



サイエンスミュージアムネット
日本語標本情報検索・GBIF情報検索システム

標本データ変換ツール 操作説明書

第3.0版
rev. 120306001

事前整形支援ツール抜粋版

2012年03月09日

独立行政法人 国立科学博物館

目次

1.	はじめに		
1.1	標本データ変換ツールについて	2
1.2	処理の流れ	3
1.3	読込可能な標本データファイル形式	5
2.	STEP1(事前整形補助ツール)	6
2.1	住所の和英、英和変換	9
2.2	住所情報の抽出・分割	11
2.3	高度・日付情報の抽出	13
2.4	幅のある日付の分割	15
2.5	アドレスマッチング用ファイル出力	17
3.	STEP2(メインツールによるマッピング)		
3.1	博物館名入力	20
3.2	機能選択(学名変換機能)	21
3.3	ファイル読込	22
3.4	項目マッピング	24
3.5	項目マッピング画面	25
3.6	項目変換	32
3.7	項目チェック	33
3.8	学名(採集、その他)ファイル出力	35
3.9	機能選択(採集変換機能)	37
3.10	機能選択(その他変換機能)	38
4.	STEP3(メインツールによるファイル合成)		
4.1	機能選択(3ファイル合成機能)	39
4.2	ファイル読込	40
4.3	項目チェック	42
4.4	登録用ファイル出力	44
	付録		
1.	高度と日付の抽出パターン	46
2.	メインツールインストール方法	47
3.	博物館情報の設定	49
4.	登録用ファイル解説	50
5.	登録用ファイルフォーマット	63

STEP1:(任意作業)

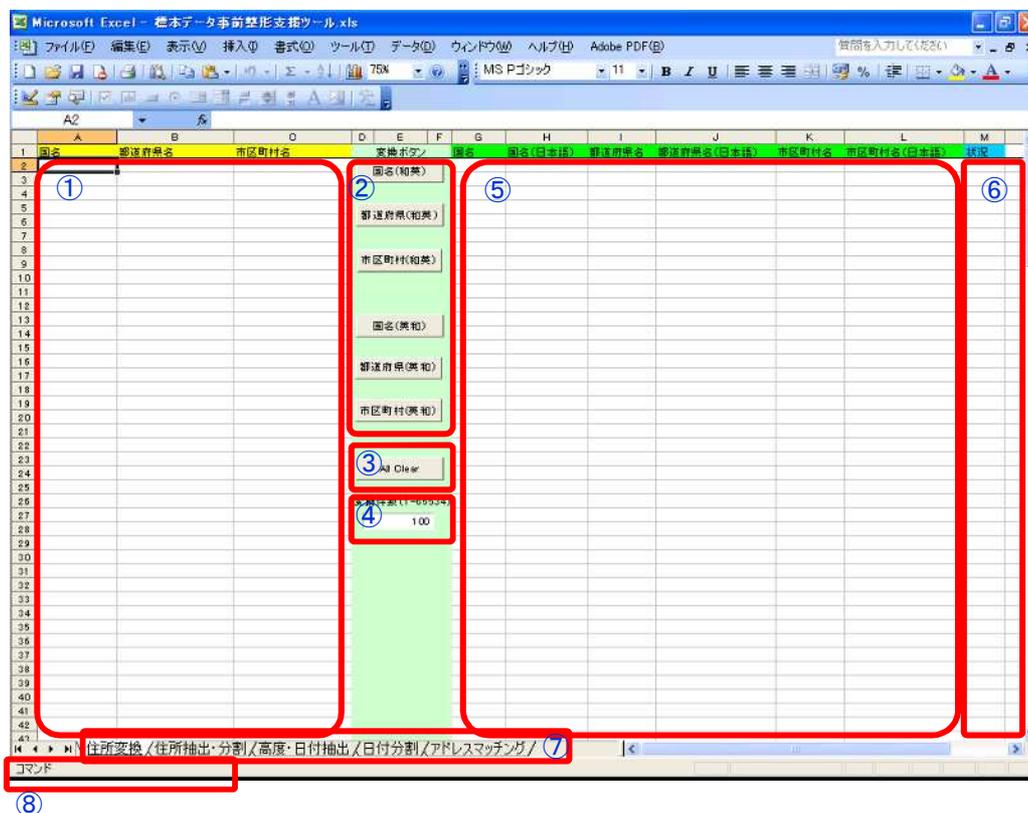
各博物館の標本データファイルを、事前整形支援ツール(Excelマクロ)を用いて整形を行います。

- ・地名の和英/英和変換
- ・地名の分割/抽出
- ・高度や日付の抽出
- ・幅のある日付の分割
- ・日本国内地名から緯度経度情報に変換するアドレスマッチングの作業を必要に応じて行います。

2. STEP1

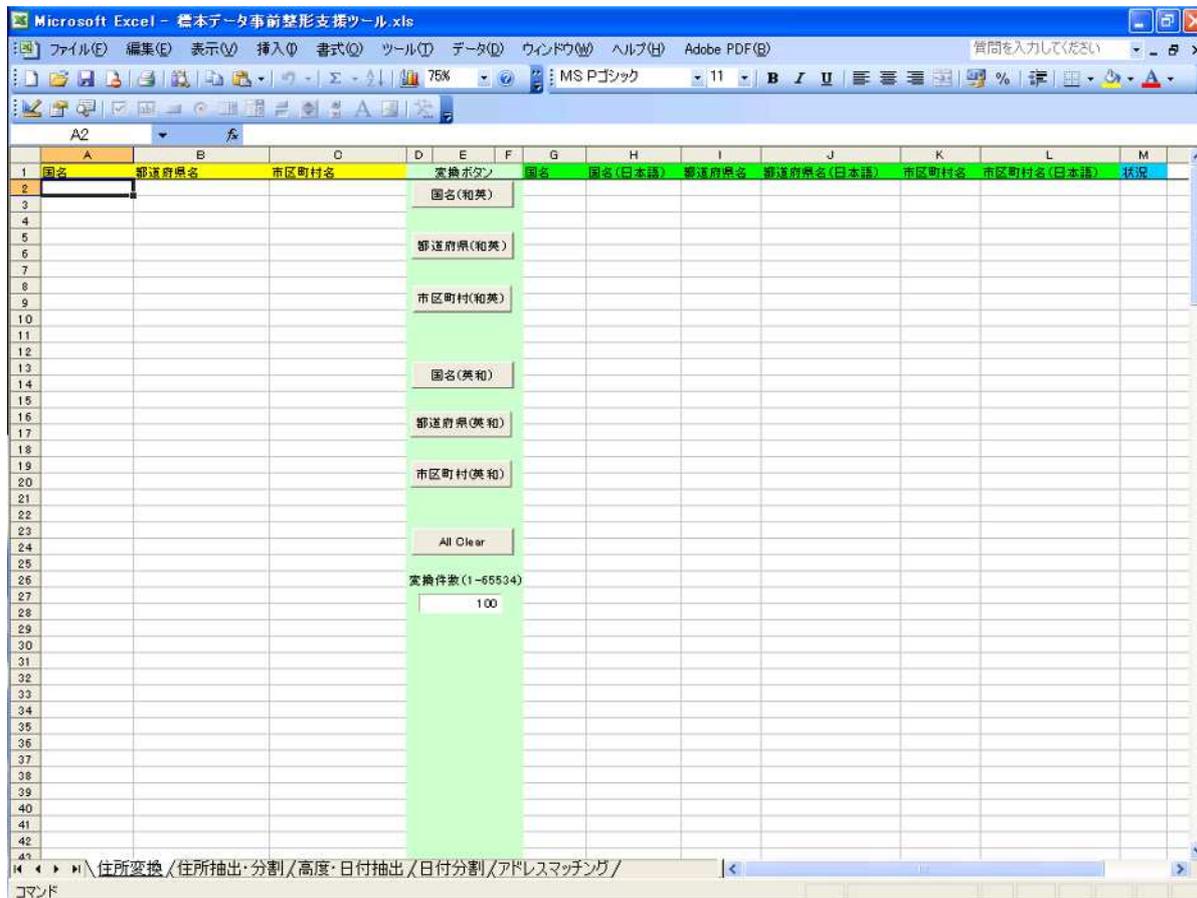
STEP1では、Excelマクロ機能を用いていろいろな情報を整形する作業を行います。必要な機能を選択して使用してください。

