

# 基于 FME 技术对浙江省天地图底层数据标准化建设

## 1. 前言

### 1.1 背景

天地图是由国家、省、市(县)三级节点构成的国家地理信息公共服务平台。随着天地图应用的日益广泛与深入,采用“超链接”和“服务聚合”的方式已无法满足跨省(市)域、跨层级(国家、省、市)的地理实体与地名地址查询、专题信息挂接、综合统计分析等应用需求。为此,国家测绘地理信息局于 2013 年启动国家、省、市节点同构试点,浙江省测绘与地理信息局于 2014 年开展数据融合工作,制定了数据融合标准。

浙江省天地图的底层检索数据作为平台的数据基础,亦作为用户与软件互交的桥梁,在使用天地图过程中扮演着极其重要的角色,当用户输入检索项时可迅速定位显示,供用户浏览、查询。浙江省天地图的底层数据包括 POI(兴趣点)、地名数据和地址数据,而目前在数据生产和常规运维过程中,数据的类型、坐标并不满足底层数据检索的要求。因此,通过 FME 对天地图数据进行标准化处理,使处理结果满足底层数据入库、检索要求。

### 1.2 总体要求

常规天地图运维数据包括省基础库、政务报送数据、县市融合数据等。数据类型包括 POI、地名、地址和电子地图,其中电子地图数据不参与底层数据的检索,因此只对前三者数据进行标准化处理。POI 和地址数据为点要素类型,地名数据涵盖了点、线、面等要素类型。少部分地方提交的数据坐标为地方独立平面坐标系。

天地图底层数据目前均以点要素类的形式存储,且坐标为国家 2000 大地坐标系。因此,在处理过程中,需要将 GIS 数据中不同类型的要素(包括线、面、三维等)统一转换成点;并将少数坐标不对应的数据转换到统一的 CGCS2000 大地坐标系下。使得最终的数据满足“同结构、同坐标”的标准数据。数据标准化流程如下图 1-1 所示。

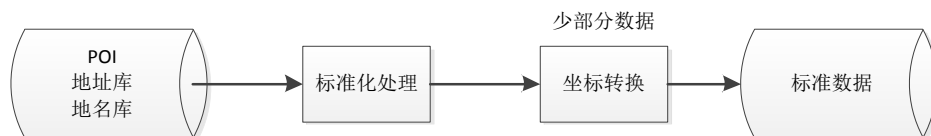


图 1-1: 数据标准化流程

标准的底层数据规定了字段、坐标等信息,具体要求如下表 1-1 所示。

表 1-1: 标准数据结构要求

数据类型	字段名称	字段说明	字段类型	字段长度	坐标系
------	------	------	------	------	-----

POI	FEATUREGUID	永久固定码	Test	32	CGCS2000 大地坐标系
	FCODE	分类代码	Test	32	
	NAME	名称	Test	64	
	ADDRESS	地址	Test	255	
	PHONE	电话	Test	30	
	WEBSITE	网址	Test	254	
	FSCALE	要素显示的最小比例尺	SHORT		
	PAC	行政区划	Test	6	
地名库	NAME	名称	Test	64	
	ADDRESS	地址	Test	255	
	OID	地名标识码	Test	128	
	addcode	地名地址代码	Test	50	
	PAC	行政区划	Test	6	
地址库	NAME	名称	Test	255	
	ADDRESS	地址	Test	255	
	ADDCODE	地址代码	Test	100	
	PAC	行政区划	Test	6	

## 2. 数据标准化

FME 对 GIS 数据标准化的过程，其实就是对 GIS 数据中的异构数据统一转换到同构的点要素，其中地址数据和 POI 数据本身数据类型为点要素，在转换到标准点要素时需要做字段匹配和空间匹配。而地名库数据包括了点、线、面，需要对各要素分别转换到标准点要素。以下将点、线、面要素的转换分别进行梳理，涵盖了常规运维数据中的地址数据、POI 数据和地名数据。数据标准化处理流程如下图 2-1 所示。

