# 基于 FME 在不动产存量数据清理中的应用

## 一、项目背景

房产登记档案补录是不动产存量数据清理项目中最为复杂和繁琐的工作之一。补录工作需要通过不动产单元号、地址和权利人等关键信息在查阅电子扫描件档案，核对原有表格信息的完整性，并根据最新的档案查漏补缺。

房产登记档案原始数据主要由 FDCQ2（房地产权）、H（户）、SH（审核）、FZ （发证）、SJ（收件）、GD（归档）、SQR（申请人）、QLR（权利人）等表格组成，表格之间通过共同的关键字段 YWH（业务号）或者 BDCDYH（不动产单元号）进行关联。

因为项目工期紧张而且数据量巨大，为了提升档案补录的工作效率，加快项目进度。我们制定了利用 FME 对数据先合并再拆分最后入库的技术方案。

## 二、总体流程

房产登记档案补录的整个处理过程主要分为四个流程：

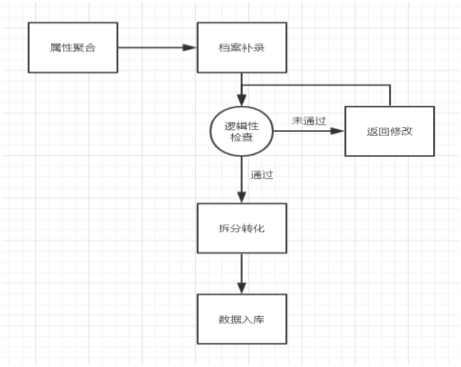
1、属性聚合；

2、档案补录；

3、质量检查；

4、成果表拆分转化入库；

整体流程图如下：



# 三、技术设计

## 1、属性聚合

这一步使用的转换器主要是 Aggregator，通过“YWH”字段进行分组将“YWH”相同的字段的其他属性聚合在一起，并用分割符“//”隔开。然后使用 FeatureMerger 转换器和其他所有表格关联挂接在一张属性表格。其中在审核阶段中，同一个业务号会对应多个审批阶段，每一个审批阶段都会产生节点名称、审核意见、审核人员和审核结束时间。为了方便补录档案，将同一类属性合并在一起。

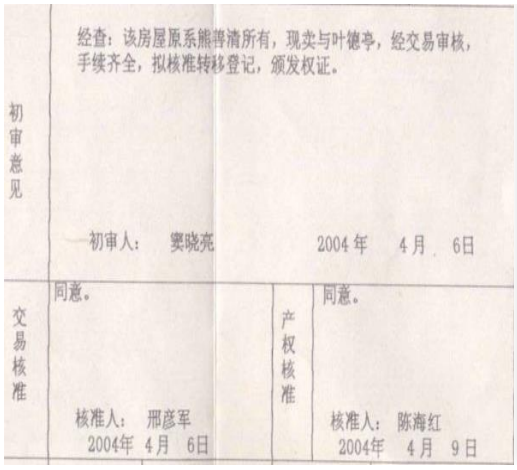



图 1 审核表部分截图

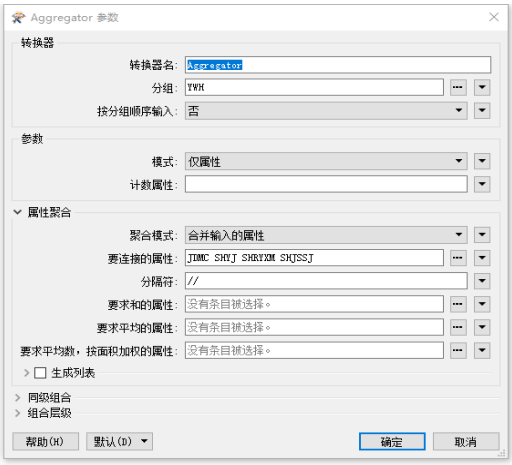



图 2 转换器参数

## 2、质量检查

房产登记档案补录的过程中，由于同一类的多条记录合并在一个字段中，可能会由于人为的失误，造成如下的几类错误：1）属性信息填写错误。审核表中的数据必须一一对应，一个审核人对应一个审批意见和审批日期，比如审核人员有三个，审批意见只填写了两个；2）时间填写错误。由于内业工作人员的疏忽，导致日期没有填写完整或者填写错误。比如 2019-3-20 错写 2019-3 或者 2019-20。

