


趣旨説明

国立科学博物館 標本資料センター
神保 宇嗣



生物標本

=ある生物がいつ、どこにいたかを示す物的証拠
=さまざまな形で利用できる資源

膨大な標本の蓄積

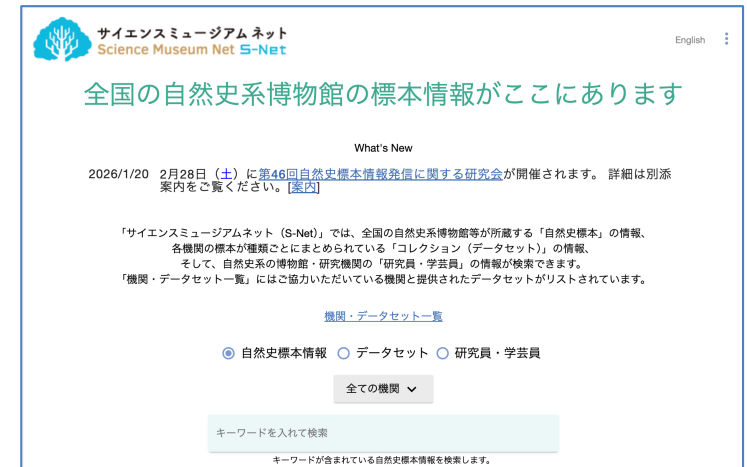


モノ＝標本としての
利活用

コト＝情報としての
利活用

サイエンスミュージアムネット (S-Net)

- 国立科学博物館が運営する自然史系標本データの共有ネットワーク
- 博物館・研究所など100以上の機関から提供を受けた約817万件のデータを公開



<https://science-net.kahaku.go.jp/>

「ミヤマカラスアゲハ」を検索



サイエンスミュージアム ネット
Science Museum Net S-Net

ホーム > 検索結果

自然史標本情報

レコード種別
PreservedSpecimen (4,486)

機関名 (日本語)
姫路科学館 (607)
ふじのくに地球環境史ミュージアム (586)
[もっと見る](#)

コレクションコード
Insecta (867)
In (672)
[もっと見る](#)

国 (日本語)
日本 (4,040)
日本国 (55)
[もっと見る](#)

都道府県 (日本語)
北海道 (771)
長野県 (473)
[もっと見る](#)

タイプ標本種別
Holotype (4)

学名
Papilio maackii (1,229)
Papilio maackii Menetries, 1858 (737)
[もっと見る](#)

