

FME 助推数据生态圈进化

西安市国土资源信息中心 李博

一、建设现状与困境

西安市国土资源局信息化建设起步较早，始于 1998 年全国第一次土地调查时期（注：国土系统信息化基本上都起步于这一时期）。形成的信息化建设成果在当时为西安市国土资源管理水平提升做出了重要的贡献。用现在的眼光来看，其建设方案受到当时信息化建设理念影响以及技术水平的限制，项目被人为分解为若干子系统建设，子系统之间相对独立，不同的数据寄生于各个子系统，数据间存在大量冗余和版本差异，系统建设交付后基本定型，没有大的升级，数据分析利用功能单一，参与数据利用的用户面窄，没有有效的形成以数据为核心的良性循环（生态圈）。

各业务部门之间数据交换方式依然以传统的人工方式为主，进行数据综合利用较为困难，整个系统进化速度缓慢（几乎是静止），围绕数据的衍生品少之又少。随着时间的推移，国土资源数据在各个系统中积累的越来越多，很不幸的导致了“明知有数据，就是利用难”的被动局面，致使国土资源数据这种特殊国有资产未能充分体现应有的价值。

随着经济社会快速发展，西安市国土资源管理工作面临诸多新挑战，包括国土资源数据爆发式增长、海量数据低效利用、国土资源管理更加科学、用户利用数据门槛过高、数据价值释放率低、数据生态圈发展缓慢等问题。现存的诸多子系统已很难支撑用户对数据高效利

用的普遍渴望。曾经零敲碎打式、大拆大建式、单打独斗式的信息化建设模式耗费了大量的资金、时间和管理成本，在信息化低效建设的怪圈里打转，这些问题，在新时期国土资源信息化建设中都将被视为不可取。

通过仔细分析，发现国土资源数据管理利用所面临的困境主要有“散、乱、难、窄”四个主要特点。“散”，是指大量国土资源数据长期以来，被散乱分布在国土资源管理中的各个环节、各个部门、各个系统当中，数据综合利用的过程似盲人摸象，难以全面掌握数据家底；“乱”，是指各类数据没有严格的统一标准，虽然一部分数据有国家标准，但还是存在很多非标准数据，这部分数据标准多依赖于信息系统承建方的厂标，各承建方之间的厂标却不尽相同，有的甚至难以兼容，导致宝贵的国土资源数据被封存在特定的系统中无法摆脱；“难”，是指数据综合利用的过程异常繁杂、艰难，数据的数量和利用需求增长速度远远超越了系统的升级节奏，越来越多的数据需求需要花费大量人工和时间成本来完成，专业技术人员颇感头痛，更不用说一般数据用户了。“窄”，是指分享渠道狭窄，花费大量成本形成的数据成果因为没有高效、便捷、丰富、简单的分享渠道，限制了数据价值的最大化，这种情况就像大山里的果农，遇到丰年成果却运不出大山，而烂在地里。

二、从建设“数据超市”入手，逐步构建数据“生态圈”

近些年，老马提出的“生态圈”概念炒的非常火，翻阅了一些资料，在此之前也有一些类似的概念被提出过，比如：“相关多元化”、

“打造产业链”、“商业帝国”之类的。既然有一些相似概念被提出，为什么要新提出一个“生态圈”的概念呢？构建所谓的数据“生态圈”能否成为国土资源信息化的解决之道呢？我想，构建数据“生态圈”必须要能解决下面几个问题。

首先是**数据共生**。国土资源数据量很大，数据出处多样，未来的机构改革后的自然资源数据量会更大，更复杂，但无论如何复杂，我们都可以将数据看做是对客观空间的分层描述和解读，既然客观上只存在一个真相，那么在数据的描述中也就应该准确的描述这个真相，无论数据的格式和出处有什么差异，它们都应共生并为真相服务。

其次是**数据共用**。作为国有资产的国土（自然）资源数据，其数据价值应该在全民应用中获得价值最大化，数据描述的真相应该帮助全民更好的理解和利用国土（自然）资源，数据共用促进全民对国土（自然）资源形成共识，全民共识促进国土（自然）资源管理和利用的科学发展。

最后是**数据共赢**。数据的共生共用，促进了数据价值的转换和释放，数据价值的传递需要更加丰富的方式，数据的非独占性决定了数据价值可以重复传递，形成数据共赢的局面，获得数据价值后受益用户会更有热情投入数据的共生共用过程，从而形成所谓的数据“生态圈”。

