

FME 在数据整合中的应用

一、项目背景

自 2013 年 11 月 20 日不动产统一登记制度正式实行以来，各地持续推进不动产存量数据整合工作，将现有的分散存放在国土、住建等部门的多源异构的不动产登记数据进行规范采集、抽取、转换、补录、整合，依据《不动产登记数据库标准》（试行）整合建设不动产登记数据库，此项整合已基本接近尾声，FME 在此项数据整合中发挥了巨大作用。

根据国土资源部关于进一步加快宅基地和集体建设用地确权登记发证有关问题的通知，2020 年将开展大范围的农村房地一体确权登记工作，新的整合工作即将开展，FME 将再次在此类项目中发光发热。

本文主要根据不动产数据整合的经验，介绍了 FME 在数据整合中的应用。并对接下来要进行的农村房地一体确权登记项目有一定的展望。

二、数据整合总体流程

不动产登记数据整合过程

不动产登记数据库整合建库主要步骤包括：准备工作、数据采集分析及规范化整理、数据整合关联和数据入库，质量控制贯穿不动产登记整合建库的全过程。

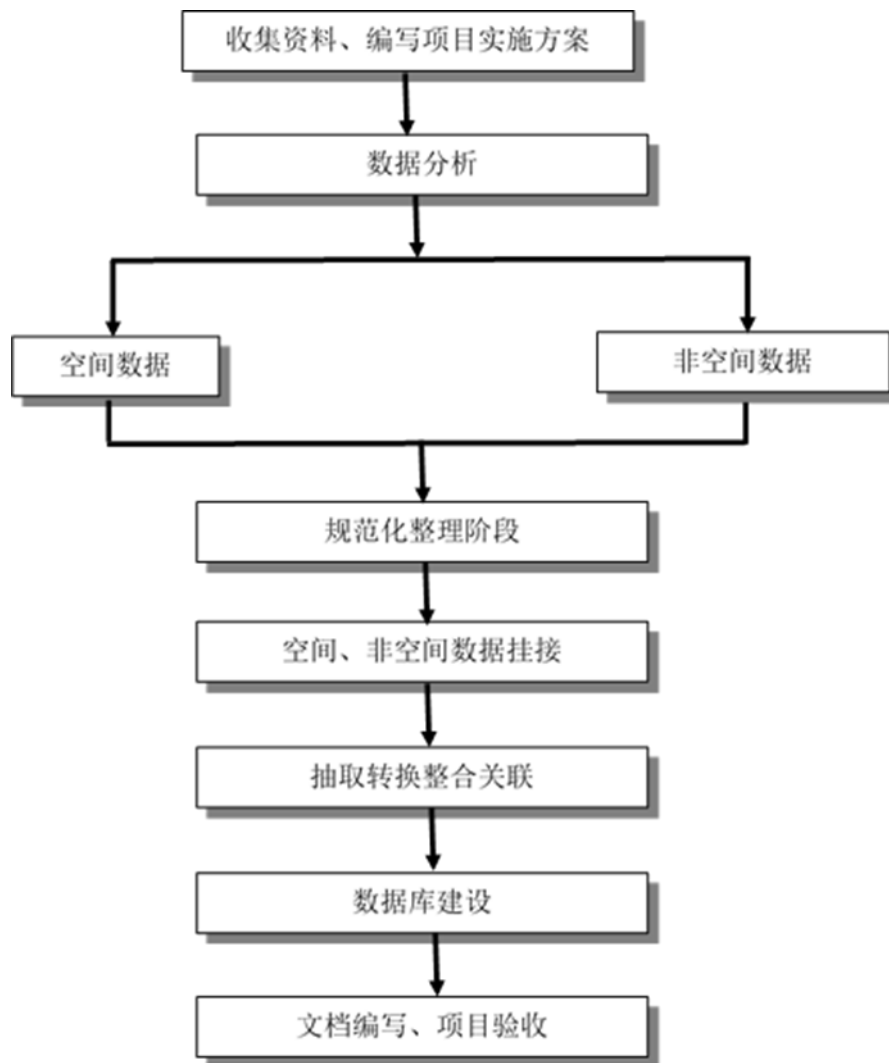


图 1 数据整合流程图

三、技术设计

数据分析及预处理

在数据整合开始之前，我们一般会收集已有的数据成果，数据成果是多源异构的，我们通过 FME 对数据进行前期分析，并进行初步处理。

这一步使用主要使用的转换器是 Joiner 和 FeatureMerger。通过关键字段对原始数据库的多张表进行关联查询，找到表与表之间的关系，分析出表格存放的数据种类。Joiner 和 FeatureMerger 转换器都可用于表与表进行关联，Joiner 转换器是阻塞性转换器，适用于少量数据在大批量数据中进行关联查询，FeatureMerger 适用于大批量数据和大批量数据关联查询，灵活结合使用有利于提高模板运行效率。

