

# 一路上有你

—我和 FME 的故事征文 tsonghua/文

## 0、前言

笔者自 09 年 GIS 专业毕业迄今都奋战在国土行业。一路走来，项目上经历了全国第二次土地调查、农村集体土地所有权调查、全国宗地统一编码以及规土一张图数据中心建设等工作分工，也从数据整理到系统开发再逐步接触项目管理，现在主要从事产品研发。

国土行业算是信息化和 GIS 应用都比较早也比较成熟的行业。从“十五”规划提出数字国土工程，开启了信息化进程；“十一五”规划实施金土工程一期，与此同时在全国范围推进第二次全国土地调查，进一步夯实基础；到了“十二五”，金土工程二期提出，站在之前的基础上，建设“一张图”数据中心，更好地利用 GIS 数据，进行更精细地管理。

十几年来，行业里国内外优秀 GIS 平台和软件辈出。一片红海的背景下，笔者从业的几年间还是遇到无法避免的各种问题。数据方面基本是贯穿采集整理入库的全过程，包括多源异构数据采集整理、数据成果质量控制、大量数据入库更新；而系统和产品，最终的目标还是通过程序的设计开发，实现以更规范的流程和更简便的方式来维护和沉淀数据，所以在系统开发测试和运维的各环节，依然离不开对数据的处理。

FME 作为空间数据 ETL 利器，从学校时期最初接触到工作的各阶段，都一直陪伴在左右，下图是钉在 Excel 上的《FME 功能函数对照表》，作为“葵花宝典”时不时翻翻。

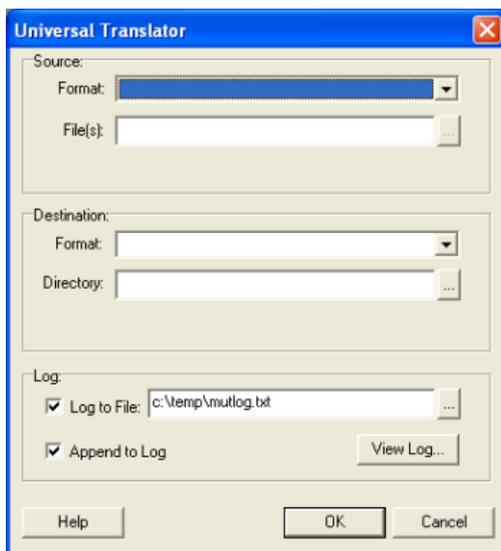


下面通过具体案例，来讲述我与 FME 的相遇、相识、相知的过程，以及其中点点滴滴的故事。由于水平有限，难免不足和谬误之处，万望指点。

## 一、相遇

### 案例一、mid\mif 文件转为 dwg

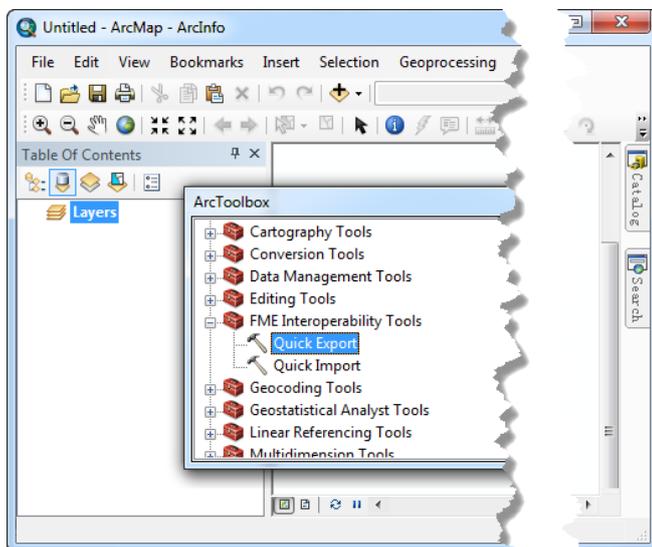
这个问题是念书时候做 GIS 实验里提到的，当时用的是 MapInfo。解决办法很简单，MapInfo 菜单下依次选择 Tools\Universal Translator，使用通用转换器来执行即可。



关键字：MapInfo、数据格式转换、Universal Translator

### 案例二、e00、dwg 文件转为 Shapefile

这个问题的背景，是在大三暑假实习时在 ArcGIS 平台遇到的，e00、dwg 等格式的 gis 数据，导入到 Esri 的 shapefile。解决办法同样很简单，在 ArcMap 的菜单里，勾选 Data Interoperability 扩展，然后使用 ArcToolbox 的 Quick Import 工具可以快速将其他格式导入到 ArcGIS 里。