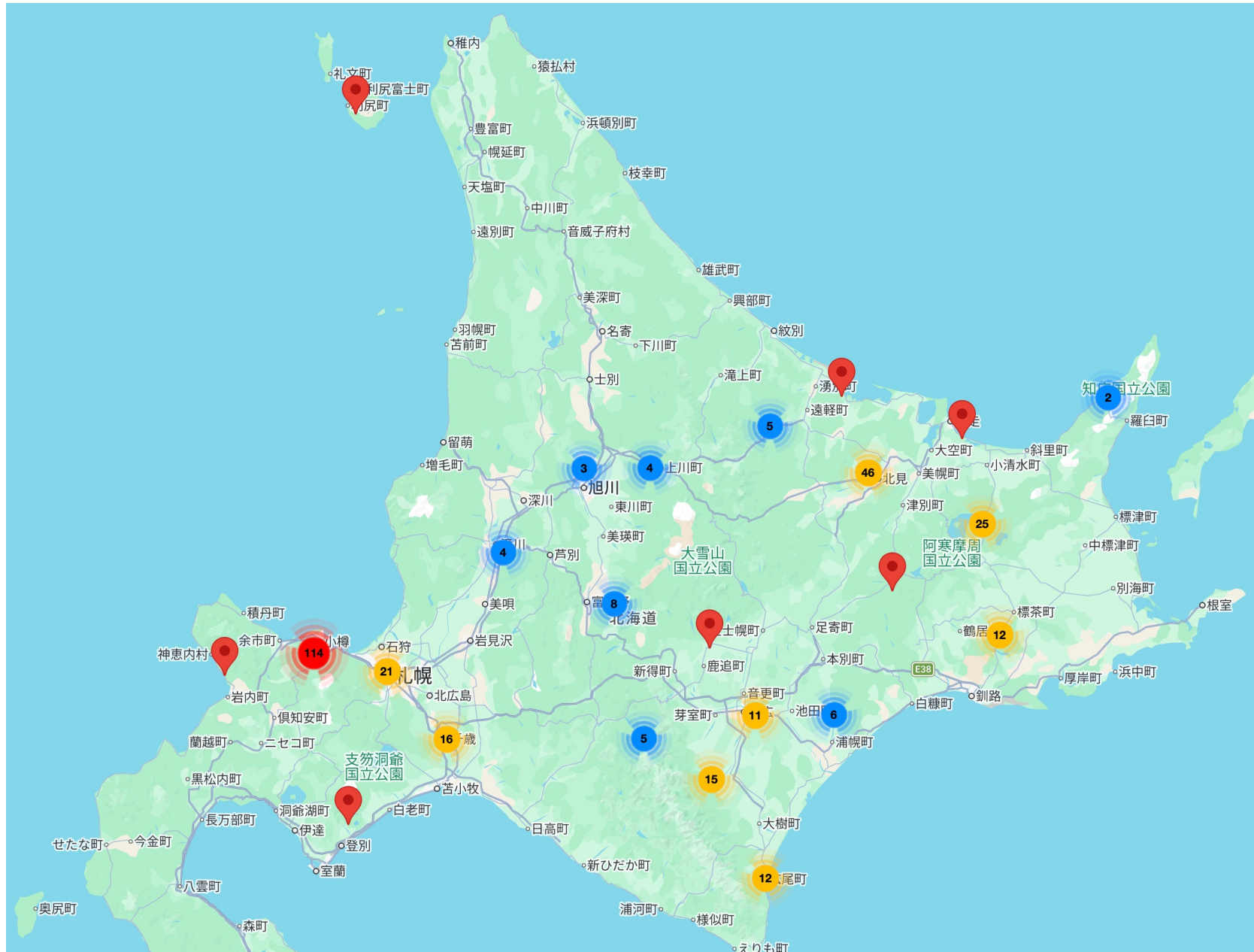
該当件数: 4,486 件 (1-20件目)

20 件ずつ表示 ソート順 ダウンロード 絞り込み検索

1 2 3 4 5 6 >>

No	学名	和名	記録年月日 (始め)	国 (日本語)	都道府県 (日本語)	機関名 (日本語)
1	Papilio maackii maackii Menetries, 1859	ミヤマカラスアゲハ・名義タ イブ亜種	*****			姫路科学館
2	Papilio maackii maackii MENETRIES, 1858	ミヤマカラスアゲハ	198005**	日本	埼玉県	岩手県立博物館
3	Papilio maackii maackii MENETRIES, 1858	ミヤマカラスアゲハ	19780611	日本	秋田県	岩手県立博物館
4	Papilio maackii maackii MENETRIES, 1858	ミヤマカラスアゲハ	19870421	日本	北海道	岩手県立博物館
5	Papilio maackii maackii Menetries, 1858	ミヤマカラスアゲハ	19810626	日本	岩手県	岩手県立博物館
6	Papilio maackii maackii Menetries, 1858	ミヤマカラスアゲハ	19780723	日本	岩手県	岩手県立博物館

北海道のミヤマカラスアゲハマップ



地図出典: Google Map (©Google)

S-Netのデータのもとになるのは…

各参加機関が所蔵する生物標本！

地名ラベル・コレクションラベル

北海道根室市東梅	地名(日本語)
Toubai, Nemuro-shi	地名(英語)
Hokkaido, JAPAN	緯度経度情報
43.2549°N, 145.4538°E	緯度経度情報
4. VIII. 2013, U. Jinbo	採集日
	採集者

地名ラベルの例

Lepidoptera Collection
NSMT-I-L-30001
National Museum of Nature and Science, Tokyo