得到表关系后，我们通过 AttributeCreator 转换器对原始字段和目标字段做映射关系。Sorter 和 Sampler 转换器对原表重复数据排序取唯一数据。最后对每条数据通过 UUIDGenerator 取一个唯一标识。

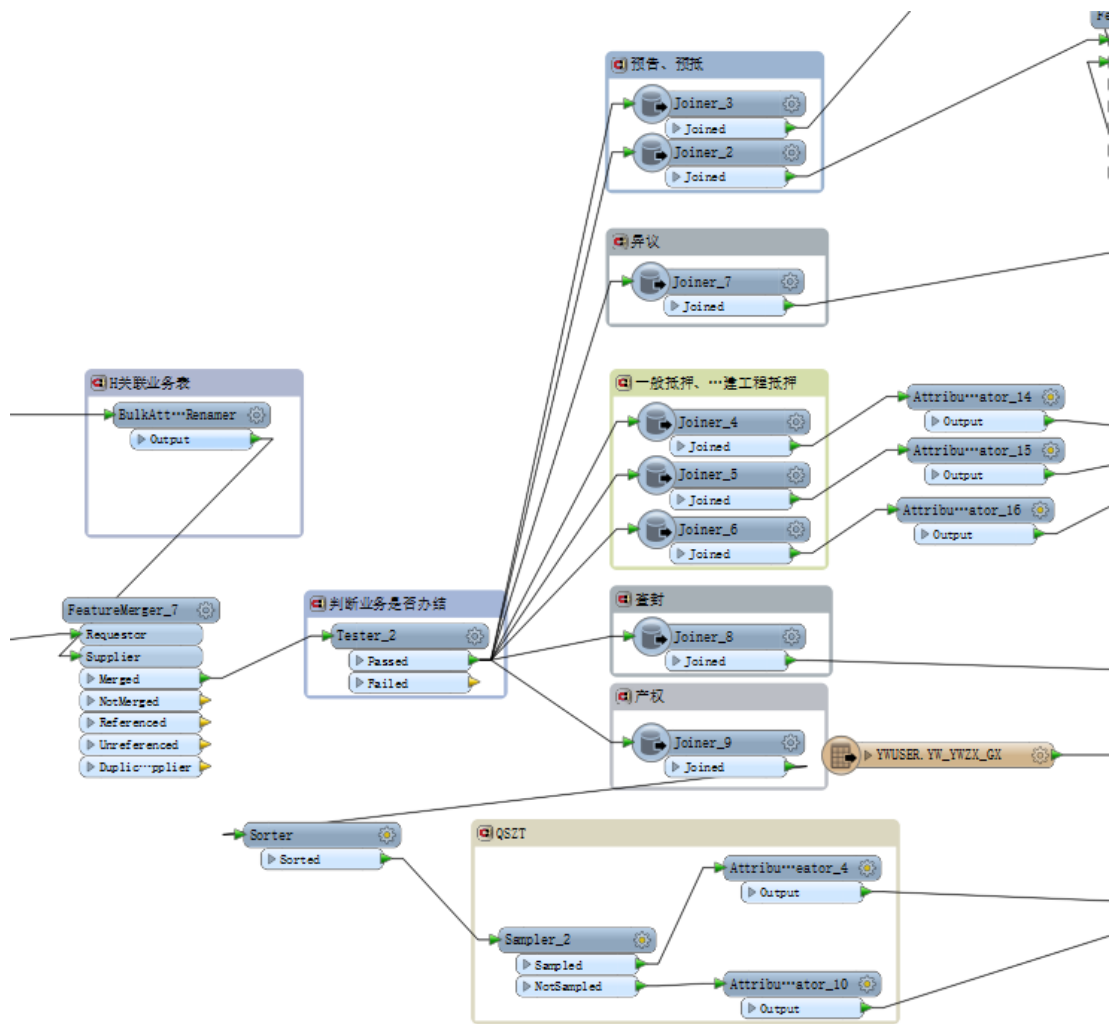
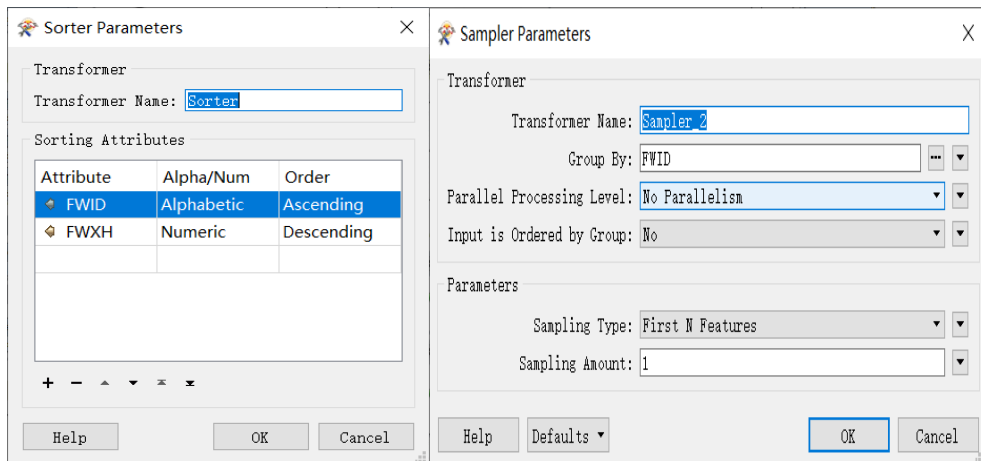


