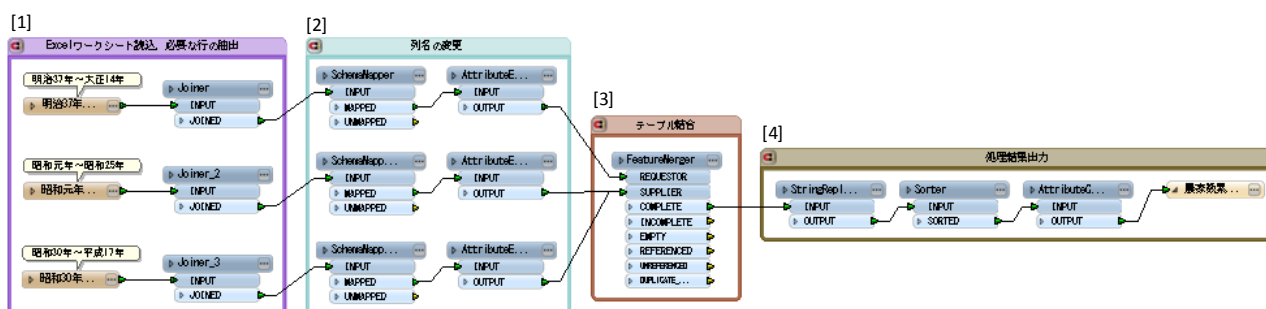


FME ワークスペースサンプル <非空間データ処理>

農林業センサス累年統計「農家数」(Excel) のデータテーブルへの変換



概要

- ウェブ上で公表されている統計データには印刷して利用することを想定して作成されたものがあり、そのままの形式ではデータテーブルとして利用するのが難しい。
- e-Stat (政府統計の総合窓口) で公表されている農林業センサス累年統計「農家数」(Excel 形式) もそのひとつであり、明治 37 年から平成 17 年までの各統計年次の全国、農業地域別、都道府県別農家数が 3 つのワークシートに分けて掲載されている。
- これをデータとして利用し易くするために、タイトル行、農業地域別集計行、空行などを除いた 1 テーブルに変換する。

入力

- Excel 形式 3 ワークシート (e-Stat からダウンロード)
 農林水産業 > 農林業センサス > 農林業センサス累年統計 > 農業編 (明治 37 年～平成 17 年)
 全国農業地域・都道府県別統計表
 表番号 2-1 農家数
 ダウンロードファイル名: j357-17-010.xls
 ワークシート名:
 1) 明治 37 年～大正 14 年
 2) 昭和元年～昭和 25 年
 3) 昭和 30 年～平成 17 年

出力

- Excel 形式 1 ワークシートに全国と各都道府県の全ての統計年次の農家数を出力する。
- ヘッダ部 1 行に統計年次等を表す簡潔な列名を出力する。
- 先頭の列には都道府県コード (全国は 0 とする) を付加する。

「農家数」表はデータ数が少なく構造もシンプルなので、手作業で必要なテーブルに変換した方が FME ワークスペースを作成するよりも早く目的が達成できると思われませんが、非空間データ処理の側面における FME の機能を簡潔に紹介するのに適当な事例として選びました。処理結果は Excel だけでなく、データベーステーブルや CSV ファイル等にも出力することもできます。

キーワード

FME ワークベンチ	FME Workbench	FME の中核であるソフトウェアの名称
ワークスペース	Workspace	FME ワークベンチで作成されたデータ処理フロー
リーダー	Reader	データ読み込みを担うワークスペース構成要素
トランスフォーマー	Transformer	データ変換を担うワークスペース構成要素
ライター	Writer	データ出力を担うワークスペース構成要素
フィーチャー	Feature	地物を表す幾何図形とその属性、テーブルの 1 レコードなど、処理対象とする事物を表すひとかたまりのデータ