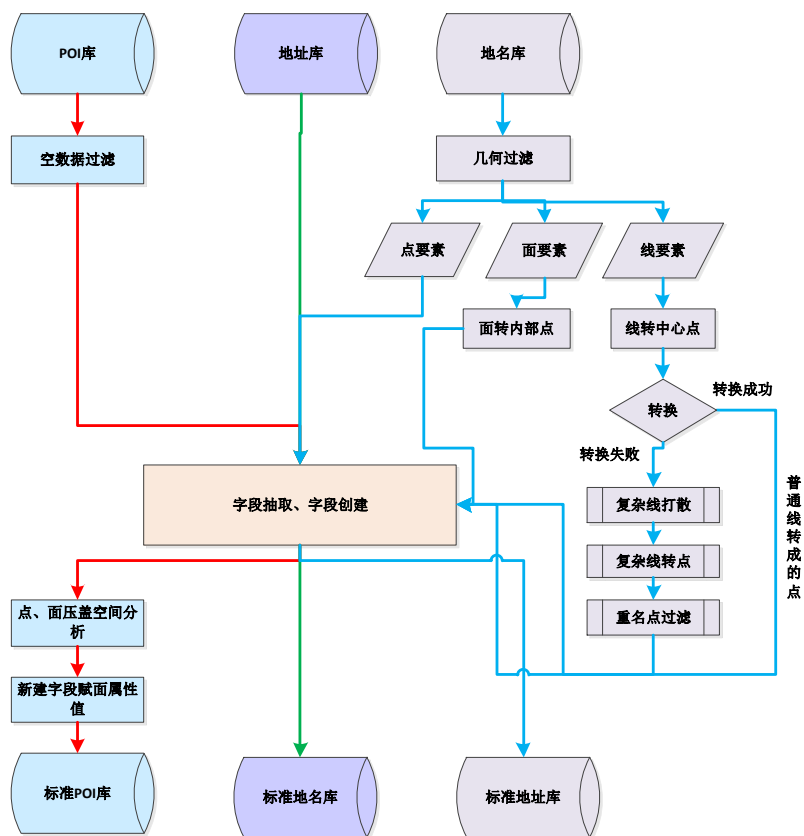


图 2-1: 数据标准化处理流程

## 1.1 点要素标准化

点要素包括地名点、地址点和 POI 点，在标准化的过程中存在一定的差异，现分别对其标准化流程进行描述。

### 1.1.1 地名、地址点标准化

地名点和地址点在转换成标准点要素过程中，只需要提取标准字段即可。地名点由于是从地名库中提取出来，因此需要先做几何点过滤处理（该步骤在下文线要素标准化中会详细介绍）。地名、地址点标准化模型如下图所示。

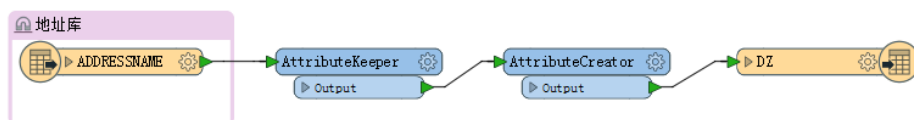


图 2-2: 地名点、地址点标准化模型

- 使用 AttributeKeeper 抽取有效字段，地名有效字段包括：ADDRESS、addcode、OID、NAME；地址有效字段包括：ADDRESS、NAME、ADDCODE。
- 使用 AttributeCreator 创建 PAC（行政区划字段），地名库中该字段存储在 OID 字段中，使用文本编辑器进行数据提取，如下图 2-3 所示。地址库中该字段存储在

在 ADDCODE，提取规则同地名库。

用属性来创建	
新属性	属性值
PAC	<input type="checkbox"/> @Left(@Value(OID),6)

图 2-3：属性创建、赋值

### 1.1.2 POI 标准化

POI 点在转换成标准点要素过程中，由于其本身无 PAC 字段且无法从其他字段中提取。因此，需要通过地名库，将 POI 点落在地名库中 BOU\_COU\_PY 面要素类上，做点、面空间叠加分析，提取 BOU\_COU\_PY 行政区划面中的区域代码字段，创建 POI 的 PAC 字段并用 BOU\_COU\_PY 区域代码字段对其赋值。POI 数据标准化模型如下图 2-4 所示。

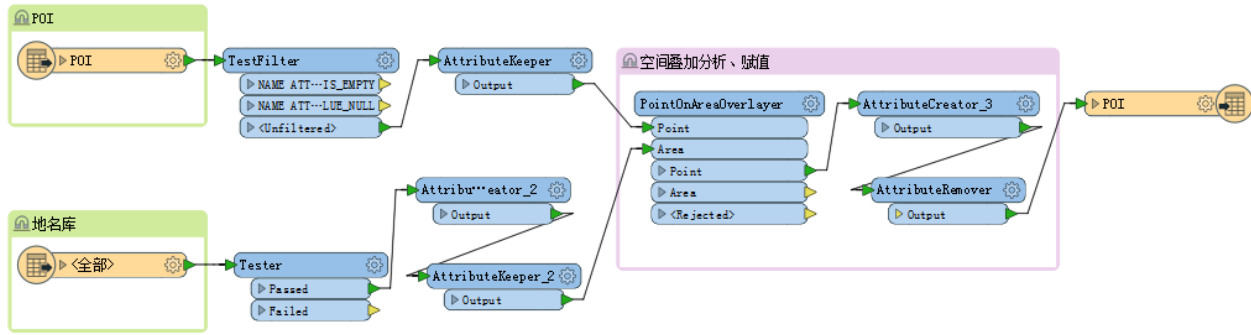


图 2-4：POI 数据标准化模型

- 使用 TestFilter 工具过滤 POI 中空要素和无属性值的要素。
- 使用 AttributeKeeper 抽取有效字段：FCODE、ADDRESS、WEBSITE、FEATUREGUID、FSCALE、NAME、PHONE。
- 使用 Tester 工具对地名库数据按照 fme\_feature\_type 进行过滤，抽取出 BOU\_COU\_PY 要素类。并对 BOU\_COU\_PY 要素类创建 PAC 字段（临时使用），并用文本编辑器提取 TNTIID 字段中的 6 位编码：@Right(@Left(@Value(ENTIID), 10), 6)，如下图 2-5 所示。

