秒				度分秒							
	秒		度		分		秒	度分秒				度分秒				度分秒				度分秒							
	秒		度		分		秒	度分秒				度分秒				度分秒				度分秒							

05 変換作業が終わった後に電子地図・地名からのマッチング作業を行う

群馬自然史博でのデータ変換作業の実際

知ると知らないで大違いー変換作業に役立つExcel関数

基本操作は「照合する」「分ける」「合わせる」

「照合する」

=IF(A_n=A_(n+1),"削除","必要"): 標本番号でソート後標本番号重複チェック

=VLOOKUP(検索元,検索列はじめ:検索列おわり,2,FALSE):

学名辞書とともに市町村合併情報、地名読みの辞書化もできる

「分ける」

=MID(照合元,抽出開始字数,抽出字数)

注意:この関数はフィールドが定型でないと恐ろしい目に遭います

「合わせる」

=CONCATENATE(番地,番地,...): =番地&番地&番地&...も同様に処理

注意:英文フィールドで「,」と直後の半角スペースを入れる場合、あらかじめ、の列を用意します。

正解に近かった収蔵データベース

システム移行の際に議論された項目と形式

: Windows95の時代からデータベースの積極利用

: 結果的にフィールドがDarwin Coreに近いものであった

→ 研究現場で欲しい情報は概ね一致する

項目の「適正なレベルでの」細分化

: システム不定形フィールドの分離の難しさ: 設計時のSE提案

階層型辞書機能の追加

: 入力の省力化のみならず誤入力・分類のゆらぎ防止

群馬自然史博におけるデータ変換作業の現実

前世紀のデータとの戦い？(笑)

- : 古い学名・分類学的取扱、地名、日本測地系
- : 1次システムは生物名の入力ミスノーチェック

ローマ字表記のゆれ、英語表現のチェック

- : 担当によって統一されていない。
- : 館収蔵システムにマニュアルはない。

データ作業員の人材難

- : 県の中心から1時間、電車は1時間に1本、
最寄駅徒歩30分、標高差60mの急勾配
館からマイカー30分圏内に大学はない。

収蔵データベースとその標準化ー今後の課題

APG III 分類と学名のGlobal Standardへの対応

:いつどこが音頭を取ってどれだけの機関が同調するか？

過去のデータのアップデートはどうする？

→自動変換ツールでもないと労多くしての危惧

Darwin core 2への対応

:分野によっては現行に比べて必要性の薄い項目を増やすだけ？

特に測位方法・精度の入力は煩雑！

学芸職員間の意志の統一と引き継ぎ

:公立博物館なるが故の人事異動

時として(来る者も去る者も)不本意な異動