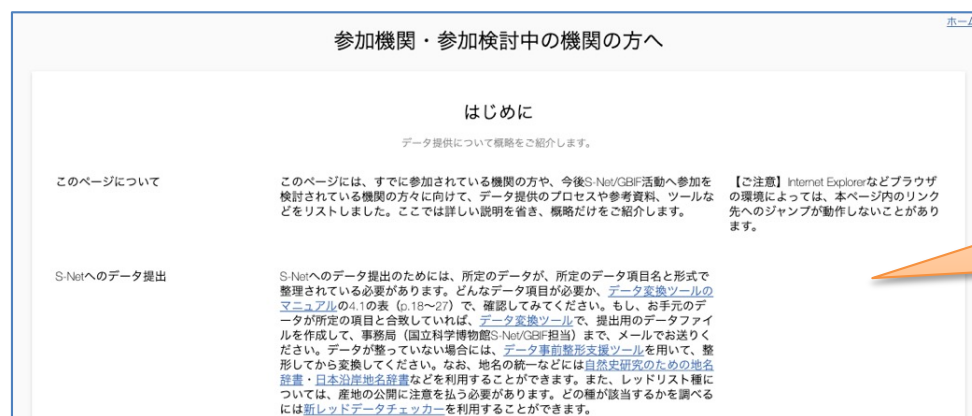
コレクションラベルの例

NSMT-I-L-30001

機関の略号 コレクションの略号 標本番号

S-Netの標本データ共有

1. 参加機関の担当者は、S-Netが提供するデータ変換ツールを利用して、自機関のデータをS-Net形式に変換
2. 参加機関の担当者は、変換後のデータをS-Net事務局へ送付
3. S-Net事務局は、最終チェック後にデータをS-Netポータルウェブサイトより公開