印刷して利用することを想定した形式の Excel ワークシート（左）からデータとして利用し易い形式のテーブル（右）に変換するのが目的です。

1 農家数

明治37年～大正14年

全国農業地域	都道府県	明治37年	注)38年	注)39年	40年	41年	42年
全 国	1	5 416 703	5 379 969	5 378 337	5 406 172	5 408 363	5 407 203
北 海 道	2	86 181	103 414	111 260	133 349	147 035	149 368
都 府 県	3	5 320 542	5 276 556	5 292 529	5 272 823	5 261 328	5 257 835
東 北	4	538 833	529 533	530 552	523 491	540 109	535 751
北 陸	5	442 366	441 099	443 097	440 061	433 673	428 048
関 東	6	1 106 944	1 090 940	1 093 408	1 107 015	1 089 303	1 084 739
東 山	7	386 074	358 512	381 380	384 248	385 477	368 877
南 東	8	472 825	465 969	465 718	470 818	453 921	457 138
東 山	9	288 045	286 459	288 332	271 949	289 905	270 974
東 海	10	649 319	638 034	632 581	632 944	629 905	633 503
中 近 畿	11	617 845	593 947	619 133	604 420	610 398	607 882
中 山 陸	12	672 104	677 296	672 833	674 942	677 245	683 114
山 陽	13	187 195	187 008	184 536	166 168	166 251	169 738
山 陽	14	504 909	510 233	508 347	508 774	510 994	513 376
西 国	15	412 891	412 082	408 136	402 450	405 033	398 333
九 州	16	380 841	384 915	393 066	377 800	375 102	373 215
北 九 州	17	626 243	626 079	625 219	617 810	614 791	618 199
南 九 州	18	254 398	256 536	257 855	258 790	280 321	282 149
沖 縄	19	---	---	---	---	---	---
青 森	20	64 835	67 373	66 282	68 923	69 399	70 413
岩 手	21	87 804	85 755	87 517	88 330	91 996	91 384
宮 城	22	88 744	84 480	84 002	83 725	84 448	84 420

入力ワークシート

都道府県CD	都道府県名	M37	M38	M39	M40	M41	M42
0	全国	5416703	5379969	5378337	5406172	5408363	5407203
1	北海道	86181	103414	111360	133349	147035	149368
2	青森	64635	67373	66262	68923	69396	70413
3	岩手	87804	85755	87517	88330	91996	91384
4	宮城	88744	84460	84002	83725	84449	84420
5	秋田	82624	79796	80229	80831	83024	77203
6	山形	88226	86262	87049	87591	86114	86320
7	福島	126800	124887	125193	124091	125130	126011
8	茨城	160206	159724	162298	163480	164483	164800
9	栃木	95530	91609	92037	92170	92600	94075
10	群馬	110338	107179	107025	108588	108414	108002
11	埼玉	167569	157670	161824	162745	147821	148442
12	千葉	163043	161382	159295	160787	161108	161715
13	東京	63944	67831	67296	70395	68382	69854
14	神奈川	78269	79086	77301	76891	76610	77127
15	新潟	197647	199142	196534	198405	196680	193763
16	富山	85563	86223	85600	84843	82794	79650
17	石川	81502	81866	83834	80744	78291	77292
18	福井	77654	73868	77099	76069	75908	75343
19	山梨	72420	70782	68401	71828	72796	74175
20	長野	195625	195697	197931	200121	197109	196799
21	岐阜	148319	142713	142655	141123	140170	140959
22	静岡	164005	155218	152219	153690	150783	155073
23	愛知	211577	213146	211033	212979	218228	215220
24	三重	125387	126857	126674	125052	120784	122251
25	滋賀	93283	93430	96602	94950	94614	95079
26	京都	87650	86088	84870	83413	82845	82010
27	大阪	95807	93406	94588	98243	97082	96429
28	兵庫	195214	185815	206391	191559	197623	195574
29	奈良	64155	63403	63579	62175	62081	62457
30	和歌山	79536	71805	73103	74380	76653	76433

出力テーブル（処理結果）

[ワークスペースからの参照用外部テーブルの作成]

都道府県コード表（都道府県コードと都道府県名の対応表。全国のコードは0）、入力ワークシートごとの新旧属性名対応表、出力テーブルのスキーマ定義表（列名、データ型など）を作成します。

都道府県コード表は"Joiner"、新旧属性名対応表は"SchemaMapper"、スキーマ定義表は処理結果出力時のライターから参照されます。

ワークスペースを作成するときこのような外部テーブルが常に必要なわけではありませんが、トランスフォーマーやライターから外部テーブルを参照する仕組みを使うことにより、処理過程におけるフィーチャーの属性名の変更や出力時のフィーチャーのスキーマ定義がダイナミックに行えるようになり、より柔軟なワークスペースとなります。