图 1 部分模板截图



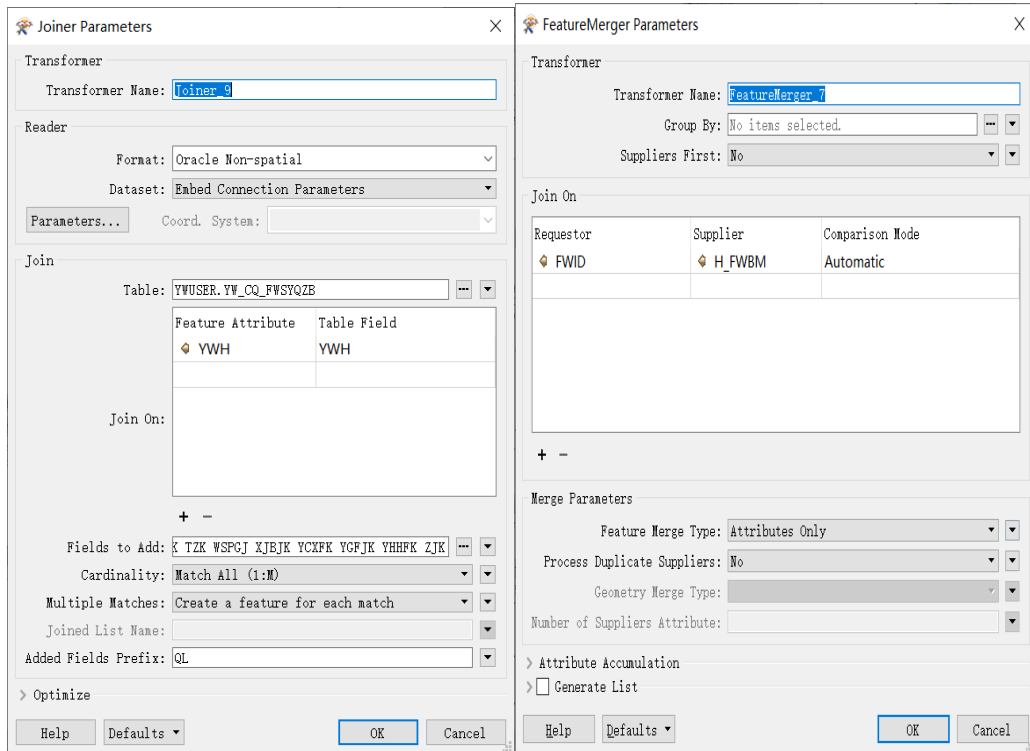


图 2 部分转换器参数截图

规范化整理

规范化整理阶段可以通过 FME 对有规律的数据进行批量整理，提高工作效率。使用 AttributeCreator 和 AttributeValueMapper 转换器规将原本杂乱的数据转化成国标字典值。如权利人证件类型，包含户口的归类为户口簿，户口簿对应的证件类型字典值 4。