用属性来创建	
新属性	属性值
PAC	<input type="checkbox"/> @Right(@Left(@Value(ENTIID),10),6)

图 2-5：属性创建、赋值

- 使用 PointOnAreaOverlay 工具进行空间点、面叠加分析，并创建 point 输出列表，对其命名，并赋 PAC 值，如下图 2-6 所示。

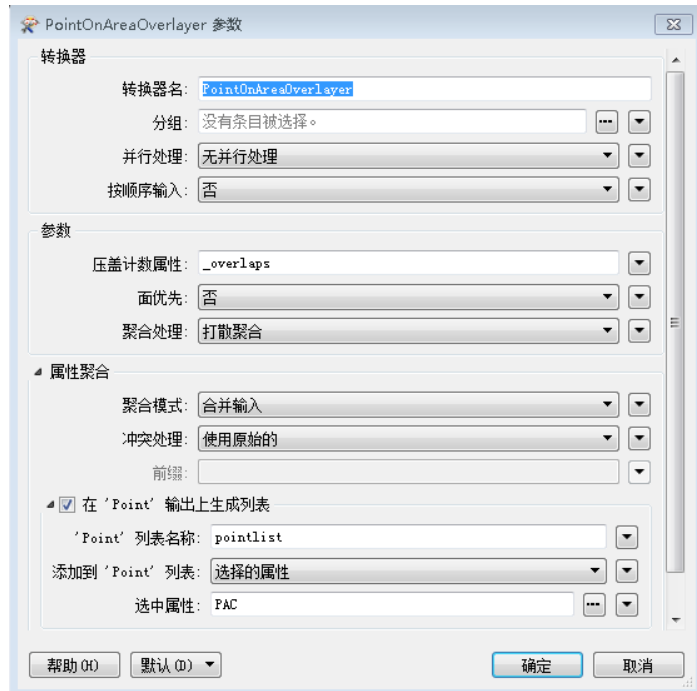


图 2-6：点、面空间压盖分析

- 使用 AttributeCreator 创建 PAC 字段（正式使用），并赋予 pointlist {0}.PAC 列表值，做最终字段属性使用，如下图 2-7 所示。

用属性来创建	
新属性	属性值
PAC	pointlist{0}.PAC

图 2-7：属性创建、赋值

- 点要素处理结果如下图 2-8 所示。

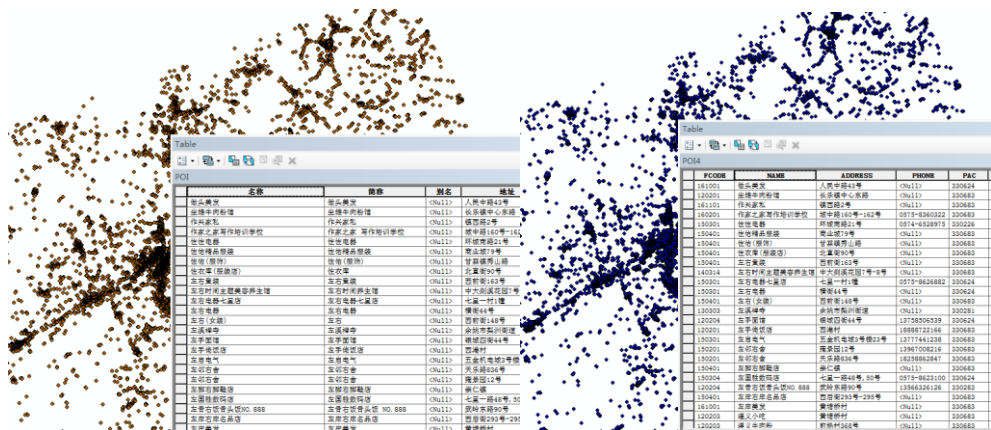


图 2-8：点要素标准化处理结果对比（字段对比）

## 2.2 线要素标准化

线要素存储于地名库中。线的标准化处理是对地名库中的线要素转换成点，由于线要素既包含了结构单一的普通线，也包括了结构多样的复杂线。在处理过程中，需要对这两种线分别进行处理，且处理方式不同。下图 2-9 为一条简单线，图 2-10 为一条复杂线（均为高亮部分）。下图 2-11 为地名库标准化模型。