<画面構成>



No.	名称	説明
①	変換前フィールド	手持ちのデータを貼り付けるフィールドです。
②	変換ボタン	様々な変換ができるボタンがあるところです。
③	All Clearボタン	シート上のデータを全て削除します。
④	変換件数入力	変換したい件数を入力します。
⑤	変換後フィールド	変換されて整形された項目が表示されるフィールドです。
⑥	状況フィールド	変換結果の状況が表示されるフィールドです。
⑦	機能選択シート	5つの機能から使用するシートを選択します。
⑧	ステータスバー	現在変換中の件数が表示される領域です。

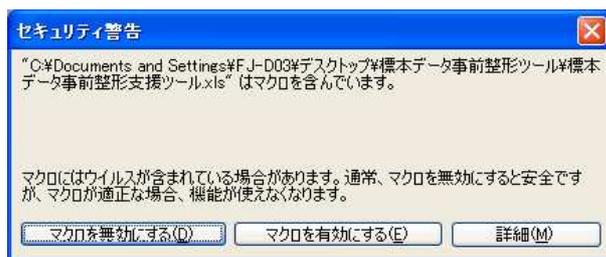
<基本的なフロー>



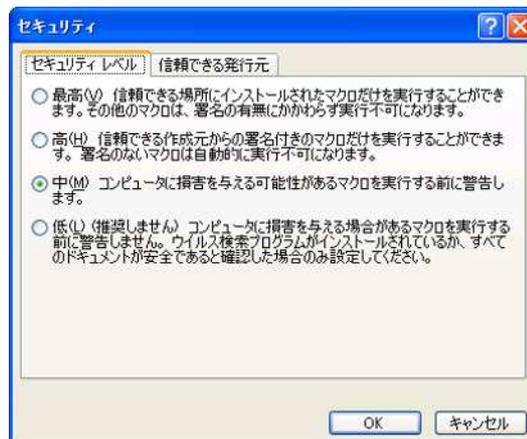
- 1) ダウンロードした”specdatapreconv*****.zip”を解凍します。(*****にはバージョンにより異なります。) Excelファイルが2つ入ったフォルダができますが、**ファイルを別のフォルダに移動しないでください。**
- 2) 標本データ事前整形支援ツール.xlsを開きます。左下のような警告画面が出たら、「マクロを有効にする」ボタンを押してください。

うまく開くことができない場合

- a. Excelのツールバーの「ツール → マクロ → セキュリティ」を選択する。(右下の画面がでます)
- b. セキュリティレベルの「中」を選んでOKを押します。
- c. 一旦Excelを終了して再度標本データ事前整形支援ツール.xlsを開いてください。

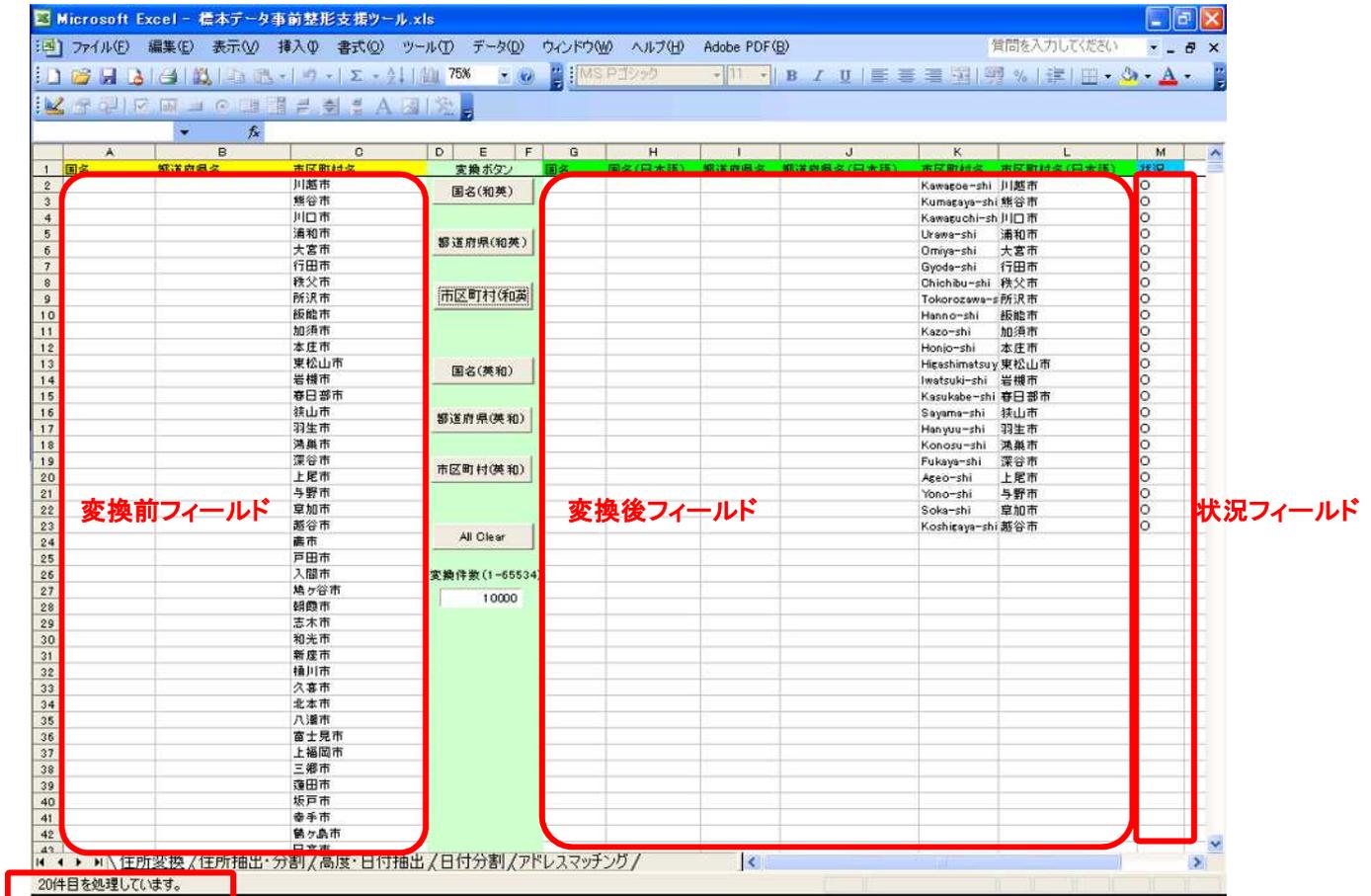


セキュリティ警告ダイアログ



Excelマクロのセキュリティ設定画面

<基本的なフロー(続き)>

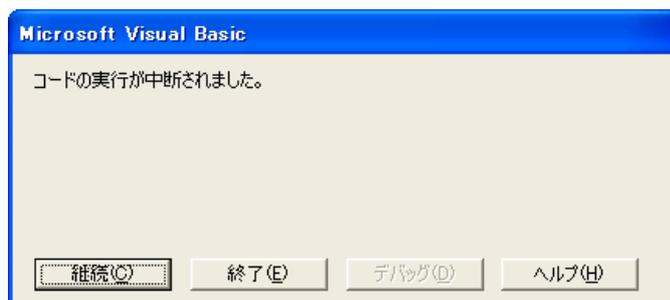


ステータスバー

- 3) もととの標本データ(CSV形式)をExcelで開いて、変換作業を行いたい行を変換前フィールドに貼り付けます。
- 4) 変換する件数を入力します。
- 5) 変換ボタンを押します。ステータスバーに変換中の件数が表示されていることを確認してください。
- 6) 変換後フィールドに変換された結果が表示されていきます。左下のメッセージが表示されたら完了です。
 完了前に強制的に終了したい場合
 a. Escボタンを押します。
 b. 右下の画面が表示されたら終了を押すと強制終了します。(継続を押すと処理を再開します。)
- 7) 変換された結果を確認したら、その行をもとのCSVファイルに追加して保存してください。
 追加保存したCSVファイルをSTEP2のメインツールで使用します。