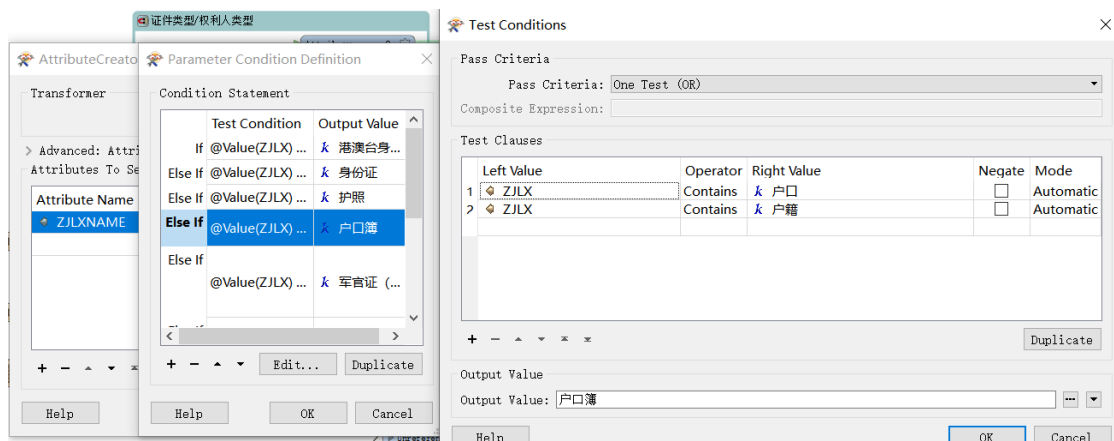


图 1 AttributeCreator 转换器参数

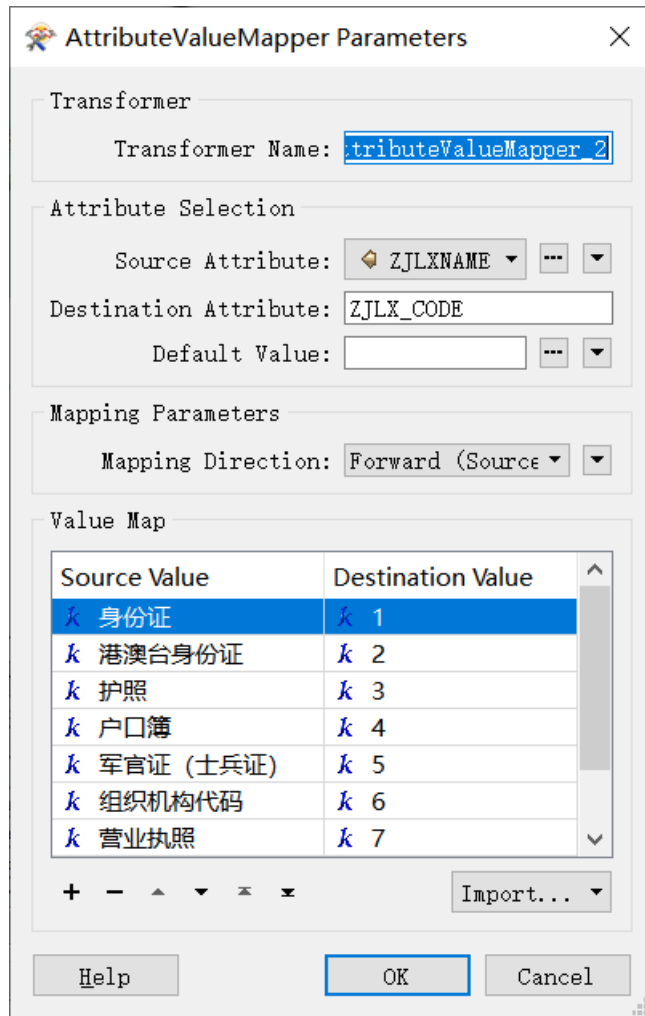


图 2 AttributeValueMapper 转换器参数

档案抽取

原数据库电子档案是通过二进制形式存储在数据库中的。这一步可以先通过 AttributeCreator 转换器新建一个 PATH 变量作为存储路径。然后通过 AttributeFileWriter 转换器解析写出档案图片。