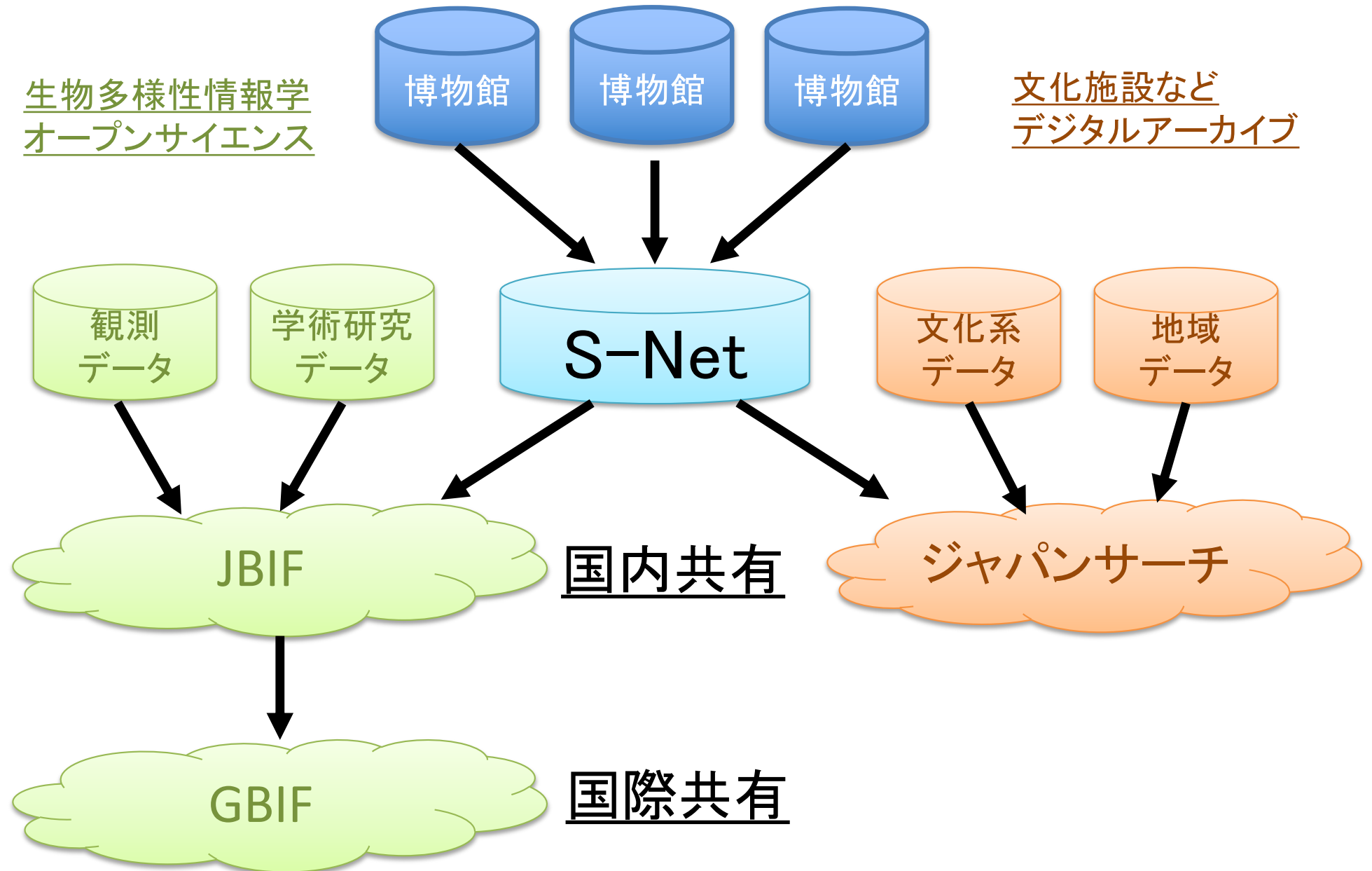
詳細はS-Netウェブサイト
サイトで確認できます

S-Net形式とは？

- S-Netのデータ形式
- 標本に関する様々な項目を用意(178項目)
 - 標本番号
 - 種名・分類
 - 採集地・日付
 - 地質年代(化石)
 - タイプ標本情報
- 多くの項目が日本語と英語で記入可能

9.	機関名 ※S-Net サイトの機関情報を確認し、同じ英語の名称を大文字、小文字の区別を含め正しく指定してください。	○	文字列・半	機関名の英文名称 (例: National Museum of Nature and Science)
10.	機関名 (日本語) ※S-Net サイトの機関情報を確認し、同じ和文の名称を正しく指定してください。	○	文字列・日	機関名の和文名称 (例: 国立科学博物館 (植物))
11.	機関コード ※S-Net サイトの機関情報を確認し、同じ機関コードを大文字、小文字の区別を含め正しく指定してください。	○	文字列・半	標本の機関コード (例: TNS, NMST, KPM)
12.	コレクションコード ※S-Net サイトのデータセット情報を確認し、同じコレクションコードを大文字、小文字の区別を含め正しく指定してください。	○	文字列・半	標本のコレクションコード (例: VS, F, AL)。コレクションコードが未設定の機関では、機関コードを入れる
13.	カタログ番号	○	文字列・半	標本番号 (例: 153801, A1-0027)
オカレンス情報				
14.	採集者番号		文字列・半	採集者によるオリジナルの標本番号
15.	オカレンス備考		文字列・半	標本等に関する補足説明 (例: found dead on the road)
16.	オカレンス備考 (日本語)		文字列・日	標本等に関する補足説明 (例: 路上轍死体)
17.	性別		文字列・半	例: male, female
18.	性別 (日本語)		文字列・日	例: オス、メス、♂、♀
19.	生活型・世代型		文字列・半	個体のライフステージ (例: juvenile, adult, sporophyte)
20.	生活型・世代型 (日本語)		文字列・日	例: 幼虫、成虫、胞子体
21.	成熟状況		文字列・半	例: pregnant, in bloom, fruit-bearing
22.	成熟状況 (日本語)		文字列・日	例: 妊娠中、開花中、結実中
23.	行動		文字列・半	採集時の個体の振る舞い (例: roosting, foraging, running)
24.	行動 (日本語)		文字列・日	例: ねぐらに滞在、摂食中、走っていた
25.	生息環境		文字列・半	例: oak savanna, pre-cordilleran steppe
26.	生息環境 (日本語)		文字列・日	例: ブナ林、海岸林
27.	個体群成立過程		文字列・半	例: native, introduced, invasive, naturalised, managed
28.	個体群成立過程 (日本語)		文字列・日	例: 土着、外来、侵略的外来、帰化、栽培、植栽
29.	処理・保存情報		文字列・半	例: fossil, cast, skin, spirit collection
30.	処理・保存情報 (日本語)		文字列・日	例: 化石、皮革、骨格、液浸
31.	カタログ旧番号・別番号		文字列・半	標本番号をつけ直す場合、以前の標本番号
32.	画像への外部リンク		文字列・半	ウェブ公開されている当該標本の画像自体へのリンク (URL)。直接リンク可能な場合のみ入力

S-Netがつなぐ世界



データクリーニングとは？

「データクレンジング」とは？

データを「きれいにする」ことで
品質を高め利用しやすくすること

「きれいにする」とは？

誤字・表記や形式の揺れを統一させること

形式的な表記揺れ

Lepidoptera Science 70: 99–148

Lepidoptera Science 70:99–148

Lepidoptera Science 70: 99–148

Lepidoptera Science 70: 99–148.