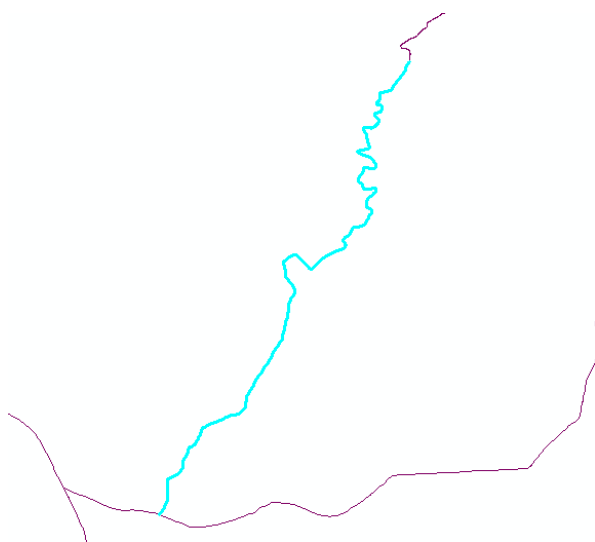


图 2-9：简单线样例

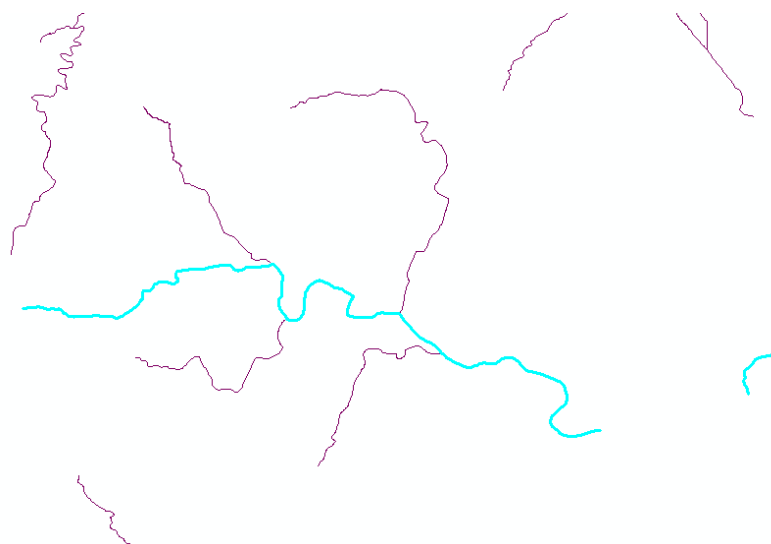


图 2-10：复杂线样例

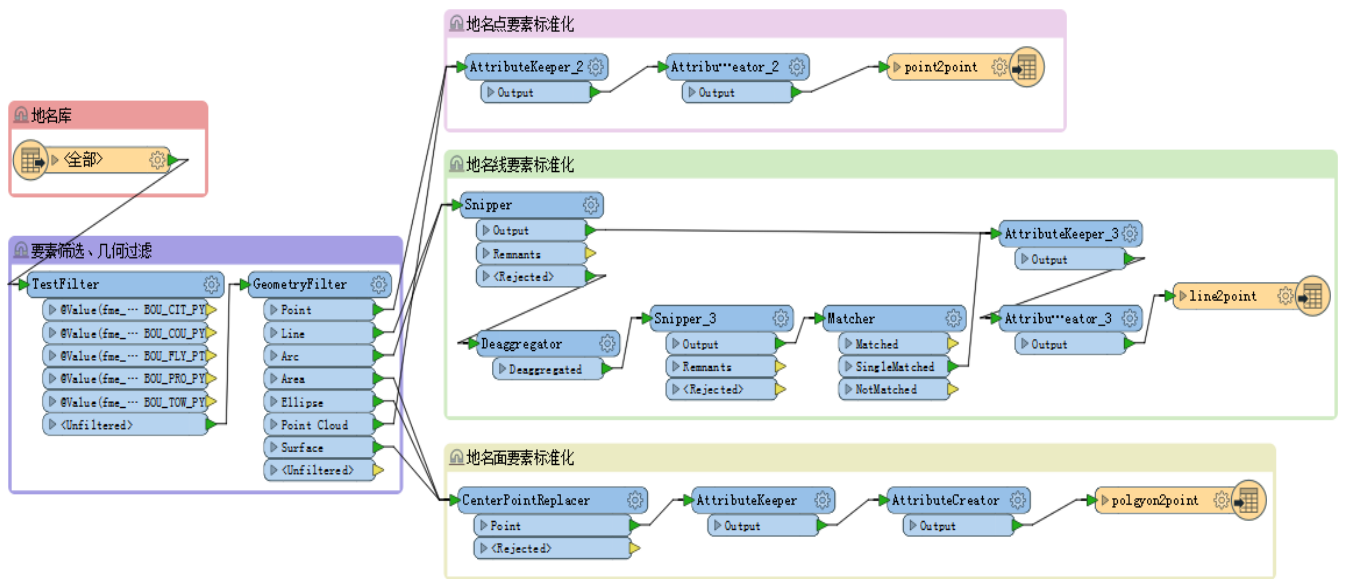


图 2-11: 地名库标准化模型

- 对地名库按照 fme\_feature\_type 属性，使用 TestFilter 工具过滤掉 BOU 类型的要素类（境界要素不参与后台检索），如下图 2-12 所示，并使用 GeometryFilter，筛选出几何类型为线的要素。

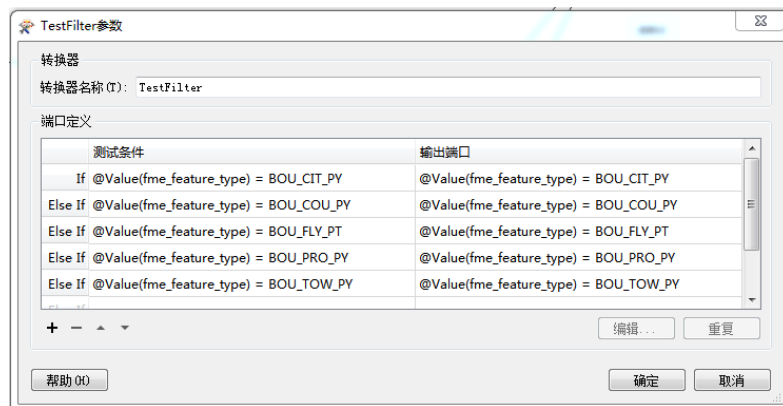


图 2-12: 境界要素过滤

- 使用 Snipper 工具，对线进行裁剪赋点，如下图 2-13 所示，间断模式使用距离百分比，并按照 50%的比例赋点，将点落在线段的中心位置。聚合参数选项选择拒绝，用来分流简单线和复杂线。简单线直接输出到下一步，复杂线则需要进一步打散处理。



图 2-13: Snipper 参数设置

- 复杂线为多线段的组合，可表示同一条路在某处断开后又重新连接或者道路分流，在空间上属于离散分布，在结构上属于同一线要素。使用 Deaggregator 工具，对复杂线进行打散处理，如下图 2-14 所示。