关键字：ArcGIS、数据格式转换、Data Interoperability

## 总结：

无独有偶，透过上面两个案例，发现行业两款优秀软件 MapInfo、ArcGIS 里，居然都有通用转换的工具，而且界面如此类似！当时经过师长的介绍，原来他们使用的都是 FME 的格式转换解决方案，果然是“尊重专业”的节奏，专业的事就该交给专业的工具来完成。

## 二、 相识

学习过程的乐趣，除了新知识的获取、新技能的锤炼，还有新朋友的结交。GIS 圈子里认识的一位好朋友 lucy1114 的《【学习笔记】 SpatialETLTool (FME) 空间数据交换引擎-抛砖引玉》，开启了我对 FME 更深层次的认识和探索——Workbench。

### 案例三、重复点删除

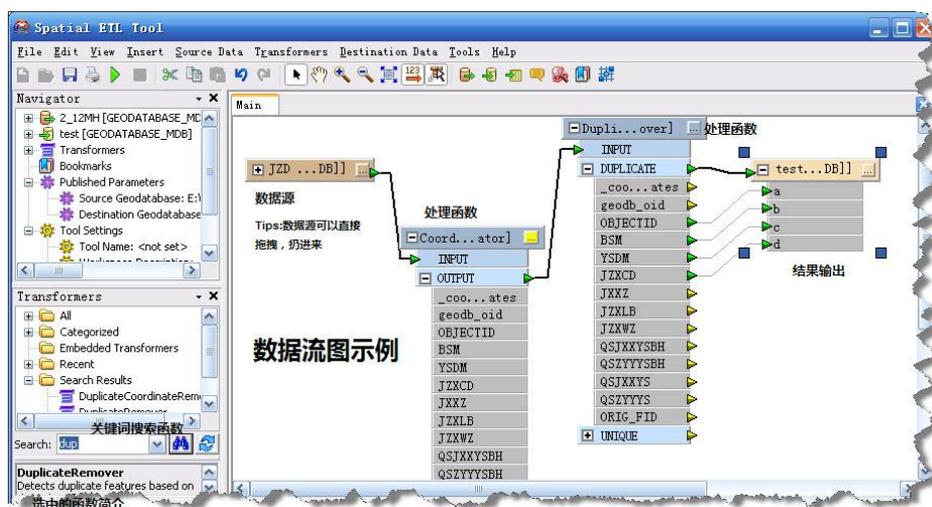
问题背景：全国第二次土地调查的成果数据需要经过质检软件检查，其中有个检查规则是不允许有重复点。

解决思路：1、将点坐标作为字段生成，格式为  $x,y$ ；2、根据属性相同，删除重复记录  
使用到的函数：

1. CoordinateConcatenator：生成点的坐标，作为属性字段

2. DuplicateRemover：根据重复属性删除点

最终 workbench 里如下图所示



关键字：重复点、DuplicateRemover、FME Workbench

此案例的完整文章，参考附件《01 删除重复点方法探讨.doc》

#### 案例四、图斑自上至下、从左到右编号

问题背景：全国宗地统一编码项目中，因权籍业务要求，对宗地图斑的编号需要满足一定的规则，也就是自上至下、从左到右编号（也有是 Z 字形编号规则，不含在本文实现的方法里）。

解决思路：1.获取图斑的某一特征点；2.根据此点的 X、Y 值来排序；3.将序好结果挂回去。

使用到的函数：

1. BoundingBoxReplacer：得到图斑的外接矩形

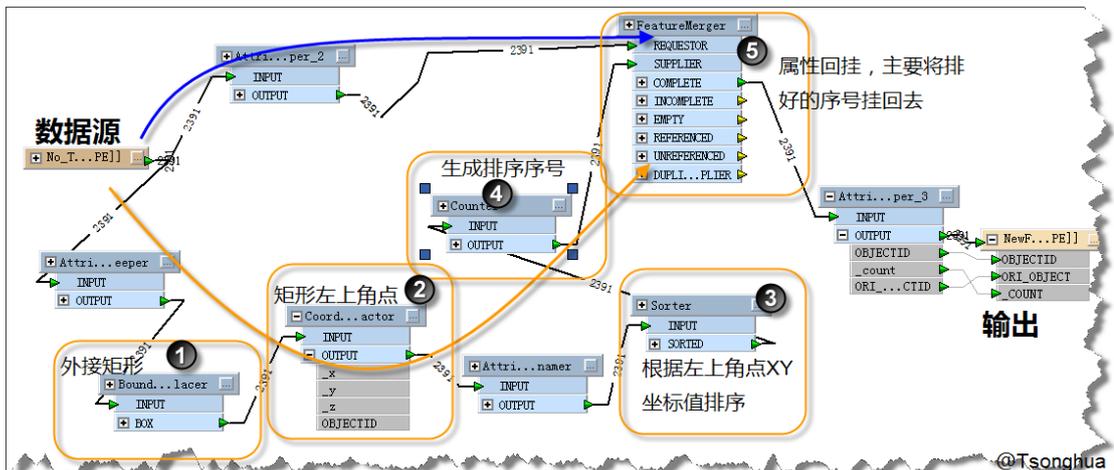
2. CoordinateExtractor：得到外接矩形左上角点坐标（该工具参数里，index 设置为 1 就是左上角点；设置为零对应矩形左下角点，其他以此类推）注：此为 fme2010 内函数，之前版本函数名为 CoordinateFetcher