Lepidoptera Science, 70: 99–148

Lepidoptera Science 70: 99–148

Lepidptera Science 70: 99–148

Lepidoptera Science 70: 99–148

Lepidoptera Science 70: 99–148

Lepid. Sci. 70: 99-148

Lepid Sci 70: 99-148

Lep. Sci. 70: 99-148

- 記号の使い方
- 空白の入れ方
- 誤記
- 略称

等々...

ファセット表示で見る表記揺れ

国 (日本語)

日本 (6,127)

台湾 (95)

[もっと見る](#)

「日本」「日本国」「Japan」
「中国」「中華人民共和国」

国 (日本語)

日本国 (88) - (25) Japan (14) 中華人民共和国 (14) 中国 (13) タイ王国 (12) インド共和国 (9) チェコ共和国 (8) 韓国 (8) ロシア (7) タイ (6) ネパール連邦民主共和国 (6) マレーシア (6) ミャンマー (5) ネパール (4) ドイツ連邦共和国 (3) ロシア連邦 (3) 朝鮮民主主義人民共和国 (3) アフガニスタン (2) アメリカ (2) キルギス (2) スリランカ (2) イギリス (1) インド (1) カナダ (1) スイス連邦 (1) スロバキア共和国 (1) デンマーク王国 (1) ニュージーランド (1) フィリピン共和国 (1) ベトナム (1) ラオス (1) 日本? (1)

都道府県 (日本語)

岩手県 (875)

北海道 (758)

「都道府県」をつけるかつけないか？

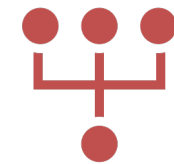
都道府県 (日本語)

沖縄県 (495) 栃木県 (339) 茨城県 (332) 宮城県 (309) 徳島県 (287) 鹿児島県 (259) 長崎県 (228) 千葉県 (226) 神奈川県 (209) 滋賀県 (198) 兵庫県 (189) 愛知県 (170) 石川県 (162) 埼玉県 (154) 群馬県 (150) 静岡県 (122) 富山県 (89) 東京 (86) 青森県 (82) 三重県 (70) 東京都 (70) 長野県 (70) 大阪府 (63) 山形県 (50) 新潟県 (47) 山梨県 (42) 福井県 (31) 香川県 (29) 福島県 (28) 千葉 (25) 岐阜県 (24) 秋田県 (22) 南投県 (16) 四川省 (15) 山口県 (14) 奈良県 (11) 鳥取県 (10) 兵庫 (8) 神奈川 (7) ガンダキ・プラデーシュ州 (6) ズリーン地方 (6) 広島県 (6) カチン州 (5) 京都府 (5) 河南省 (5) チェンマイ (4) 北京市 (4) 群馬 (4) ナン州 (3) ヒマーチャル・プラデーシュ州 (3) ラダック連邦直轄領 (3) 佐賀県 (3) 慶尚北道 (3) 高知県 (3) 高雄県 (3) ジャンムー・カシミール蓮舫直轄領 (2) ハワイ州 (2) バンコク (2) 和歌山県 (2) 基隆市 (2) 宮崎県 (2) 岡山県 (2) 島根県 (2) 彰化県 (2) 愛媛県 (2) 新北市 (2) 新疆ウイグル自治区 (2) 福岡県 (2) 香港 (2) 香港特別行政区 (2) 鹿児島 (2) no data (1) パラワン州 (1) ブリティッシュコロンビア州 (1) ホークスベイ地方 (1) メガラヤ州 (1) ラムドン省 (1) ラムプーン県メーター郡 (1) ヴァレー州 (1) 南モラヴィア州 ホドニーン地区 (1) 台中県 (1) 台湾 (1) 広東省 (1) 新竹県 (1) 海南省 (1) 熊本県 (1) 西ベンガル州 (1) 雲南省 (1)

なぜ困る？

S-Netには、何カ国・地域のモンシロチョウの標本が登録されているのか？

- 同じ国・地域の表記が統一していなければ、正しい答えに行き着けない。検索もうまくいかない
- S-Netでできるだけ統一されていれば、利用者の手間を減らせ、検索時の漏れも減らせる