为了实现上面的共生、共用、共赢目标（以下简称“三共”），我提出构建国土（自然）资源“数据超市”，以“数据超市”为依托，实现“三共”目标，以“数据超市”为基础，不断进化，持续衍生新

的数据产品（服务），围绕“数据超市”不断扩大用户面，将数据利用规则不断加强显化，理顺数据提供者、数据运营者（数据超市）、数据利用者的关系，逐渐形成良性循环的数据“生态圈”。

三、 FME 助推数据“生态圈”进化

构建数据生态圈，会面临很多难题，是一个持续的过程，除了建设者思想到位，还需要有充分的技术准备。我们在“数据中心”核心软件选型过程中做了认真调研，最终，选择了 FME 系列产品作为“数据中心”的重要基础支撑软件，主要基于以下几点考虑：

1、 FME 支持丰富的数据格式，超过 325 种数据格式转换，并且每年在版本升级中都会追加新的数据格式。在“数据超市”建设过程中会考虑到数据共生（共存）的问题，因为现实中数据的格式丰富，全部转换成统一的格式进行存储成本很高，而且很难持续，FME 丰富的格式包容性，赋予“数据超市”更强的数据包容性，无论数据源是怎样的格式，都能很好的处理，并包装成相应的数据产品（服务），较好的实现数据共生（共存）。

2、“数据超市”是一个具有生命的生态系统，数据的变化驱动“数据超市”的进化，所以，“数据超市”需要成熟的基础支撑软件，为持续进化提供必要的能力。FME 拥有完整的产品线，包括桌面版、服务器版以及云版（听说安图已经考虑引进国内），为“数据超市”的持续进化准备就绪，几乎不用考虑随着“数据超市”发展壮大后更换基础支撑平台的问题。

3、“数据超市”面向当前和未来，用户的计算需求丰富多样，

如果采取传统的代码开发模式来满足用户不断增长的计算需求必然带来巨大的综合成本提升。FME 经过很长时间的发展，已经开发了 500 多个成熟的转换器，每个转换器只需通过简单的拖拽连接，就可以完成功能的开发，这个特性无疑为“数据超市”提供了一条取之不尽用之不竭的高速“数据产品”生产线，也使数据“生态圈”低成本快速进化成为可能。

4、 对于更加高级的计算需求，可能会存在模板化开发不能满足的情况（极少数情况），这时，FME 提供了开发接口，用于代码开发，比如，“数据超市”通过代码调用 FME 提供的服务，编写自定义脚本，开发自定义转换器及插件。尽管 FME 提供了丰富的开发接口，但这丝毫没有削弱 FME 模板化开发的魅力，正所谓，模板与接口相辅相成，相得益彰！

5、 以往，数据处理的门槛很高，高到一般用户不得不专门去学习，甚至放弃。“数据超市”拥有一套数据产品化思维，将数据到数据产品的过程对用户全部透明，通过 FME 强大的可视化模板开发功能，将丰富的数据产品按照需求包装起来，最终发布出来的数据产品只是用户最关心的那一部分，而且，一旦发布，全体共用（需要权限），实现数据价值的重复释放。值得庆幸的是，这个包装发布过程非常清晰，容易上手。

6、 “数据超市”可以科学、持续的进化，根本上说还是要“以人为本”，运营和维护“数据超市”的团队非常重要，通过合作，我们体会到无论是 SAFE 公司还是 FME 的国内代理安图公司都是非常具

有活力的实干型企业，他们对产品拥有执着的追求，每年推出一个新版本，优秀的帮助系统、可靠的本地化团队等，FME 不仅是一套工具，它更是一种发展理念，同时传递出一种能力。

以上这些都成为我们考虑选用 FME 的重要因素。

数据“生态圈”的建立是一个复杂的过程，我们也在摸索中前行，除了要抓住“共生、共用、共赢”这个核心，还要选择非常优秀的，具有前瞻性的软件平台作为支撑，共建、共享、合作、共赢这是时代的主旋律，用户对获得数据价值的渴望就是我们数据管理者的前行动力，构建数据“生态圈”，必须选择好“四梁八柱”，步步为营，快速迭代，创新发展。