图 2-14: Deaggregator 参数设置

- 再次使用 Snipper 工具，对被打散的线重构点，同样在聚合参数选择中选拒绝。
- 打散线重构生成的点，属性依附于复杂线。同一条复杂线（已经打散）可能会衍生出许多属性相同的点，需要做点与点之间的匹配筛选处理，过滤重复点，使得一个点只对应一条复杂线。（例 1：环城南路复杂线转换成点后，出现多个名为环城南路的点，只保留一个；例 2：同名为环城南路的线，可属于杭州、亦可属于宁波，同名但位置不同均保留）。因此，在属性匹配模式中，选择匹配全部属性，使得全部属性相同的点进行匹配，避免出现上述例子中的情况。且只比对属性，不对比空间关系，因此在比对几何对象中选择“无”。下图 2-15 为 Matcher 参数



设置。

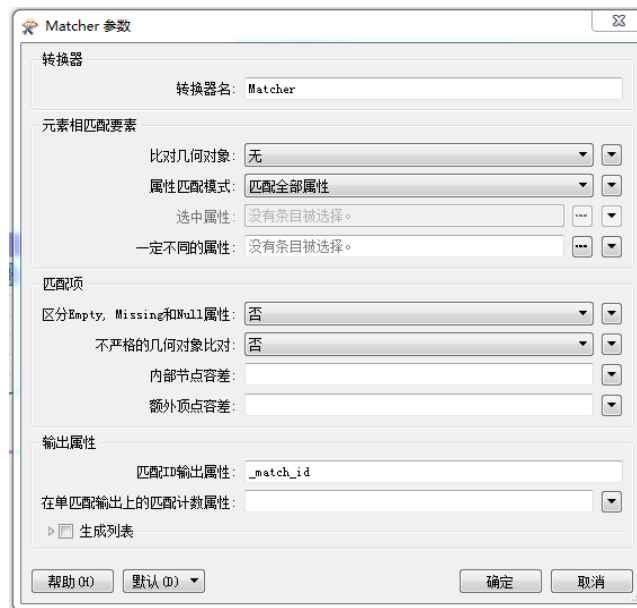


图 2-15: Matcher 参数设置

- 使用 AttributeKeeper 保留 ADDRESS, addcode, OID, NAME 字段，使用 AttributeCreator 创建 PAC 字段并提取 OID 字段的值，提取规则同地名库。
- 线要素处理结果如下图 2-16 所示：