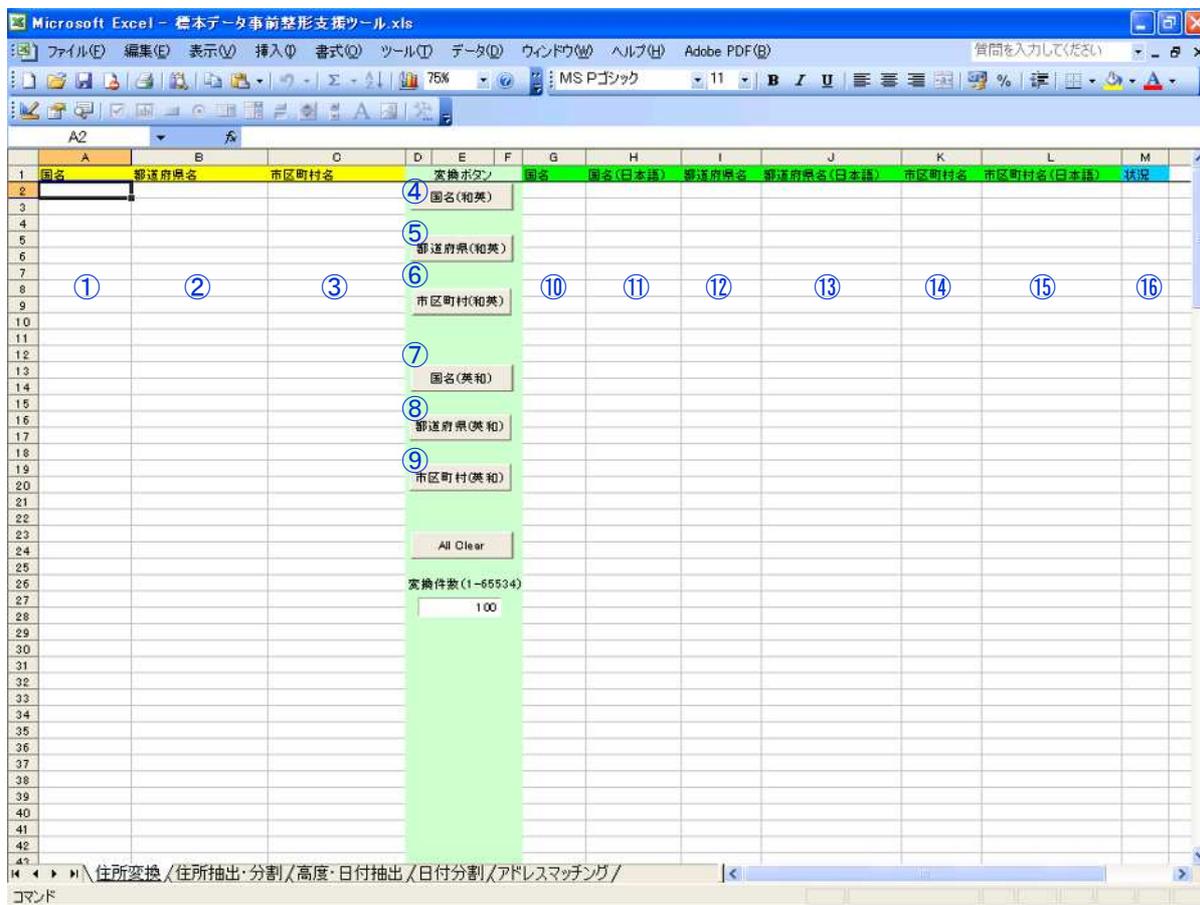
処理完了時ダイアログ



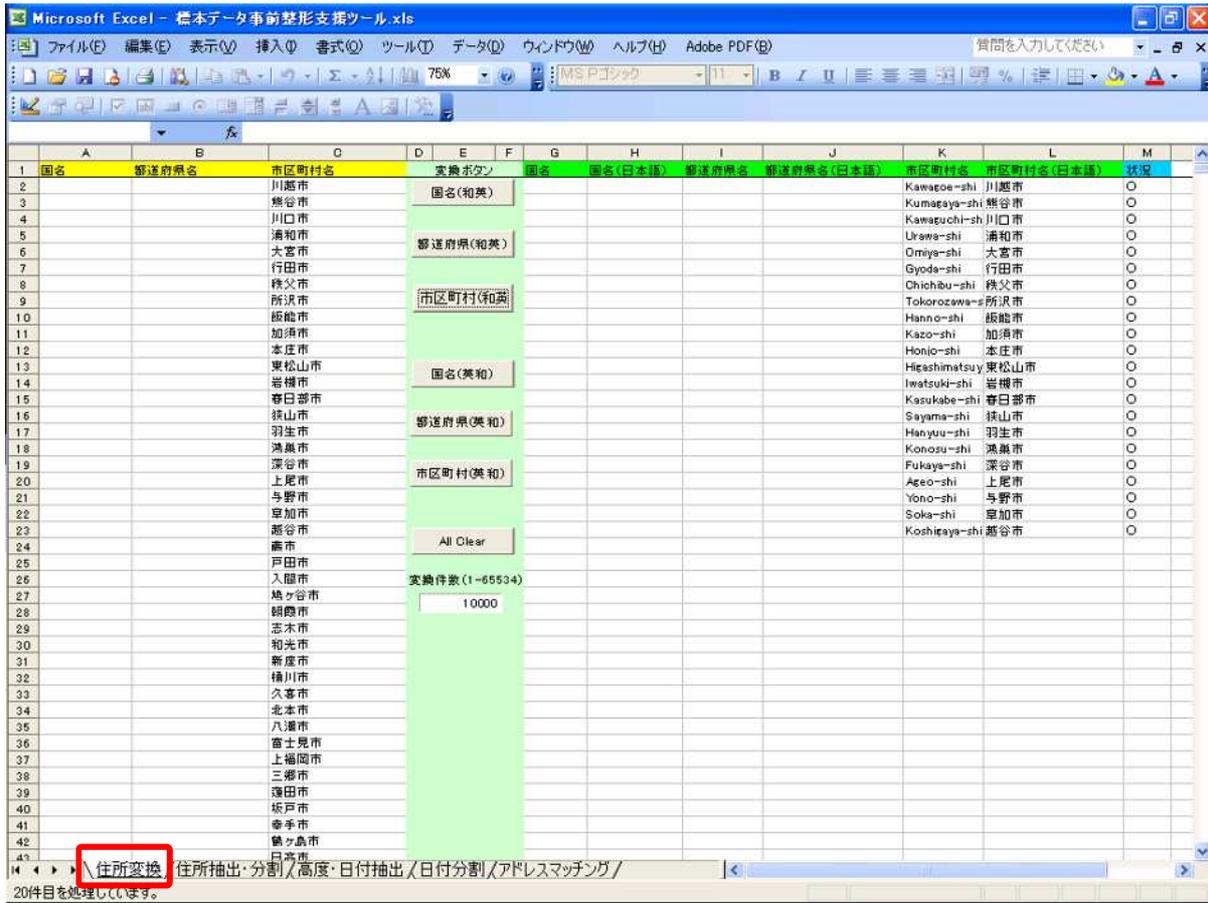
Escボタン押下時の強制終了ダイアログ

2.1 住所の和英、英和変換

ここでは、日本語の住所情報から英語の住所情報を、もしくはその逆の変換作業を行います。



No.	名称	説明
①	国名フィールド(変換前)	元の国名データを貼り付けるフィールドです。
②	都道府県フィールド(変換前)	元の都道府県データを貼り付けるフィールドです。
③	市区町村フィールド(変換前)	元の市区町村データを貼り付けるフィールドです。
④	国名和英変換ボタン	国名を日本語から英語に変換するときのボタンです。
⑤	都道府県和英変換ボタン	都道府県を日本語から英語に変換するときのボタンです。
⑥	市区町村和英変換ボタン	市区町村を日本語から英語に変換するときのボタンです。
⑦	国名英和変換ボタン	国名を英語から日本語に変換するときのボタンです。
⑧	都道府県英和変換ボタン	都道府県を英語から日本語に変換するときのボタンです。
⑨	市区町村英和変換ボタン	市区町村を英語から日本語に変換するときのボタンです。
⑩	国名フィールド(英語・変換後)	変換後の国名(英語)が表示されるフィールドです。
⑪	国名フィールド(日本語・変換後)	変換後の国名(日本語)が表示されるフィールドです。
⑫	都道府県フィールド(英語・変換後)	変換後の都道府県(英語)が表示されるフィールドです。
⑬	都道府県フィールド(日本語・変換後)	変換後の都道府県(日本語)が表示されるフィールドです。
⑭	市区町村フィールド(英語・変換後)	変換後の市区町村(英語)が表示されるフィールドです。
⑮	市区町村フィールド(日本語・変換後)	変換後の市区町村(日本語)が表示されるフィールドです。
⑯	状況フィールド	変換結果の状況が表示されるフィールドです。



