

自然资源大数据整合建库研究 -以宁夏回族自治区为例

王伟光
DOI:10.32629/gmsm.v3i1.488

[摘要] 本文提出了整合、集成和规范现有的自然资源各类数据库的技术方法,为提高自然资源用途管制能力提供了参考依据。
[关键词] 大数据; 数据整合; 信息化

引言

经过持续的自然资源信息化建设,宁夏初步建成了一批数据资源,提供了多种共享服务,建设了多套业务系统,形成了一批涵盖测绘、土地、地质矿产资源、林草资源、水资源以及地质环境等业务的自然资源信息化成果,为自然资源管理提供了一定的信息化支撑。

新时期自然资源管理面临着由土地资源、矿产资源、地质环境和耕地保护管理全面向“山水林田湖草”系统治理的转变。对照形势要求和发

展需求,自然资源数据管理和应用各自为政的现象还比较突出,数据共享、开放和服务的水平还需要进一步提升。

目前对于自然资源基础平台建设的相关研究较多,不少学者对于自然资源基础平台的建设目标任务、构架体系及相关技术都有比较系统的探索,但是科学分析自然资源数据整合方法的文章并不多见。

1 研究思路

以宁夏自然资源厅内现有各类自然资源数据为基础,全面汇集与梳理全区自然资源多源异构数据,开展全区自然资源数据调查分析评价、数据库整合标准、数据库标准、数据库建设规范等相关标准研究和制定。推进自然资源数据的整合、集成、共享,实现全区自然资源数据的统一管理,做到“统筹规划、整体设计、急用先行、全区共享”,充分发挥全区资源各类数据的应用价值。

2 技术路线

按照国家和宁夏自然资源数据中各类数据对应的现行标准规范,开展数据整合工作,形成具备统一空间参考和质量标准的数据。按照制作的对照映射关系,利用整合处理软件进行自动化处理,实现数据字段结构统一、内容值域统一、数学基础规范、文件命名规范、目录组织规范、数据关系建立等要求。数据标准化整合技术路线如图1所示:

3 整合方法

3.1 数据收集与分析

自然资源数据分析与检查的目的是以土地资源数据、矿产资源数据、基础测绘数据、地质环境灾害数据为分析对象,从数据内容、格式、坐标系、数据质量和差异性等方面进行分析,在摸清各类自然资源数据现状的前提下,有针对性的对各类数据进行数据整合处理。具体分析内容如下:

(1) 数据内容分析

主要是依据自然资源相关标准和规程文件,整理调研收集各类数据,梳理各类数据的内容和体系结构,对收集的待整合对象的属性内容进行确认,包括数据属性字段项是否满足业务需要,字段内容是否有缺失,并进行数据的时效性分析,进行空间数据的一致性检查、拓扑检查,判断数据对象