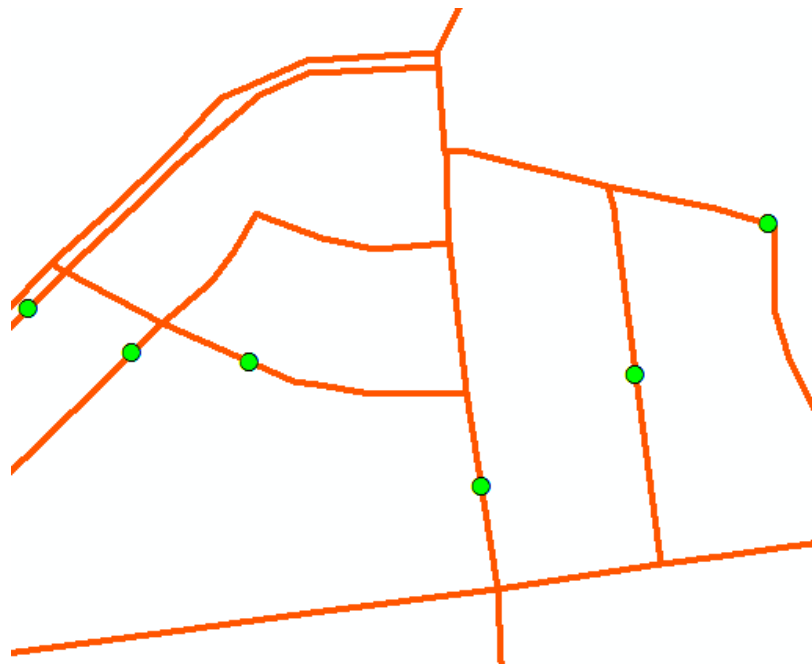


图 2-16: 线要素转点处理结果

### 2.3 面要素标准化

面要素存储于地名库中，对面要素处理过程，主要是将面转换成面的内部点。

- 对地名库按照 fme\_feature\_type 属性，使用 TestFilter 工具过滤掉 BOU 类型的要素类，如下图 2-16 所示，并使用 GeometryFilter，筛选出几何类型为面的要素。
- 使用 CenterPointReplacer 对面创建内部点，如下图 2-17 所示。



图 2-17: CenterPointReplacer 参数设置

- 使用 AttributeKeeper 抽取有效字段: ADDRESS、addcode、OID、NAME，并使用 AttributeCreator 创建 PAC 字段，抽取 OID 字段中的数据，方法同地名地址点抽取。
- 面要素处理结果如下图 2-18 所示。

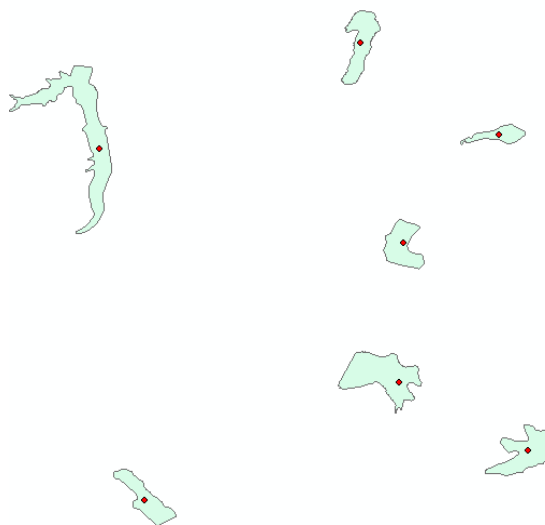


图 2-18: 面要素标准化处理结果

### 3. 坐标转换

对常规的运维数据进行标准化过程中，无需进行坐标转换。地方区县提交的数据为地方坐标系，需要做一定的坐标转换和投影反算，最终转换成 2000 国家大地坐标系。坐标转换过程如下图 3-1 所示。



图 3-1：坐标转换基本流程

### 3.1 四参数计算

地方提供的控制点是建立在地方独立坐标系到 CGCS2000 平面坐标系 120 度带的过渡关系上，利用坐标转换公式可计算得到仿射变换需要的参数。坐标转换公式如下图 3-2，控制点信息如下图 3-3 所示。

$$\begin{aligned}
 X &= \Delta x + (1+k) * (\cos \alpha * x + \sin \alpha * y) \\
 Y &= \Delta y + (1+k) * (-\sin \alpha * x + \cos \alpha * y) \\
 \text{令:} \\
 \Delta x &= a \\
 \Delta y &= b \\
 (1+k) * \cos \alpha &= c \\
 (1+k) * \sin \alpha &= d \\
 \text{则:} \\
 X &= a + cx + dy \\
 Y &= b - dx + cy
 \end{aligned}$$

