

FME 在不动产存量数据整合中的应用实践

引言

2016 年，不动产统一登记全面“落地”；不动产存量数据整合工作作为不动产登记信息平台推进的难点，其具有数据量大、多源异构、介质不一及历史沿革问题产生的数据质量参差不齐等诸多难题。而 FME 可谓不动产存量数据整合的“神器”，它提供了完整、强大的空间 ETL 解决方案，数百种空间数据格式的转换与数据重构等强大功能。

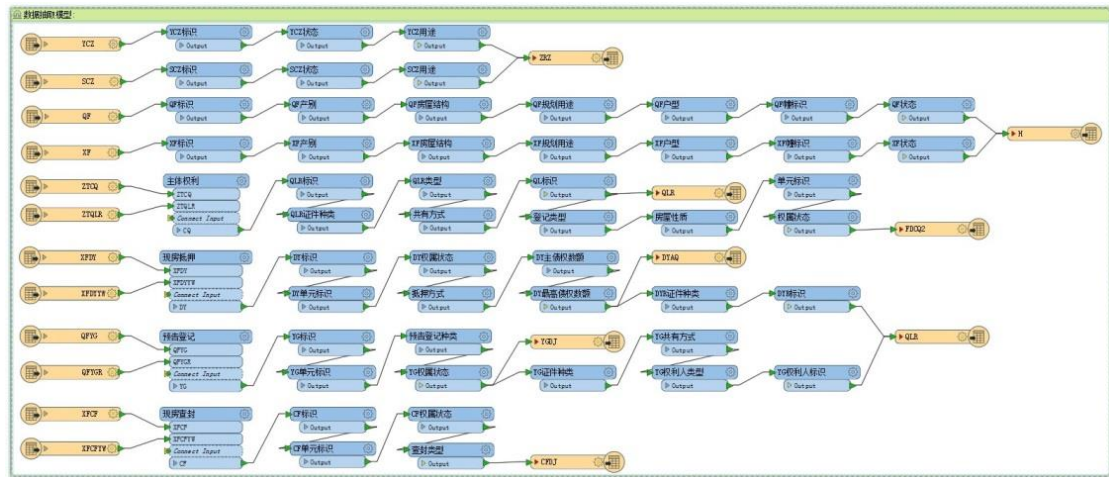
数据对象

按照国土资源部《不动产登记数据库标准》以及《不动产登记数据整合建库技术规范》规定，存量不动产登记数据是指实施不动产统一登记之前形成的土地、房屋、林地、草原、海域等不动产登记数据，包括空间数据、属性数据和其他数据。

一般各地先整合房地数据：从纵向理解，以宗地数据为底图->开展幢落宗->搭建楼盘表->建立唯一不动产单元；从横向理解，基于不动产单元审批的所有权、用益物权、担保物权、其他不动产权利与权利人等业务挂接与清理；形成满足登记发证和信息查询与共享等要求的不动产数据库。

数据抽取模型

通过 FME 建立存量数据抽取模型，可以将上述多种格式数据进行读取、梳理、关联、重构、清理并将规范化处理后的数据传输至目标数据库中；下面笔者以逻辑关系最为复杂、数据量最大、群众办理业务最关注的房产数据为例，浅谈一些基于 FME 处理不动产存量数据的实践体会。如图：



某省会级城市房产数据抽取模型截图（Oracle 数据源转国标 mdb）

- **数据抽取方面**

相较于利用编程语言开发的抽取工具而言，其具有简洁性、周期短、图形化、性能高、易维护等优势，无论是否掌握编程技能，对于数据整合工作你都可以从容应对。

- **模型优化方面**

由于房产数据一般逻辑复杂、涉及表格较多、关联关系重现等要求，数据抽取模型应尽量高效、美观、易维护为准则；初期构建数据抽取模型往往看上去很复杂，其实也可以很简洁，这里我为了采取模型界面优化辅用了 SQL 建表语句作为预处理脚本，生成一套临时源表，达到了转换器的缩减与复用及减少连接线的效果，也不改

变原始数据。这里也可以采用 Junctions 和 Tunnels 工具进行模型界面优化。

- **模型性能方面**

处理地市级及以上城市的存量房产数据，许多表格数据量上百万条，对数据抽取的效率要求非常高，这里可做如下三点优化：1、减少多表间的数据关联与查询操作。2、运行模型前删掉库索引，加快写入速度。3、64 位 FME 版本与留有足够可用空间。

- **转换器方面**

数据抽取笔者常用转换器有 AttributeCreator：用于标识字段的创建、状态字段的赋值、其他所需字段的属性创建；AttributeValueMapper：用于字段按国家数据字典标准的映射更新、其他所需字段的属性映射更新；InlineQuerier：用于通过 SQL 语句实现的数据内容过滤与多表关联查询；AttributeFilter：用于数据属性内容为空等过滤处理；StringConcatenator：用于坐落拼接重组等数据属性连接转换；通过上述及其他转换器，从而实现单元-权利-权利人的关联、字段国标化处理、属性内容维护、数据迁移重构等多方面目标。

- **数据维护方面**

可从下面三个方面考虑：1、Writer Mode 包含 INSERT、UPDATE、DELETE 三个选项针对不同情形进行维护处理。2、根据具体数据问题维护数据抽取模型，重新读入（整表维护）、调连接线（字段维护）、调转换器（批量维护）等。3、FME 日志分析，浏览转换信息可迅速定位错误源；转换统计可做整体参考与数据量统计输出。

- **模型复用方面**

当数据抽取模型确定，各地方同一类房产数据库通过模型微调可以实现快速存量数据抽取并用于上线运行，数据质量高、节省时间、复用性强。

结束语

FME 的强大毋庸置疑，为集行业应用与大数据特性于一身的不动产存量数据整合带来了极大的便捷，产生了许多成功与高效的案例。今后我们还将结合 FME 对产业链上游的测绘数据生产及下游的数据挖掘与辅助决策分析等方面做更深入的应用实践。2017 年，让我们与 FME 携手同行，创造更多的精彩。