的基本属性、时空属性以及业务属性的逻辑正确性等。

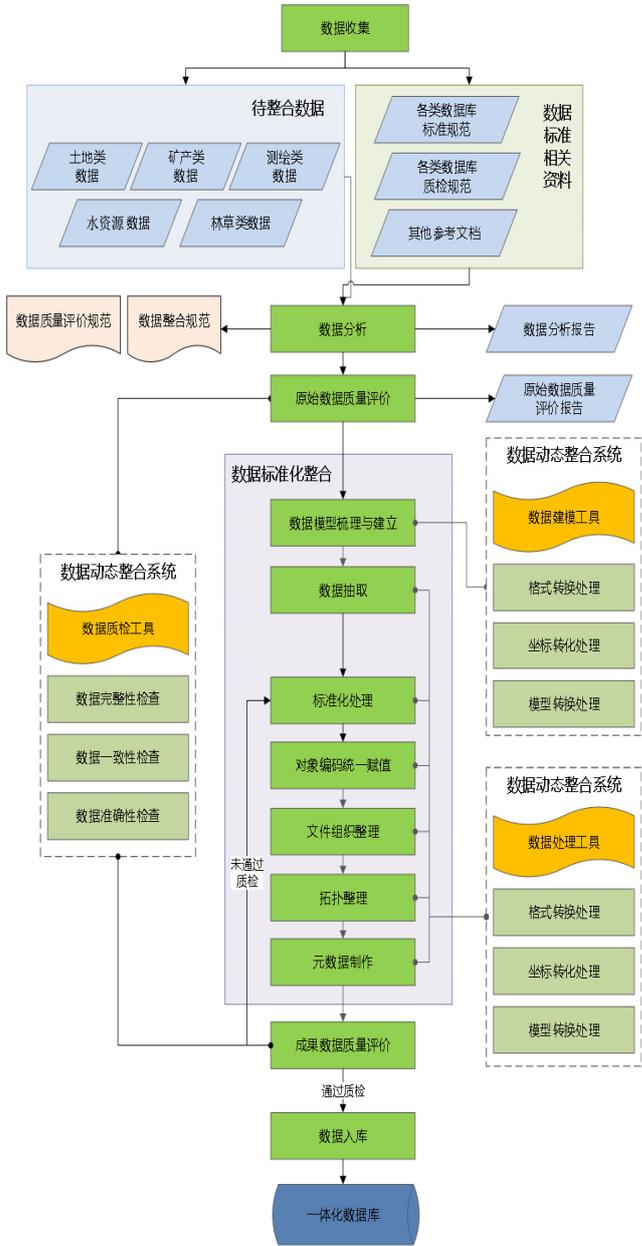


图1 数据标准化整合技术路线图

(2) 数据格式和参考坐标分析

以调研收集的各类数据为分析对象,全面梳理数据格式类型以及其采用的坐标参考系,形成数据格式和参考坐标分析表,为每类数据标准化整合处理提供基础。

(3) 数据质量状况分析

为保证后续工作的顺利进展,首先要对调研收集的各类数据进行初步的分析,从数据的内容完备性、质量准确性、规范一致性、数据时效性等方面判断数据质量情况。

(4) 数据差异性分析

分析同一类数据在不同数据库中的一致性与差异性,是进行数据整合的前提与基础,因此需要以数据标准为依据,对各类数据的分类标准、编码体系、空间布局等进行分类梳理,找出其关联与区别。当同一类型的多套数据满足数据一致性要求时,保留现势性高、数据内容完备、数据质量较优的一套数据;当同一类型的多套数据差异较大时,通过整合处理形成符合一套标准的多套数据^[1]。

3.2 数据关系梳理

对多源异构的数据进行关系梳理,构建现状数据与目标数据模型的对照映射关系,判断是否涵盖目标对象数据模型中涉及的全部属性内容,并建立对应关系,明确数据抽取原则及优先级排序。需要准备的工作内容如下:

(1) 数据模型文件制作

目标数据模型文件制作:依据统一的自然资源数据整合类规范,制作各类自然资源数据模型,包含数据集、数据分层、属性结构、地物分类、属性值域等内容。

现状数据模型文件制作:对收集的各类现状数据,利用数据整合工具,直接导出数据模型文件,包含了数据分层、属性结构、地物分类等内容。

(2) 数据映射关系对照表制作

基于已有的现状数据模型及目标数据模型文件,完成数据图层、属性字段、地物分类、内容值域等信息的对照关系映射,输出每类数据的映射关系文件,为后续自动化抽取现状数据到目标数据模型做准备。

3.3 数据整合处理

(1) 数据格式与参考坐标统一

参考数据分析梳理的各类自然资源数据的数据格式类型以及其采用的坐标参考系,完成不同数据格式如SHP、GDB\MDB、DWG\DXF等的数据转换,统一为CGCS2000坐标系,GDB数据格式。

(2) 数据结构与内容规范化整理

以标准化数据模型设计为依据,基于完成的数据映射关系对照表,完成数据图层、属性字段结构、内容值域等转换处理,使之符合一体化数据库模型的设计。

(3) 对象编码统一赋值

为保证空间数据对象在整个数据资源中的唯一性,按照统一规则对自然资源数据空间对象进行编码赋值,实现空间对象的分类和标识,便于后续基于编码建立空间对象之间的关联关系,为业务关联、业务分析等创造条件。

(4) 文件命名与文件组织目录整理

依照一体化数据库设计和质量控制规定,对文件名称、文件内容等规

范化整理,处理完成之后,按照数据库入库的文件组织目录要求进行整理,便于自动检查与入库^[2]。

3.4 数据质量控制

依照相关的数据库设计和质量控制规定,对文件名称、文件内容、数据内容、元数据结构与内容等进行处理。处理完成后,还应校验处理后的数据是否符合规范,进行整体质量控制,最终保证基础信息整合结果的正确无误。

质量检查主要包括以下方面:

(1) 数据完整性检查

依据数据规定,需要检查数据是否完整、是否存在缺项漏项等情况;对数据现势性进行把控,现势性较低的数据应作为历史数据进行归档。

(2) 数据结构一致性

整合后的数据与数据库模型设计保持一致,通过数据质量检查模块中的方案检测可完成一致性自动检测。

(3) 数据内容一致性

整合后的数据内容与整合前的数据内容保持一致;主要通过数据图层集各类地物数量统计、叠加分析等方式进行一致性检查。

(4) 数据概念一致性

概念无歧义,概念与标准一致,保证非空间表格数据结构与标准保持一致。

(5) 数据格式一致性

表格数据格式与相关要求和标准一致,采用Excel、XML格式或Access格式(.mdb);空间数据格式采用CGCS2000坐标系下的GDB格式。

(6) 数据逻辑一致性检查

检查各层是否有重复要素,各要素之间的关系是否合理,有无矛盾存在等。

数据质量检查过程,可借助数据质量检查模块辅助进行。数据质量检查合格后的数据成果,才能进行数据入库^[3]。

5 结论

新形势下开展自然资源信息化工作,要以全覆盖、全要素、立体调查监测为基础,整合已有的各类自然资源数据,形成相互关联的自然资源数据底板,从而实现自然资源全程、全覆盖动态监管,提高用途管制能力。

通过自然资源基础数据整合,将分散的海量自然资源数据汇聚整合为集中管理的数据资源体系,形成统一集中管理的自然资源数据源,提供快速灵活服务的全方位支撑,促进全区自然资源空间数据在各级自然资源主管部门的互通共用,推进各级政府部门间的数据共享,实现自然资源数据在自然资源业务部门内和政府部门间的充分共享和有序开放。

[参考文献]

[1]孙玉婷.“多规合一”背景下的“规划一张图”拼合研究[J].城市勘测,2018,(02):18-20.

[2]王家耀,武芳,郭建忠,等.时空大数据面临的挑战与机遇[J].测绘科学,2017,42(07):1-7.

[3]刘瀚斌.自然资源监管要用好大数据[N].中国环境报,2018-01-19(3).

作者简介:

王伟光(1985--),男,宁夏石嘴山人,汉族,本科学历,工程师,身份证号:640202198510270510,研究方向:自然资源信息化,国土空间规划。