图 3-2：坐标转换公式

点号	XX地方独立坐标系		XX地方2000坐标系	
	$X_i$	$Y_i$	$x_i$	$y_i$
T1	3168098.042	475930.865	3168022.964	490679.918
T2	3162382.306	485963.049	3162319.854	500719.343
T3	3156173.72	484088.195	3156108.878	498852.293
T4	3157517.942	440264.544	3157397.983	455027.094
T5	3164831.918	442236.536	3164714.363	456989.856
T6	3159958.749	445735.64	3159845.649	460495.063
T7	3169511.73	451965.962	3169406.406	466713.294
T8	3166781.172	459523.513	3166685.399	474274.248
T9	3155850.949	455647.832	3155750.349	470412.349
T10	3172122.22	464495.467	3172032.704	479239.444
T11	3163579.274	463292.836	3163488.264	478047.594
T12	3168754.867	469727.971	3168671.962	484476.192
T13	3163432.976	472041.759	3163352.99	486796.69
T14	3159344.538	467619.088	3159258.984	482379.169
T15	3158299.187	477276.4	3158225.785	492037.801
T16	3174956.775	478125.336	3174884.483	492865.732
T17	3164610.747	481655.42	3164542.878	496408.887
T18	3175226.906	483959.379	3175161.99	498699.454

图 3-3：地方控制点的相关坐标系

说明：由于涉密原因，上述数据已经做过加密处理，并统一赋点号。以下仿射变换参数同样做过加密处理。

### 3.2 仿射变换

#### (1) 仿射变换

利用 Affiner 工具，将上述计算得到的参数输入到工具中，进行仿射变换。从地方独立坐标系转换到 CGCS2000 平面 120 度坐标系。Affiner 参数设置如下图 3-4 所示，坐标转换过程如下图 3-5 所示。

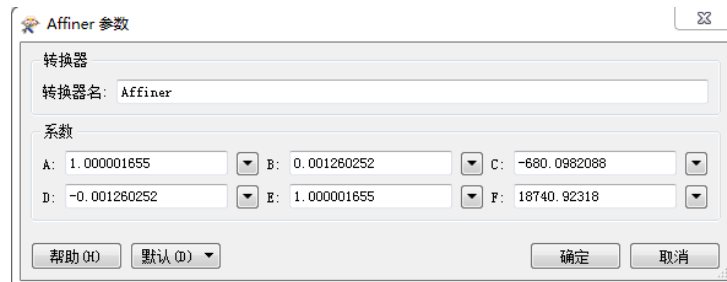


图 3-4：仿射变换参数设置



图 3-5：地方独立坐标系转到 CGCS2000 坐标系

### (2) 精度校验

坐标转换后的数据需要与参考数据进行对比校验，误差在特定范围内，满足精度要求。精度校验结果如下图 3-6 所示。

XX地方2000坐标系		坐标转换后		误差		
xi	yi	Xa	Ya	$\Delta X$	$\Delta Y$	d
3168022.964	490679.918	3168022.968	490679.9871	0.00449511	0.069079403	0.0692255
3162319.854	500719.343	3162319.866	500719.391	0.012112	0.047954041	0.04945999
3156108.878	498852.293	3156108.907	498852.3582	0.029045849	0.065232611	0.071406966
3157397.983	455027.094	3157397.902	455026.9406	-0.080567961	-0.153372495	0.17324641
3164714.363	456989.856	3164714.376	456989.7184	0.01274643	-0.137560098	0.138149383
3159845.649	460495.063	3159845.608	460494.9697	-0.040568304	-0.093347142	0.101781512
3169406.406	466713.294	3169406.457	466713.2628	0.05102103	-0.031195684	0.059802309
3166685.399	474274.248	3166685.419	474274.2675	0.019918737	0.019506163	0.027879141
3155750.349	470412.349	3155750.293	470412.3549	-0.055509742	0.005924473	0.055825002
3172032.704	479239.444	3172032.742	479239.4987	0.037674878	0.054671017	0.066395155
3163488.264	478047.594	3163488.266	478047.6319	0.001914675	0.037944007	0.037992285
3168671.962	484476.192	3168671.977	484476.255	0.01537354	0.063046025	0.064893351
3163352.99	486796.69	3163352.994	486796.7538	0.00351925	0.063799529	0.063896518
3159258.984	482379.169	3159258.975	482379.2279	-0.008928348	0.058939849	0.059612258
3158225.785	492037.801	3158225.793	492037.8733	0.007986866	0.072332385	0.072772
3174884.483	492865.732	3174884.478	492865.818	-0.004564538	0.085980949	0.086102024
3164542.878	496408.887	3164542.882	496408.9464	0.004103422	0.059426097	0.059567601
3175161.99	498699.454	3175161.962	498699.5302	-0.027753631	0.076205668	0.081102207