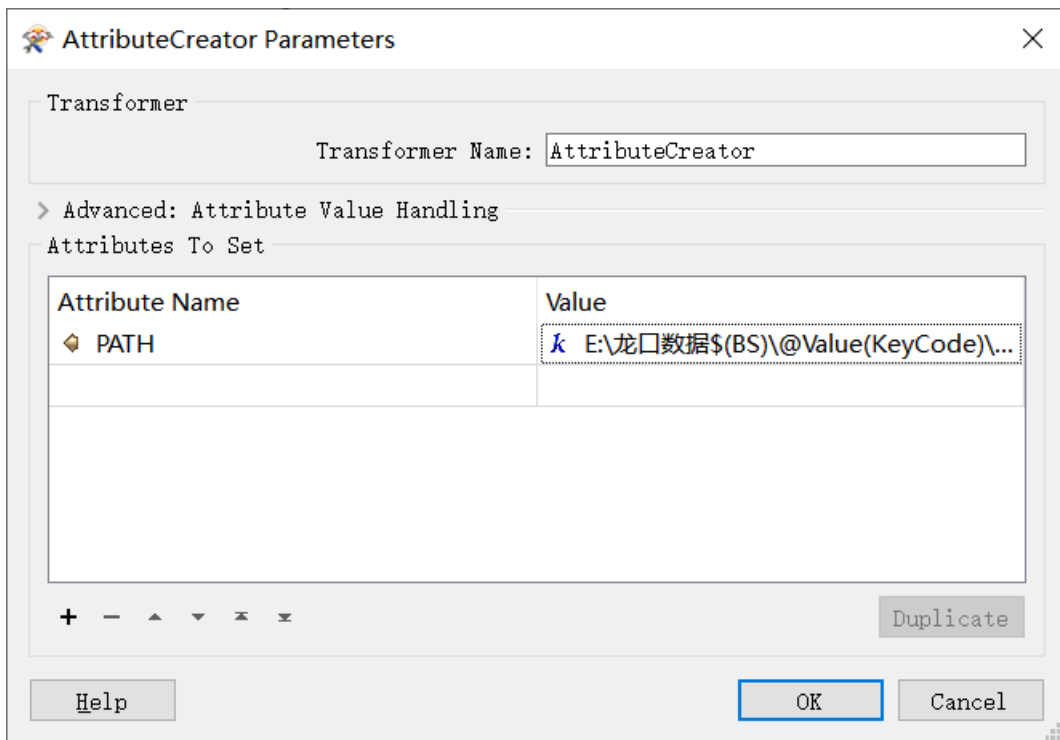


图 1 AttributeCreator 转换器参数

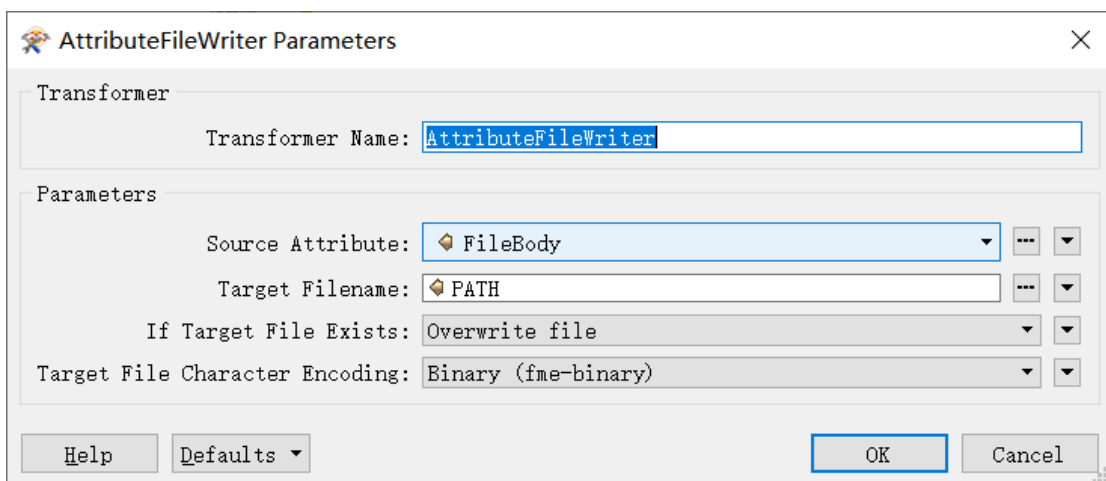


图 2 AttributeFileWriter 转换器参数

四、FME 在农村房地一体确权登记项目中应用的展望

农村房地一体确权登记项目主要包含两部分工作：外业调查与测绘，内业数据整合。在处理内业数据时，FME 必将派上用场。

使用 FME 的空间数据处理相关转换器对外业测绘成果进行预处理与叠加分析；运用不动产数据整合所用到的相关转换器对外业调查数据进行规范化处理，实现数据整合；利用 FME 对整合成果进行入库，实现农村房地数据与已有不动产数据融合；最后根据数据情况，利用 Directory and File Pathnames 读模块或其他文件读取转换器对电子档案进行获取，并完成挂接。

五、结语

FME 是一个强大的工具。在对数据批量操作的时候，首先想到的就是 FME，FME 使我的工作更高效，让枯燥的数据处理分析工作充满了更多乐趣。最后希望能借这个平台认识更多数据大牛交流数据分析处理经验。