3.sorter：排序，根据上一步的到的左上角点坐标 Y 倒序、X 升序

4.counter：计数器，根据排序结果生成新的序号

5.FeatureMerger：相当于 ArcMap 里的 Join，这里将排序好新生成的序号挂回去。

最终 workbench 里如下图所示



关键字：宗地、自上至下从左到右、编号、FME Wrokbench

此案例的完整文章，参考附件《02 图斑自上至下从左到右编号方法再思考.doc》

总结：

Workbench 是 FME Desktop 的精髓。对于一个棘手的问题，关键是厘清处理思路；思路清晰之后，可以从“宝典”里找对应的函数模块。对于不会编程的 giser 可谓是最大的福音，FME 直接提供功能完备的处理片段，自己组装即可。很像是小朋友搭积木，用一个个合适形状的小积木，一个个摆上去就能搭建起自己心目中的“建筑”。

与 ArcGIS 的 ModelBuilder 相较，FME Workbench 在功能函数的拆分粒度上更加精细、提供各种快捷的操作来提高效率，比如“Copy Attributes From FeatureType”快速从输入或者其他输出来复制整个 schema、“Auto Connection”快速自动做字段映射（在新版本里，同名的字段自动 connect 不需要重新连一把，这个更加便捷）。

### 三、 相知

涂子沛先生在其著作《大数据》里这么描述：

所谓的“信息管理系统”也就是实现某一特定业务管理功能的软件。软件的构成，主要有两部分，一是程序(也可称为代码)，二是数据(或称为数据库)。程序和数据的关系，就好像发动机和燃料，所有的程序，都是靠数据驱动的；数据之于程序，又好比血液之于人体，一旦血液停止流动，人就失去了生命，代码也将停止运行。数据的生命力，甚至比程序更持久。程序可以不停地升级、换代甚至退出使用，但保存数据的数据库却会继续存在，其价值很可能与日俱增、历久弥新。

数据是 GIS 的血液，更是整个信息化的核心。一张图数据中心项目实施过程，主要包括了数据建库、管理系统以及数据更新三部分。在负责整个项目管理的时候，我发现除了基础的数据转换和数据处理之外，FME 还能提供更多的惊喜。

#### 案例五、国土部土地复垦数据抽取、构面

问题背景：建库的时候，有一类业务数据是从国土资源部农村土地整治监测监管系统里下载的 excel 来自动构面和处理属性。原始数据格式如下：

	A	B	C	D	E	H	I
1	[属性描述]						
2	格式版本号	①	数据生产单位	国土资源部	数据产生日期	地坐标系	几度分带
3	投影类型		计量单位	米	带号	0.01	转换参数
4	[地块坐标]						
5	16.6.4313991010.1_地块1_面...						
6	1	1	12.329		49.23		
7	2	1	13.158		63.63		
8	3	1	14.653		68.44		
9	4	1	17.488		10.45		
10	5	1	14.677		85.48		
11	6	1	172.22		79.68		
12	7	1	135.241		78.65		
13	8	1	130.819		69.88		
14	9	1	170.665		64.78		
15	10	1	175.667		64.14		
16	11	1	169.28		700.8		
17	12	1	164.74		89.48		
18	13	1	172.698		65.03		
19	14	1	11.297		36.37		
20	15	1	10.399		41.26		
21	1	1	12.329		49.23		