图 3-6：Affiner 转换成果精度校验

### 3.3 投影反算

在完成平面坐标系系统间的相互转换，并保证转换精度的情况下，基于 WGS 84 椭球做投影反算 (CGCS2000 平面 120 度带投影已经在 FME 投影文件中自定义,CGCS2000 大地坐标还未自定义，而 WGS84 椭球与 CGCS2000 椭球近似，可替换使用)，转换过程如下图 3-7 所示。



图 3-7：投影反算过程

### 3.4 FME 的两种坐标转换方式比较

FME 中存在多种坐标转换工具,常规的有 Affiner、3D Affiner 仿射变换工具,在 MY FME 的插件中还有 4PReprojector、7PReprojector。后两者在应用过程中,直接输入四参数或者七参数即可转换,无需像 Affiner 需要根据四参数来计算公式的常数。但是在实际应用过程中,发现两者的转换精度存在一定的差异,下文通过对两者转换方式的比较来确定何种方式更利于坐标转换。

利用坐标转换公式,计算得到四参数,并在 4PReprojector 工具中直接输入,如下图 3-8 所示。投影反算过程如下图 3-9 所示。



图 3-8:4Preprojector 工具窗口

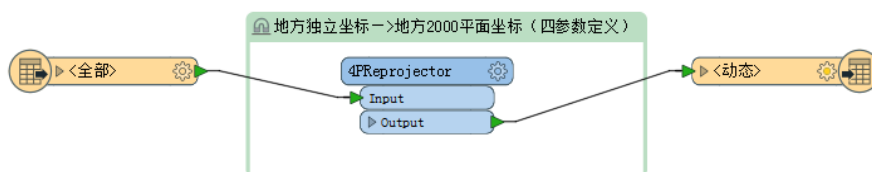


图 3-9: 4PReprojector 四参数转换过程

4PReprojector 转换后的数据(已做过加密处理)与原始参考数据进行对比较验,存在较大误差,如下图 3-10 所示,综合对比 Affiner 的处理结果,Affiner 仿射变换工具得到的结果精度更高,更适合坐标转换。

XX地方2000坐标系		4PReprojector转换后		误差		
xi	yi	Xb	Yb	$\Delta X$	$\Delta Y$	d
3168022.964	490679.918	3167436.584	494603.331	-586.380	3923.413	3966.990
3162319.854	500719.343	3161721.054	504635.665	-598.800	3916.322	3961.836
3156108.878	498852.293	3155512.412	502760.944	-596.466	3908.651	3953.899
3157397.983	455027.094	3156855.673	458937.156	-542.310	3910.062	3947.491
3164714.363	456989.856	3164169.71	460908.991	-544.653	3919.135	3956.800
3159845.649	460495.063	3159296.607	464408.211	-549.042	3913.148	3951.478
3169406.406	466713.294	3168849.748	470638.338	-556.658	3925.044	3964.321
3166685.399	474274.248	3166119.35	478195.968	-566.049	3921.720	3962.360
3155750.349	470412.349	3155189.014	474320.518	-561.335	3908.169	3948.276
3172032.704	479239.444	3171460.52	483167.817	-572.184	3928.373	3969.824
3163488.264	478047.594	3162917.527	481965.371	-570.737	3917.777	3959.131
3168671.962	484476.192	3168093.274	488400.408	-578.688	3924.216	3966.654
3163352.99	486796.69	3162771.421	490714.318	-581.569	3917.628	3960.560
3159258.984	482379.169	3158682.875	486291.726	-576.109	3912.557	3954.745
3158225.785	492037.801	3157637.734	495949.085	-588.051	3911.284	3955.243
3174884.483	492865.732	3174295.382	496797.657	-589.101	3931.925	3975.811
3164542.878	496408.887	3163949.406	500327.977	-593.472	3919.090	3963.770
3175161.99	498699.454	3174565.642	502631.708	-596.348	3932.254	3977.217

图 3-10: 4PReprojectpr 转换成果精度校验

#### 4. 结束语

随着天地图平台的升级,今后平台的底层检索数据会逐渐由点类数据向复杂结构数据过度,将会支持点、线、面、实体等不同数据结构的数据,方便用户快速检索,更加直观地浏览。

FME 作为一个强大的数据转换、处理工具,它的应用已经延伸到地理信息行业的方方面面。在天地图的运维工作中,简到格式的转换,繁到构建复杂的数据判重、数据融合模型,处处都可以见到 FME 的影子。也有越来越多的用户开始使用 FME 去编写程序,相比较传统开发模式,FME 更加容易上手且便捷、灵活。在每年的 FME 软件升级过程中,会出现很多新的功能模块,也对原功能模块进行了优化,带给用户丰富的体验。作为一个日常频繁使用 FME 的用户,真心希望 FME 技术能被更多的人发掘,服务于更加广泛的群体以及个人,发展地越来越好。