- 1) 「住所変換」のシートを選択します。
- 2) もととの標本データのうち、国名か都道府県か市区町村を貼り付けます。(英語もしくは日本語)
- 3) 変換したい内容にあったボタンを押します。変換後フィールドに変換後の項目が表示されます。変換後の行には日本語、英語両方が表示されます。
- 4) 変換された内容を、状況フィールドを参考にしながら確認してください。元のデータとの照合はExcel行番号を照らし合わせながら行うことができます。

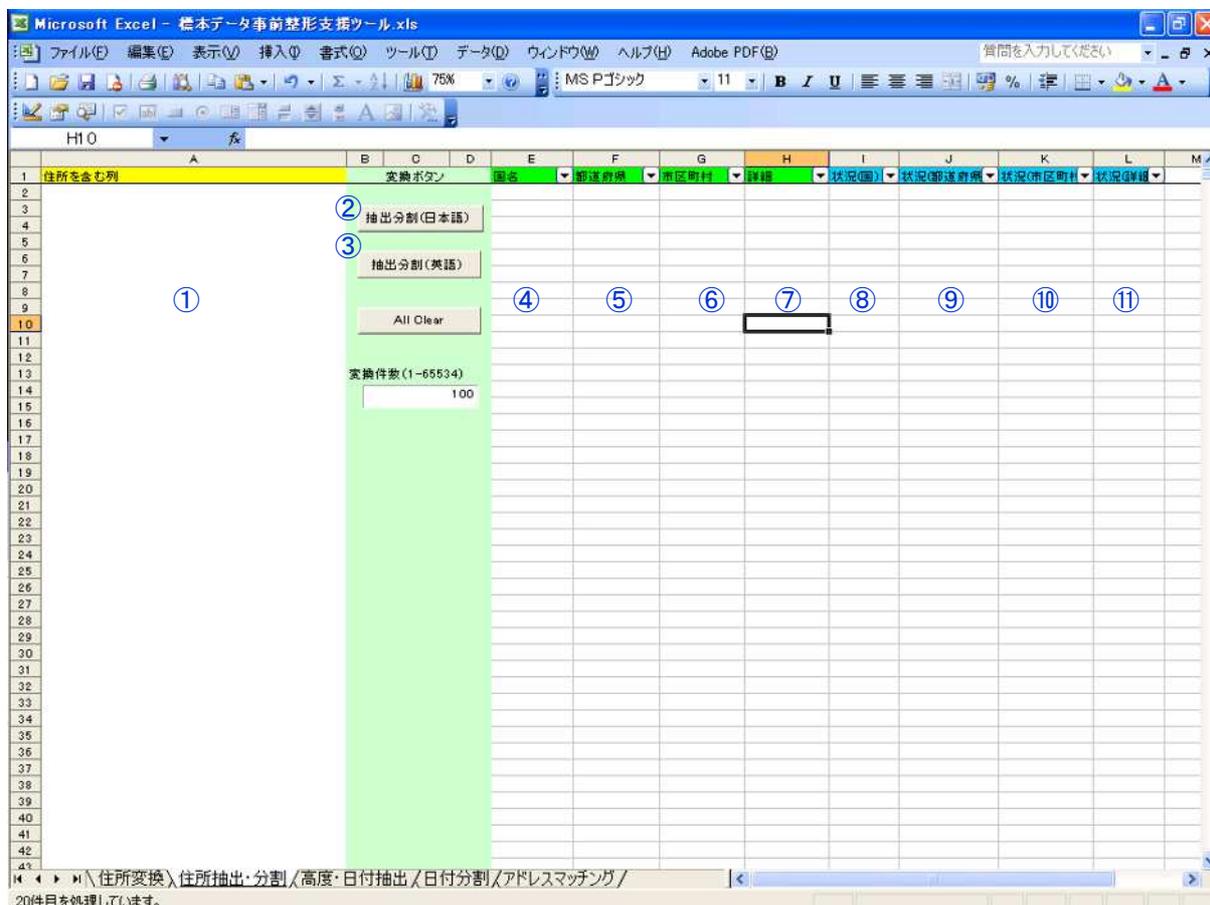
「住所変換機能」での状況フィールドについて

- ・O → 正しく変換が完了している項目です。
- ・不可 → 住所マスターデータにないために変換ができなかった項目です。
- ・空欄 → もともと空欄だった項目です。

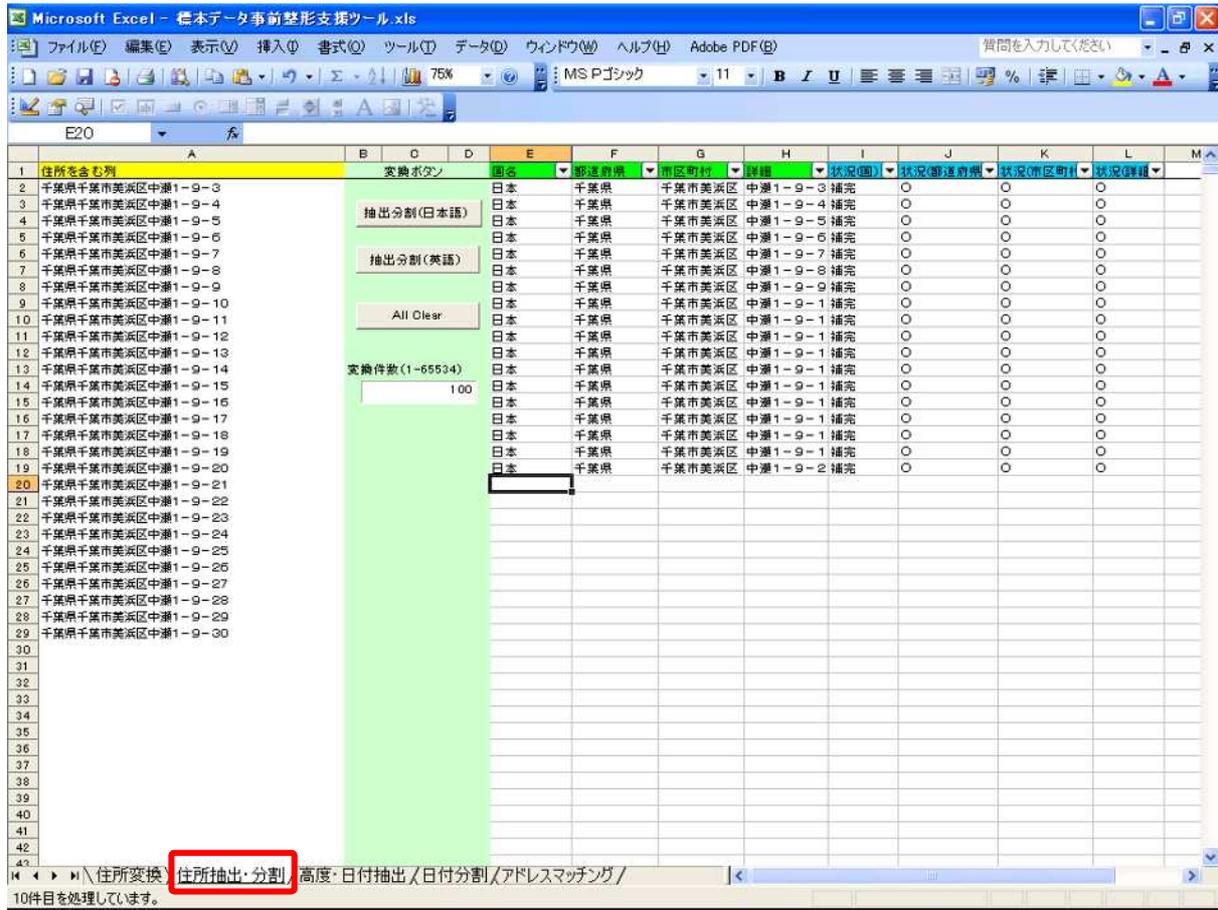
※市区町村の変換は一般的な性能のパソコンにおいて、1万件あたり30分～1時間程度の時間を要します。