[1] Excel ワークシート読込，必要な行の抽出

入力ワークシートごとに読込用のリーダーを配置し、それぞれ"Joiner"によって都道府県コード表を用いて全国と都道府県の行のみを抽出します。

Excel 用のリーダーでワークシートを読み込むとき、ワークシートの先頭行が列名であるときはそれを入力フィーチャーの属性名にできますが、入力ワークシートの先頭行は列名ではないので、ここではリーダーがデフォルトで設定する属性名 F1, F2...を使用します。

"Joiner"は外部のテーブルを読み込み、結合キーを指定することによって外部テーブルのレコードを入力フィーチャーにその属性として結合します。データベースにおけるテーブルの内部結合（inner join）に相当します。

入力ワークシートの F2 列に都道府県名等が記載されているので、それをキーとして都道府県コード表と結合することにより、都道府県コード表に記載した全国と都道府県の行だけが抽出されるとともに、都道府県コードがフィーチャーの属性に追加されます。

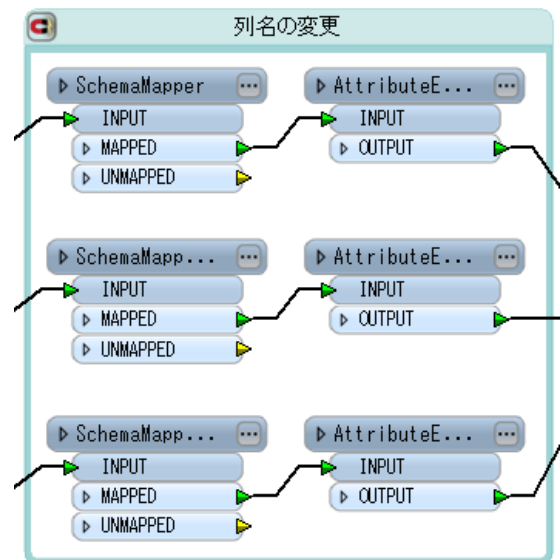


[2] 列名の変更

前述したように入力フィーチャーの属性名(列名)は F1, F2... という記号化された名前となっており、3 つのワークシートから読み込まれたフィーチャーを結合するときには衝突するので、"SchemaMapper"によって属性名を変更します。"SchemaMapper"は、外部テーブルとして作成した新旧属性名対応表を参照して属性名を変更します。

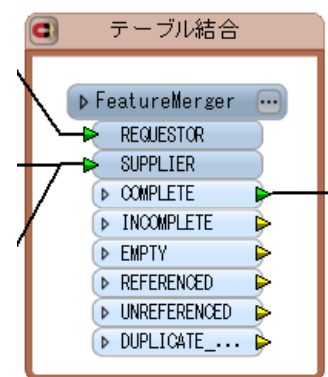
"SchemaMapper"による変更後の属性名を後続のトランスフォーマーで参照できる(見える)ようにするには、"AttributeExposer"を使います。

なお、フィーチャーの属性名は"AttributeRenamer"でも変更することができ、その場合は新旧属性名の対応をトランスフォーマーのパラメータ設定画面上で設定します。変更する属性名が少ないときは"AttributeRenamer"が簡便です。



[3] テーブル結合

都道府県コードをキーとして、"FeatureMerger"によって3つのワークシートから読み込まれたフィーチャーをひとつのフィーチャーに統合します。これによって、明治37年から平成17年までの各統計年次の農家数を属性として持つフィーチャー(出力テーブルのレコード)ができ、3テーブルを結合したことに相当する結果が得られます。



[4] 処理結果出力

入力ワークシートの数値データ欄は3桁ごとのスペース区切りの書式が設定されています。そのような書式付きのデータは処理過程では文字列として扱われ、そのままでは数値データとして出力することができないので、"StringReplacer"によってスペースを削除しておきます。

次に、"Sorter"によってフィーチャー(レコード)の出力の順番を都道府県コード昇順とするためにフィーチャーをソートします。

一般的には出力テーブルのスキーマ(列名やデータ型等)はライターのプロパティ画面で設定しますが、実行時に外部テーブルを参照して設定するオプション(「ダイナミックスキーマ」と呼ばれます)もあります。ダイナミックスキーマを使うときは、同じスキーマで出力するフィーチャーの fme_feature_type (フィーチャータイプ名) 属性には同じ名前が設定されている必要があるため、ここでは"AttributeCreator"によってフィーチャータイプ名を設定しました。

Excel形式ファイル書込用のライターを接続してダイナミックスキーマのオプションを設定すれば、ワークスペースは完成です。ワークベンチのメインメニューまたはツールバーの操作で実行することにより、はじめに示した出力テーブルが作成されます。