Excel 文件结构，包括三部分：

- 1、头文件，介绍属性描述、投影类型、精度等
- 2、属性信息
- 3、图形信息，包括点序号和 x、y 坐标值

其中，2 和 3 共同组成一个图形。经分析，发现有的复垦文件，里面不止一个

	A	B	C	D	E
1	[属性描述]				
2	格式版本号		数据生产单位	国土资源部	数据产地
3	投影类型	高斯克吕格	计量单位	米	数据产地带号
4	[地块坐标]				带号
5	33, 0. 5197418040, 1, 沪				
6	1	1	6. 334	46. 72	
7	2	1	6. 657	50. 35	
8	3	1	5. 097	50. 93	
36	31	1	5. 526	30. 92	
37	32	1	0. 476	332. 9	
38	1	1	6. 334	46. 72	
39	5, 0. 0094897280, 2, 沪				
40	1	1	6. 297	02. 86	
41	2	1	9. 982	92. 79	
42	3	1	6. 004	88. 57	
43	4	1	3. 394	98. 65	
44	1	1	6. 297	02. 86	
45	10, 0. 2432807790, 3, 沪				
46	1	1	0. 464	10. 76	
47	2	1	31. 424	64. 65	
48	3	1	5. 484	54. 54	
49	4	1	7. 216	47. 52	

解决思路：1、构造分组字段，来区分不同的图形；2、构造图形；3、构造属性，批量处理所有 Excel。

使用到的函数：

1.StringSearcher：这个转换器，旨在找出每个图形的属性部分，特征是以 “@” 结尾，所以用了正则表达式 “(\*)@”；

2.Counter：将属性过滤出来后，用这个转换器来计数；

3.VariableSetter：这个是关键的关键。fme 读取数据的时候，是自上而下读的。所以 counter 计数的是属性的排序号，其实也是这个属性对应的所有图形的序号。

将此序号设置为 Global，全局通用，这样就保证了分组的实现；

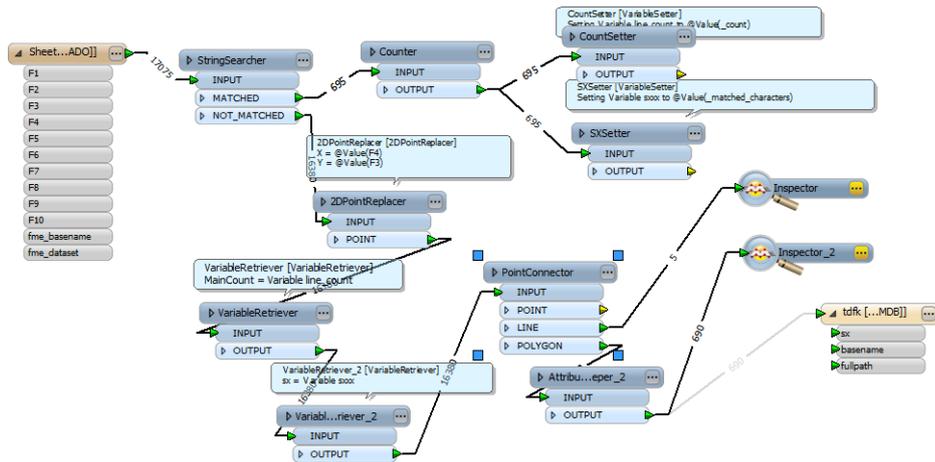
4.VariableReceiver：接收上面 VariableSetter 转换器的结果。挂到点上，进而挂到图形上。

5.fme\_basename：获取 Excel 的文件名；

6.2DpointReplacer：参数是 X、Y。这个是 Excel 的 D 和 C 两列

与此同时，整个转换方法，一个是计数器需要设置，目的是为了展点后构面的分组依据；

Workbench 里设置如下：



关键字：数据中心、建库、FME Workbench、分组、全局

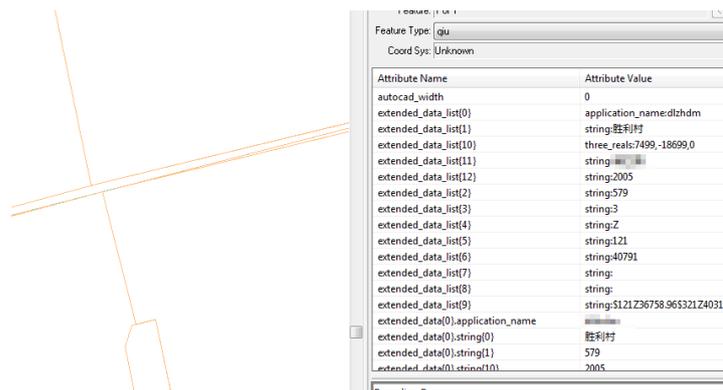
参考材料：[使用 FME 将勘测定界坐标交换格式文本文件转换成 GIS 格式](#)