2.2 住所の抽出・分割

ここでは、一つの項目に含まれてしまっている住所情報を国名、都道府県、市区町村、詳細に抽出・分割をする作業を行います。



No.	名称	説明
①	住所を含む列フィールド(変換前)	元の住所を含んでいるデータを貼り付けるフィールドです。
②	抽出分割(日本語)ボタン	元の住所が日本語のときに抽出・分割をするときのボタンです。
③	抽出分割(英語)ボタン	元の住所が英語のときに抽出・分割をするときのボタンです。
④	国名フィールド(抽出後)	抽出後の国名が表示されるフィールドです。
⑤	都道府県フィールド(抽出後)	抽出後の都道府県が表示されるフィールドです。
⑥	市区町村フィールド(抽出後)	抽出後の市区町村が表示されるフィールドです。
⑦	詳細フィールド(抽出後)	抽出後の住所の詳細が表示されるフィールドです。
⑧	国名状況フィールド	国名抽出結果の状況が表示されるフィールドです。
⑨	都道府県状況フィールド	都道府県抽出結果の状況が表示されるフィールドです。
⑩	市区町村状況フィールド	市区町村抽出結果の状況が表示されるフィールドです。
⑪	詳細状況フィールド	住所の詳細抽出結果の状況が表示されるフィールドです。



- 1) 「住所抽出・分割」のシートを選択します。
- 2) もとからの標本データのうち、住所情報を含む列を貼り付けます。(英語もしくは日本語)
- 3) 日本語であれば「抽出・分割(日本語)」ボタンを、英語であれば「抽出・分割(英語)」ボタンを押します。
- 4) 変換された内容を、状況フィールドを参考にしながら確認してください。
元のデータとの照合はExcel行番号を照らし合わせながら行うことができます。

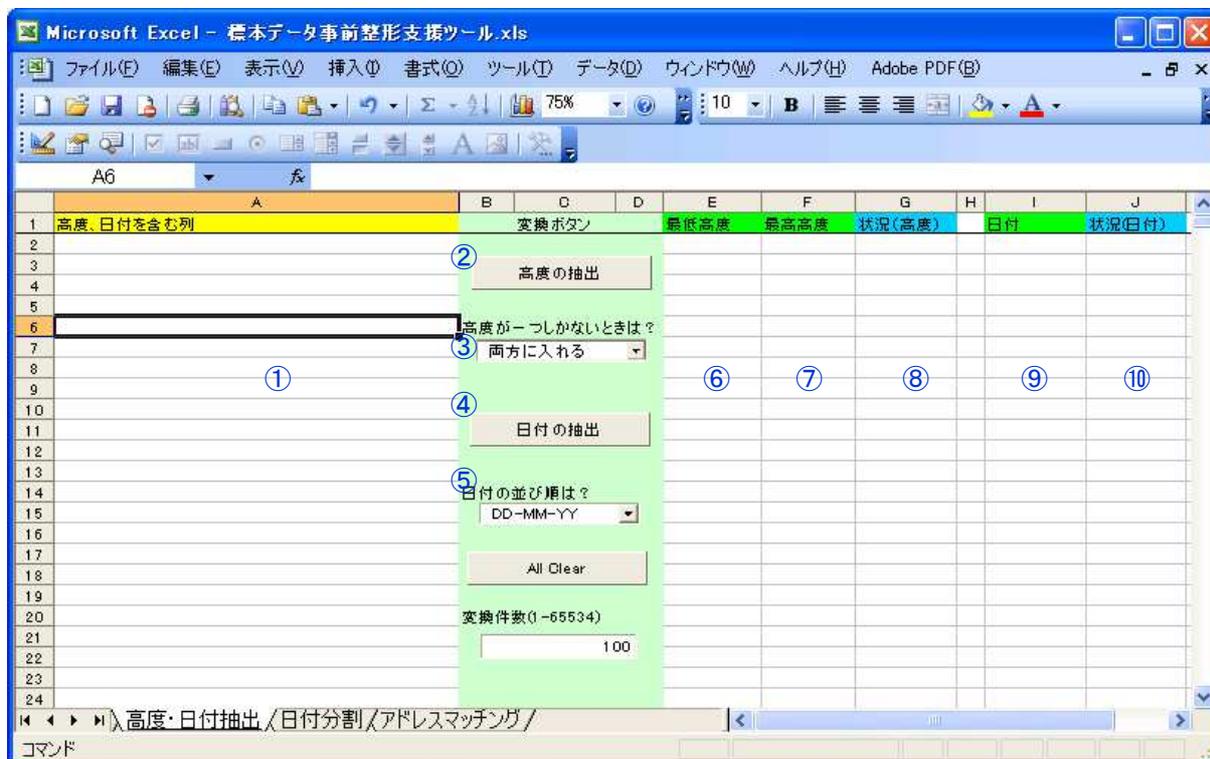
「住所抽出・分割機能」での状況フィールドについて

- ・○ → 正しく変換が完了している項目です。
- ・補完 → 抽出されなかったが、下位の住所情報から上位の住所情報を補完した項目です。
都道府県から国名(日本)を、市区町村から都道府県を補完しています。
- ・上書き → 抽出したが、下位の分類によって改めて変更しなおした項目です。
- ・警告 → 抽出はされているが、同一名称の市区町村などで変換が間違っている可能性のある項目です。確認をして間違いがあれば手で修正してください。
- ・△ → 都道府県において、正しい抽出ができていない可能性がありますので、確認をして間違いがあれば手で修正してください。
- ・不可 → 住所マスターデータにないために抽出ができなかった項目です。
- ・空欄 → もともと空欄だった項目です。

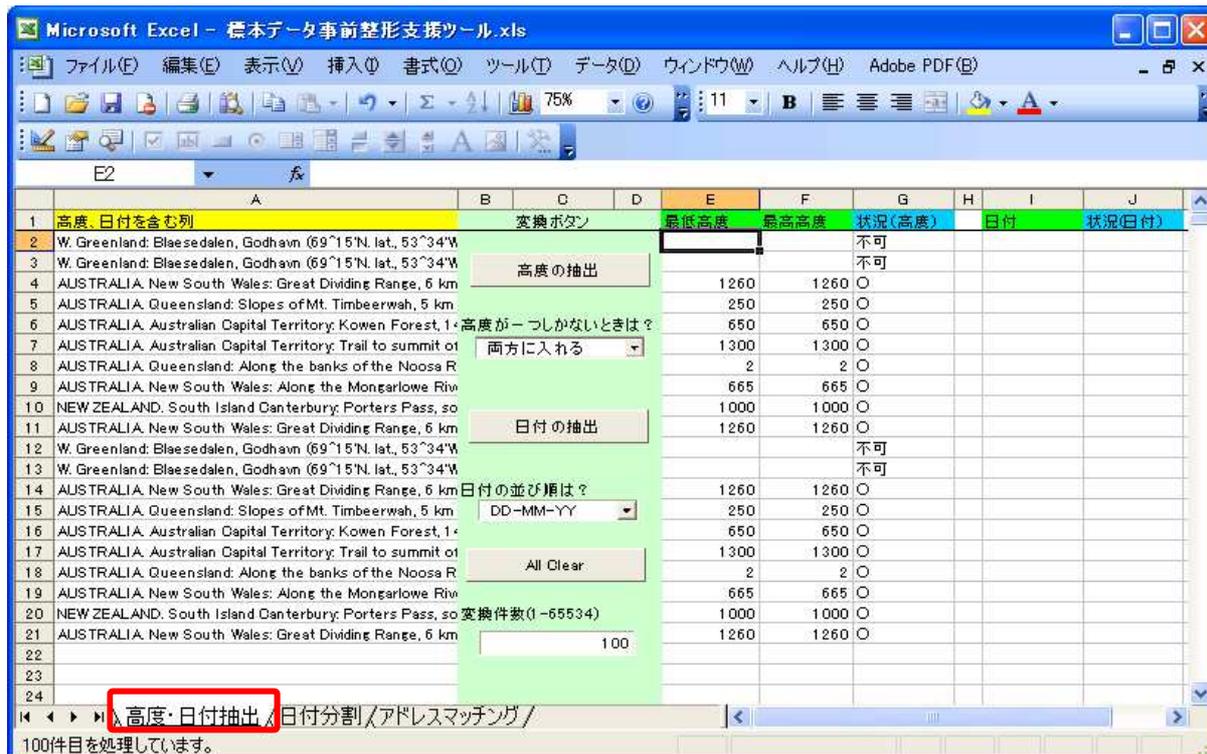
※変換は一般的な性能のパソコンにおいて、1万件あたり30分～1時間程度の時間を要します。