质量检查的用到的转换器有 AttributeSplitter、ListElementCounter 和 Tester。设计思路如下：

（1） 通过 ListElementCounter 转换器获取列表属性的总成员数

（2） 然后用 Tester 转换器判断成员数是否一致，不一致的数据输出到不合格的表格中。

（3） 如果一致，继续通过 DateTimeConverter 转换器，转为 FME 标准格式的日期型数据。如果时间转换失败则说明日期填写存在问题需要返回修改。

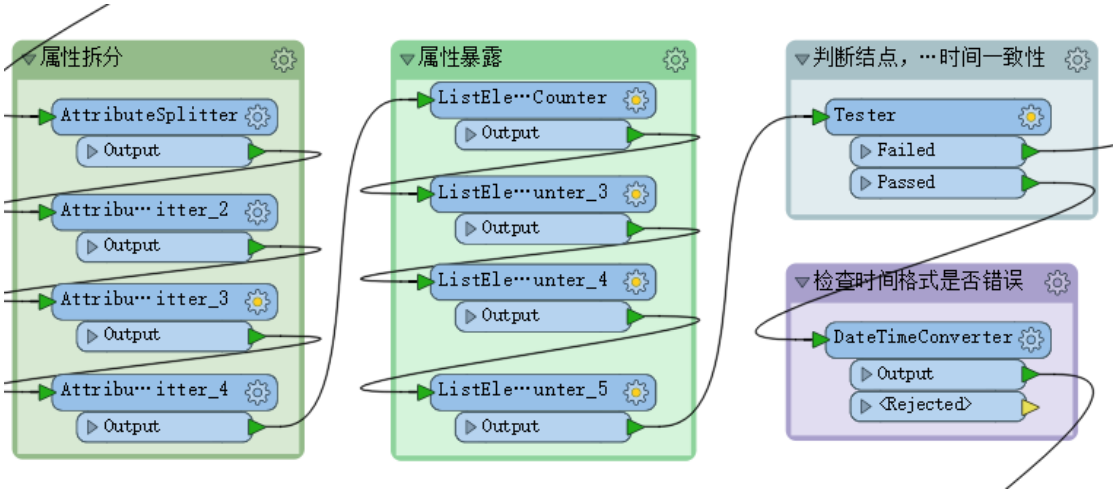



图 3 质量检查转换器

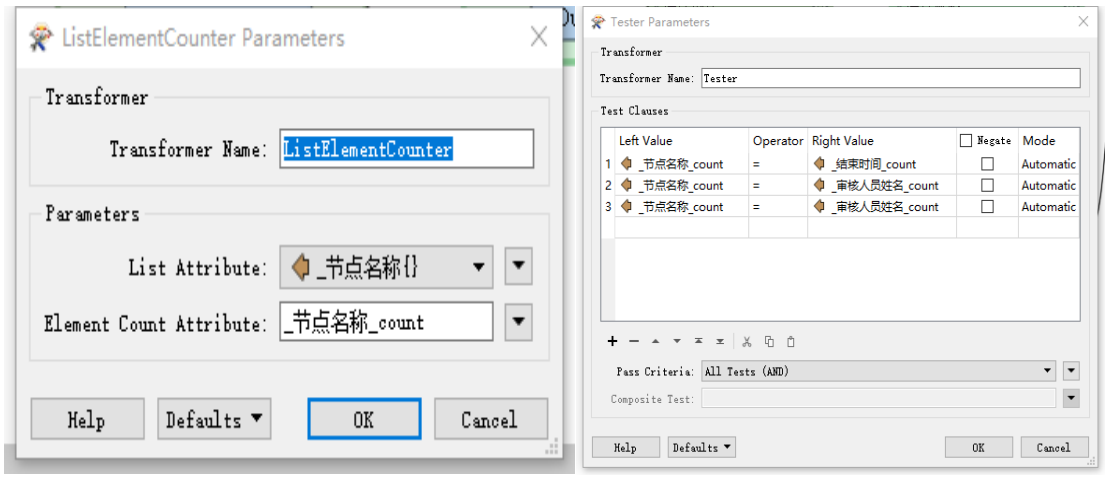



图 4 转换器参数

## 3、成果表拆分

房产登记档案补录的成果表拆分为原始表的标准格式，需要使用的转换器有 AttributeSplitter，ListExploder，FeatureMerger 等。

具体思路如下：

（1） 首先拆分属性，使用用“//”分隔符将 JDMC、SHYJ、SHRY 和 SHJSSJ 四个聚合在一起的属性拆分；

（2） 然后暴露属性，用 ListExploder 转换器将列表中的所有属性都变为输出要素的非列表属性。此外，在原始列表中该属性的元素索引编号将作为属性添加到要素中。