### 案例六、带 EED 属性的 CAD 与 shapefile 互转

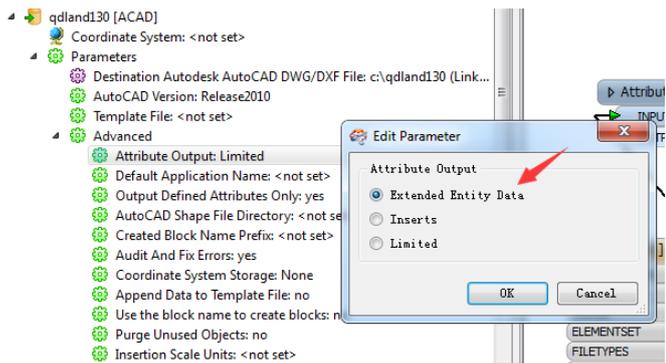
问题背景：数据采集目前还是 CAD 占主流，所有在 GIS 项目建设过程中无法避免与 CAD 的互转操作。而 CAD 里存储属性，除了注记、属性块，还有 EED (Entity Extended Data 实体扩展属性)。

解决方案：

- 1、读 EED 属性，一般是借助 FME Universal Viewer 查看，看看属性在 Extended\_data\_list {} 什么位置，以及有多少个，然后在读取的时候，AttributeExposer 暴露出来，右键展开即可。



- 2、写 EED 属性，更简单。设置 Attribute Output 即可。



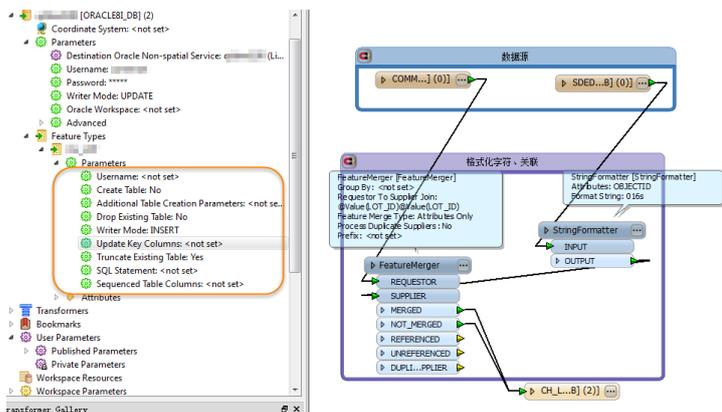
不过这里有个缺陷，就是 Application\_ID 没法根据不同的属性来设置不同值。

关键字：CAD 转换、EED 属性

### 案例七、Oracle 属性数据联查更新

问题背景：数据迁移中遇到需要将 B 表的 OBJECTID 字段格式化后根据关联字段更新到 A 表的问题，两张大表联查更新 (Update)，SQL 语句各种优化还是执行了十几个小时没结束。突然想到，空间数据 ETL 工具 FME，也能处理非空间数据，尝试下，效果很好。

解决方案：读写数据源都是 Oracle，连接方式也是常规的，暂按下不表。令我感到惊喜的是，在写模块的设置里，对表是否删除、重建、Truncate、Update KEY 等等都有非常人性化的设置，因为之前表里已经有各种所以、分区等，所以必须设置为不删除、不重建、处理前需要 Truncate 等，尝试如下，



### 案例八、抽取 Oracle 坐标串构面

问题背景：在开发的过程，有个设计是将拉框的范围的坐标串插入 Oracle 某表的字段里。测试人员想验证结果的正确性，需要反向将坐标范围生成范围。对于没有编程经验的人员而言，这个问题就显得比较棘手。

解决方案：FME 读取 Oracle 内容，通过 StringResearcher 来分割出坐标组、XY 值，继而



## 五、 尾声

以上洋洋洒洒许多文字，想述说的是一个 FME 使用者的切身感受。FME 并非能包治百病的膏药，而是在遇到合适的场景，提供一个可能的解决方案。

感谢，一路有你，且行且珍惜。

## 附、联系方式

作者：雷中华（网名 tsonghua、tj051181）

单位：上海南康科技有限公司

邮件：[gis.lei@qq.com](mailto:gis.lei@qq.com)

[tj051181@gmail.com](mailto:tj051181@gmail.com)

地址：上海市北京西路 95 号东楼 503 室 #200003

电话：182 2138 5328

## 附、已发 FME 文章清单

01 删除重复点方法探讨.doc

02 图斑自上至下从左到右编号方法再思考.doc