2.3 高度・日付の抽出

ここでは、一つの項目に含まれてしまっている高度・日付情報を数値や8桁表記の日付に抽出する作業を行います。



No.	名称	説明
①	高度や日付を含む列フィールド	高度や日付を含んでいるデータを貼り付けるフィールドです。
②	高度の抽出ボタン	高度の抽出・分割をするときのボタンです。
③	高度の抽出仕様選択	抽出時に高度情報が1つしかない場合に、最高高度に入れるか、最低高度に入れるか、両方に入れるか選択できます。
④	日付の抽出ボタン	日付の抽出・変換をするときのボタンです。
⑤	日付の抽出仕様選択	元のデータの日付の順番が、年-月-日の順か月-日-年の順か、日-月-年の順番かを選択します。
⑥	最低高度フィールド	抽出後の最低高度が表示されるフィールドです。
⑦	最高高度フィールド	抽出後の最高高度が表示されるフィールドです。
⑧	高度状況フィールド	高度抽出結果の状況が表示されるフィールドです。
⑨	日付フィールド	抽出後、8桁表示に変換された日付が表示されるフィールドです。
⑩	日付状況フィールド	日付抽出結果の状況が表示されるフィールドです。



- 1) 「高度・日付抽出」のシートを選択します。
- 2) もとものの標本データのうち、高度や日付を含む列を貼り付けます。
- 3) 高度であれば「高度が一つしかないときの処理」を選択し、「高度の抽出」ボタンを押します。
日付であれば「日付の並び順」を選択し、「日付の抽出」ボタンを押します。
- 4) 変換された内容を、状況フィールドを参考にしながら確認してください。
元のデータとの照合はExcel行番号を照らし合わせながら行うことができます。

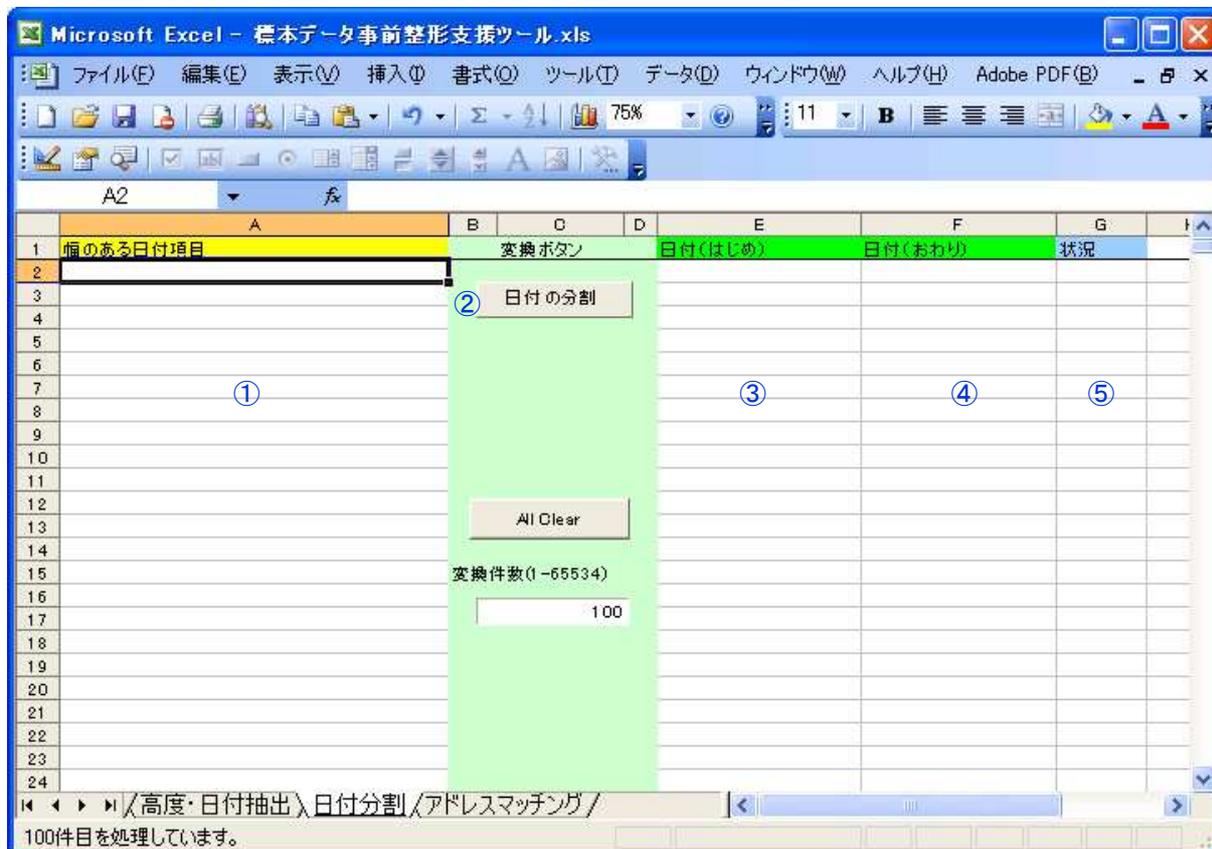
「高度・日付抽出機能」での状況フィールドについて

- ・○ → 正しく抽出が完了している項目です。
- ・警告 → 抽出はされているが、その後にまだ高度(日付)情報が存在する可能性のある項目です。
確認をして誤抽出があれば手で修正してください。
- ・不可 → 存在しない、もしくは抽出パターンに合わないため抽出できなかった項目です。
- ・空欄 → もともと空欄だった項目です。