表記揺れあれこれ

- 形式的な違い 「日本」「日本国」
「チョウ目」「チョウ」
- 正式名と通称 「中華人民共和国」「中国」
- 典拠の違い 「チョウ目」「鱗翅目」
- 追加情報 「チョウ目(鱗翅目)」「チョウ目チョウ類」
- 誤記 「チョウ目」

目名 (日本語名)

チョウ目 (4,186)

鱗翅目 (1,120)

目名 (日本語名)

[チョウ \(501\)](#) [チョウ目チョウ類 \(210\)](#) [チョウ目 \(鱗翅目\) \(94\)](#)

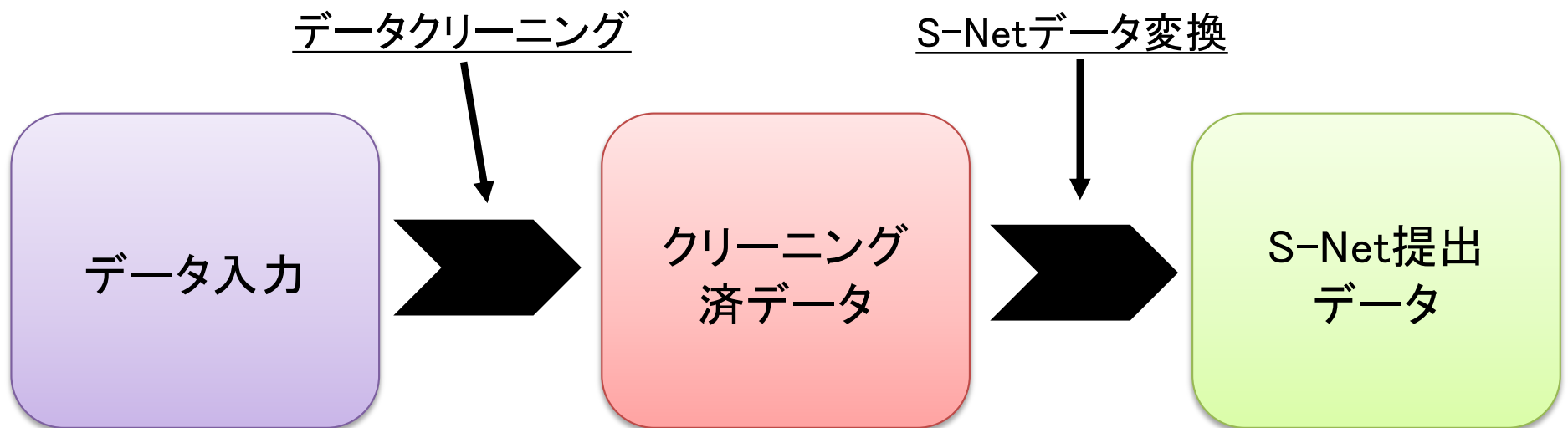
(2021年時点の結果)

目名 (日本語名)

[チョウ \(555\)](#) [チョウ目チョウ類 \(210\)](#) [チョウ目 \(鱗翅目\) \(92\)](#) [チョウ目 \(29\)](#) [チョウ* \(1\)](#)

いつクリーニングするのか？

- S-Net変換前(マスターデータ)作成時
- あとから見つかった修正が必要な箇所は、クリーニング済データに反映させたあと、再度クリーニングを実施
- S-Net変換ツールにも簡易クリーニング機能あり



どんなことをするのか？

- 形式の統一
 - 全角・半角
 - ひらがな・カタカナ、大文字・小文字
 - 余計な空白や記号の除去
- 表記の統一
 - 同じデータの書き方を統一する(名寄せ)
 - 対応するデータを同じように入れる
(例: 和名から学名を入力)

データクリーニングのツール

- 表計算ソフトウェア
 - Microsoft Excel, Google Spreadsheetなど
- データクリーニングソフトウェア
 - Open Refine
- 専用のプログラム・ウェブ上のツール
 - 事前整形支援ツール・変換ツール
 - レッドデータチェッカー
 - GBIFの学名チェッカー
- テキストエディタ
- プログラム(マクロ・スクリプト)による自動化