（3） 最后属性挂接，通过 FeatureMerger 转换器将“YWH”和“element index”作为匹配的约束条件，多次挂接过完成数据表的拆分。

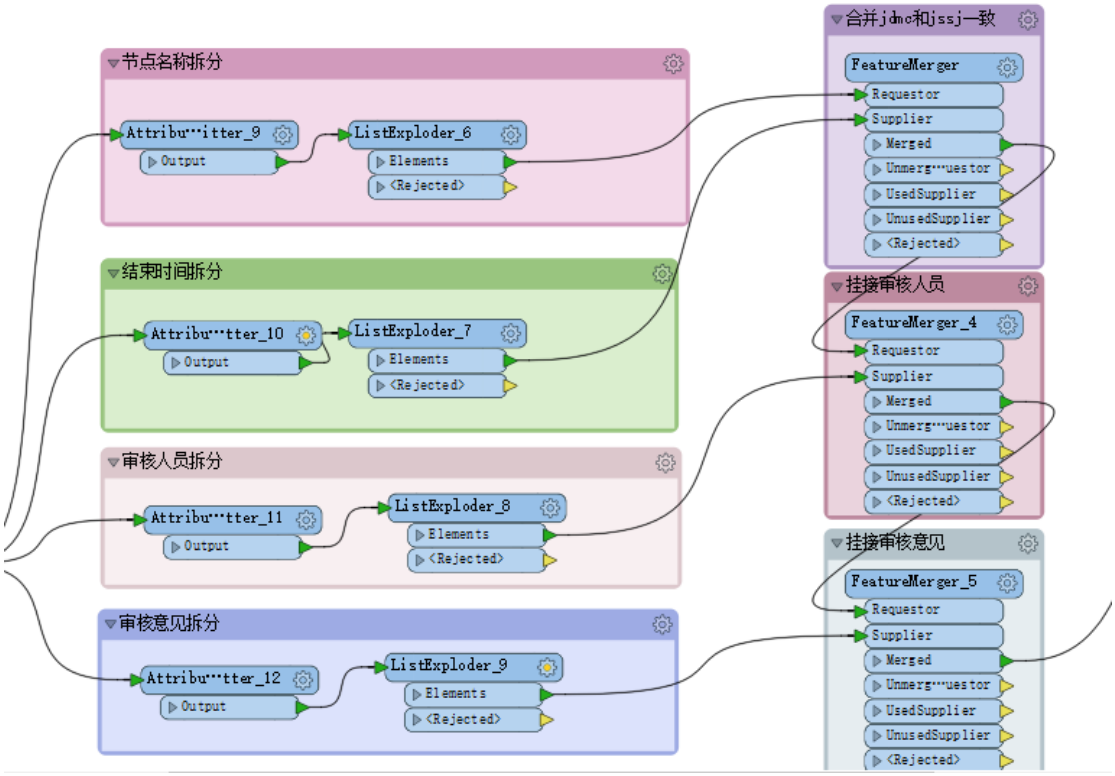



图 5 表格拆分转换器

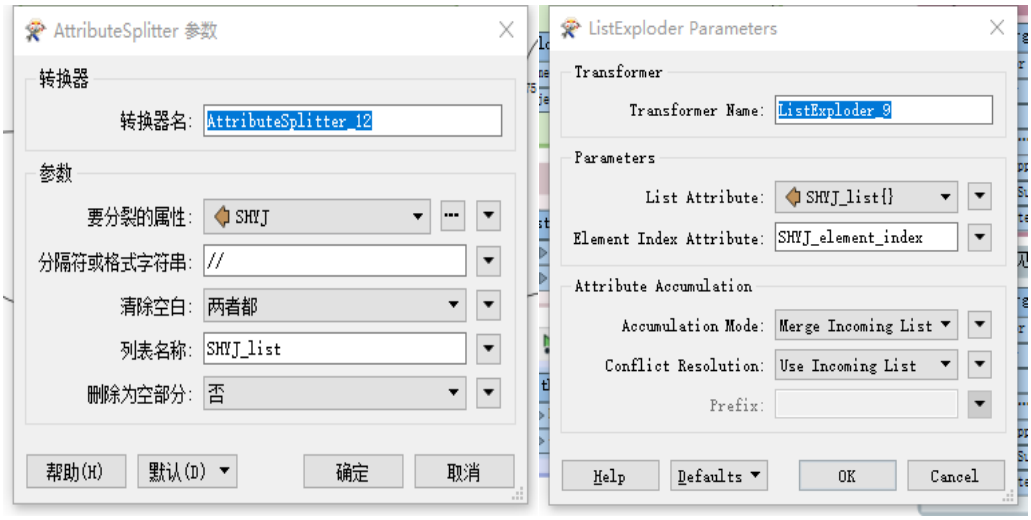



图 6 转换器参数

# 四、结束语

整个处理流程包含了数据合并，批量赋值、属性挂接、正则表达式匹配、字典映射、条件判断、拆分、逻辑性检查等，总计 110 多个转换器。FME 提供的方案实在是简单、易用、可靠、0 代码，对于没有编程基础的数据处理人员来讲是个强大的工具。在不动产存量数据清理的实际使用中，搭配多核高性能的工作站，运用 FME 的并行处理功能，极大的提高了不动产存量数据清理整合的效率。