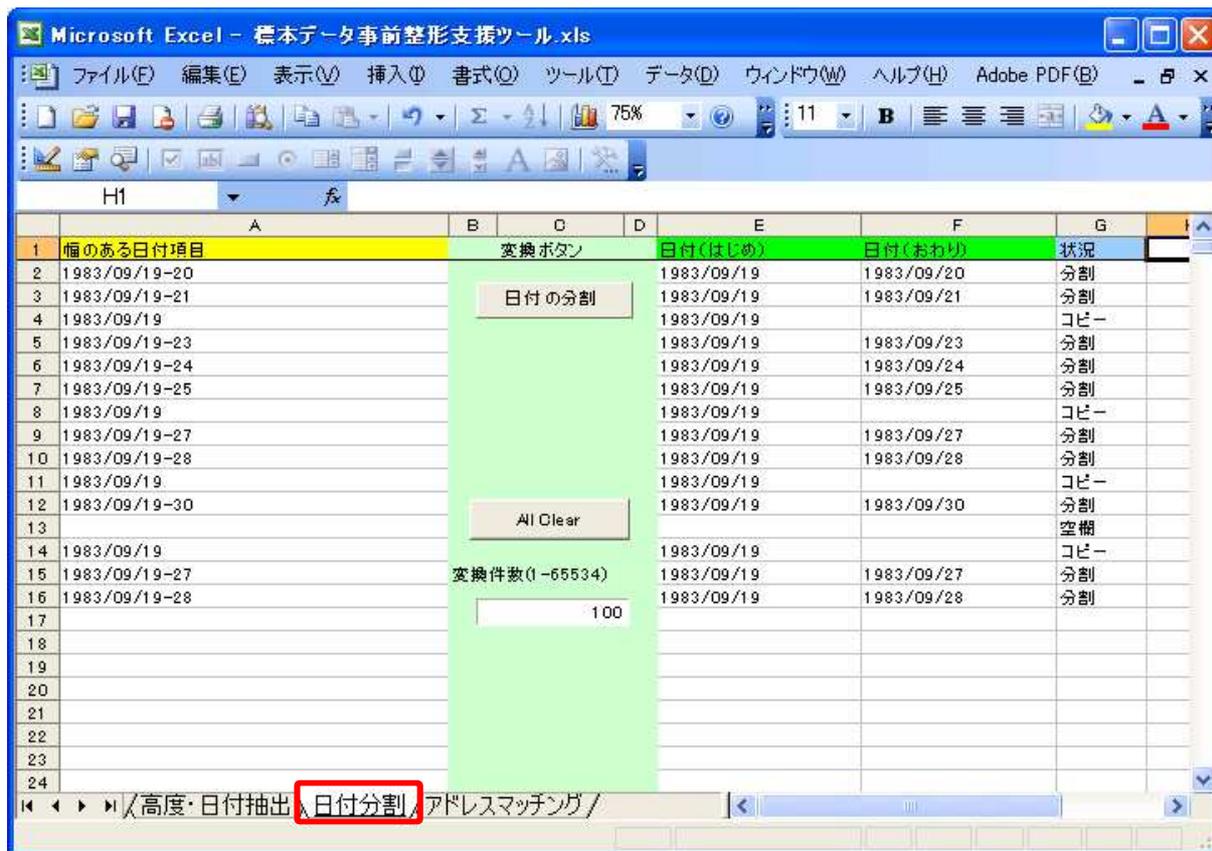
※抽出できる高度や日付の表記方法に関しては、付録1を参考にしてください。

2.4 幅のある日付の分割

ここでは、ハイフンでつながれた幅のある日付項目を「日付(はじめ)」と「日付(終わり)」に分割する作業を行います。幅のない通常の日付が含まれていても、「日付(はじめ)」にコピーされるだけなので、問題はありません。



No.	名称	説明
①	幅のある日付列フィールド	幅のあるデータを貼り付けるフィールドです。
②	日付の分割ボタン	日付の分割をするときのボタンです。
③	日付(はじめ)フィールド	分割後の日付(はじめ)が表示されるフィールドです。
④	日付(おわり)フィールド	分割後の日付(おわり)が表示されるフィールドです。
⑤	日付分割状況フィールド	日付分割結果の状況が表示されるフィールドです。



- 1) 「日付分割」のシートを選択します。
- 2) もともとの標本データのうち、幅のある日付を含む列を貼り付けます。
(貼り付けると表示がおかしくなる場合 → FAQ(よくある質問)を参照してください。)
- 3) 「日付の分割」ボタンを押します。
- 4) 変換された内容を、状況フィールドを参考にしながら確認してください。
元のデータとの照合はExcel行番号を照らし合わせながら行うことができます。

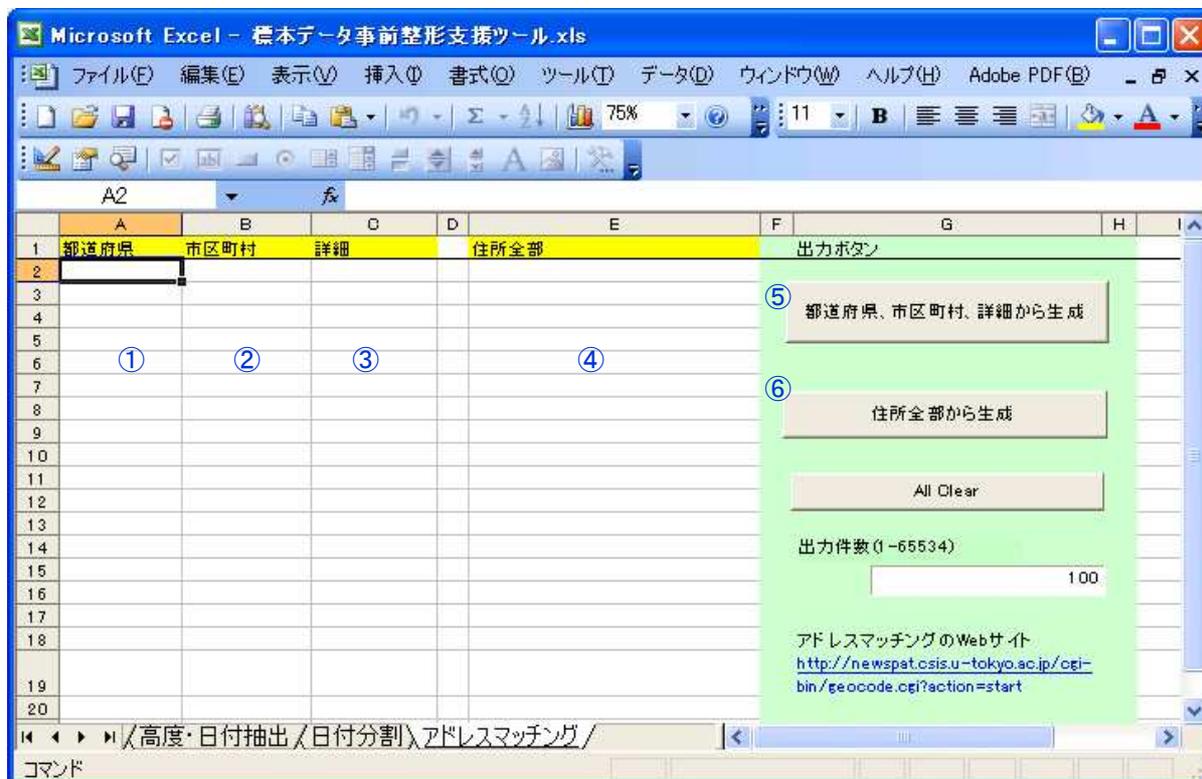
「日付分割機能」での状況フィールドについて

- ・分割 → 正しく分割が完了している項目です。
- ・警告 → 抽出はされているが、その後にまだ日付情報が存在する可能性のある項目です。
誤変換の可能性がありますので、変換内容を確認してください。
- ・コピー → 幅のある日付でないために、日付(はじめ)にそのままコピーされた項目です。
- ・空欄 → もともと空欄だった項目です。

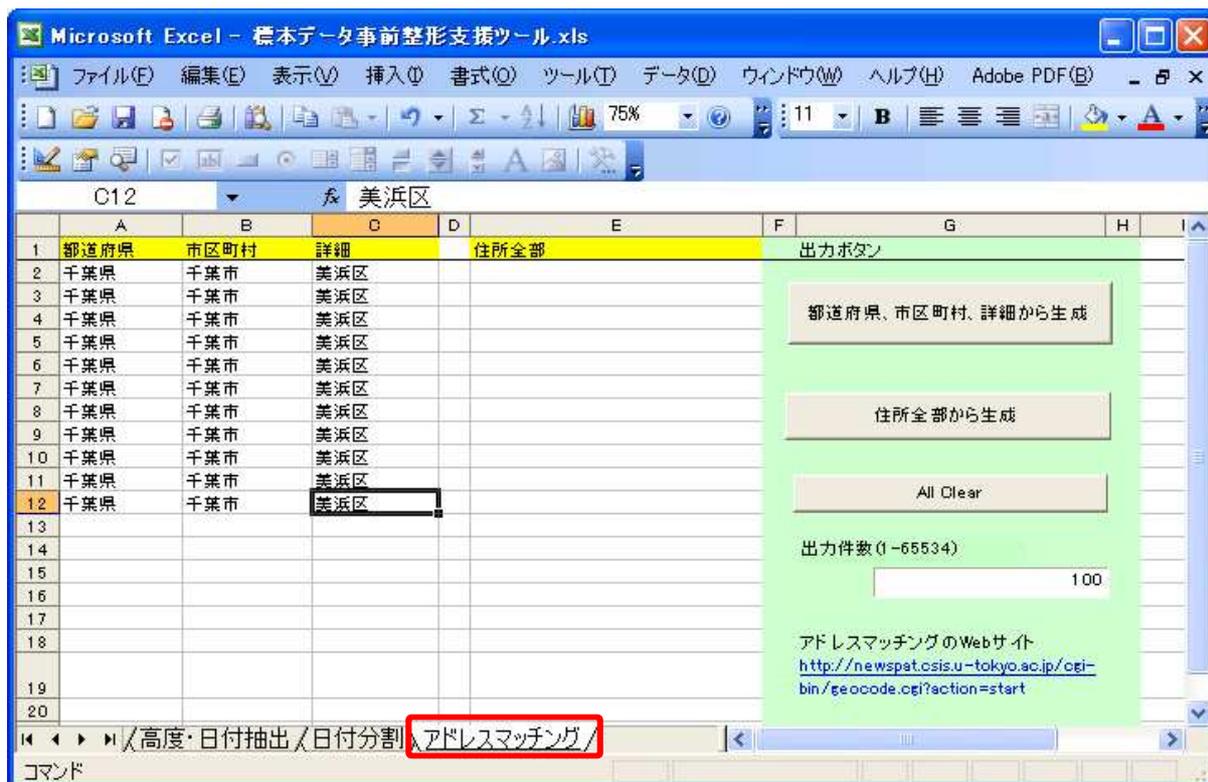
※抽出できる幅のある日付の表記方法に関しては、付録1を参考にしてください。

2.5 アドレスマッチング

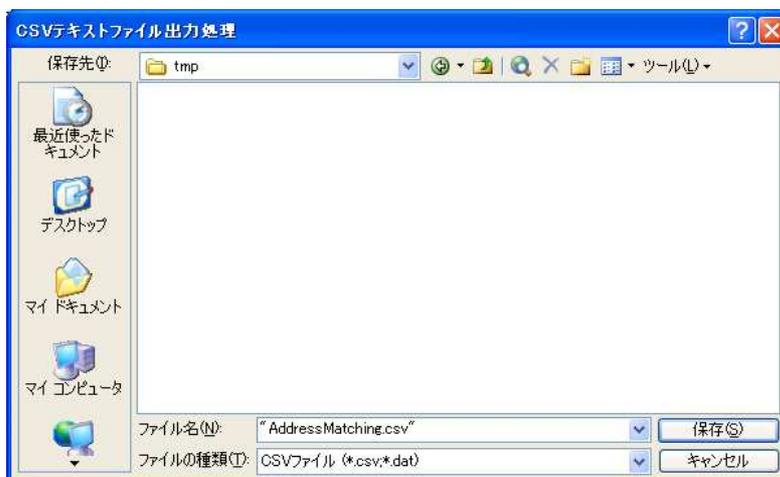
ここでは、アドレスマッチングサービスを利用するためのファイルを生成し、緯度経度情報を取得する作業を行います。



No.	名称	説明
①	都道府県フィールド	都道府県データを貼り付けるフィールドです。
②	市区町村フィールド	市区町村データを貼り付けるフィールドです。
③	詳細フィールド	住所の詳細データを貼り付けるフィールドです。
④	住所全体フィールド	住所全体のデータを貼り付けるフィールドです。
⑤	都道府県、市区町村、詳細から生成ボタン	都道府県、市区町村、詳細に入力した情報を用いてアドレスマッチング用のファイルを出力するボタンです。
⑥	住所全部から生成ボタン	住所全体に入力した情報を用いてアドレスマッチング用のファイルを出力するボタンです。



- 1)「アドレスマッチング」のシートを選択します。
- 2)住所が都道府県以下分割されたデータを用いる場合は、都道府県フィールド、市区町村フィールド、詳細フィールドにそれぞれ対応したデータを貼り付けます。まとめられた住所データを用いる場合は、住所全部フィールドに対応したデータを貼り付けます。
- 3)「都道府県、市区町村、詳細から生成」ボタン、もしくは「住所全部から生成」ボタンを押します。
- 4)下の様なCSVファイル保存画面が表示されますので、任意のフォルダに保存してください。

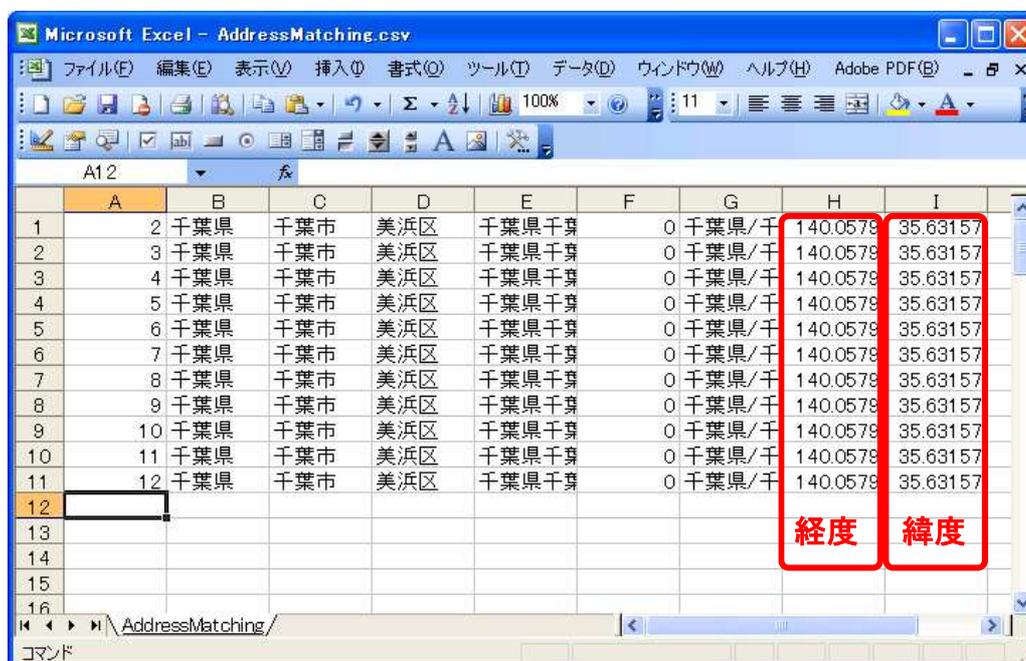


アドレスマッチング用CSVファイル保存画面

パラメータ設定

対象範囲?	全国街区レベル(経緯度・世界測地系)	
住所を含む カラム番号?	5	
入力ファイルの 漢字コード?	自動設定	
出力ファイルの 漢字コード?	入力ファイルと同じ	
出力ファイルの フォーマット?	csvフォーマット	<input type="checkbox"/> x,yを反転?
変換したい ファイル名?	参照...	
送信 クリア		

- 5) アドレスマッチングサービスのサイトにアクセスします。
(<http://newspat.csis.u-tokyo.ac.jp/geocode/>)
- 6) 「住所を含むカラム番号」を5に設定し、変換したいファイル名に先に保存したCSVファイルを入力します。
- 7) 送信ボタンを押すと再度CSVファイルの保存画面が出てきますので、任意のフォルダに保存します。
- 8) 保存されたCSVファイルを開くと、下の図の用になります。H列が経度、I列が緯度になっていますので、元の標本データに緯度経度情報としてこの情報を追加してください。



	A	B	C	D	E	F	G	H	I
1	2	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
2	3	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
3	4	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
4	5	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
5	6	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
6	7	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
7	8	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
8	9	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
9	10	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
10	11	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
11	12	千葉県	千葉市	美浜区	千葉県千葉	0	千葉県/千	140.0578	35.63157
12									
13									
14									
15									
16									

経度 緯度

アドレスマッチングサービスから出力されたCSVファイル