

地理国情普查中属性数据的质量控制

赵晶晶

新疆维吾尔自治区第一测绘院

DOI:10.12238/gmsm.v4i4.1155

[摘要] 地理国情普查获得了大量的地理空间数据,这些数据将被用于建设地理信息数据库,而地理信息数据库是我国各部门进行建设和开发的重要参考数据,本文介绍了地理国情普查的质量控制的内容要点,介绍了属性数据质量的定义和误差类型,最后介绍了属性数据质量控制方法,旨在为提高地理信息数据的质量提供一些理论性的参考。

[关键词] 地理国情普查; 属性数据; 质量评价

中图分类号: P208 **文献标识码:** A

Quality Control of the Attribute Data in the Geographical National Conditions Census

Jingjing Zhao

The First Surveying and Mapping Institute of Xinjiang Uygur Autonomous Region

[Abstract] Geographic national conditions census has obtained a large number of geospatial data to be used to build geographic information database, which is an important reference data for the construction and development of various departments. This paper introduces the main points of quality control of the geographical census and the definition and error types of attribute data quality, aiming to provide some theoretical reference for improving the quality of geographic information data.

[Key words] geographic national conditions census; attribute data; quality evaluation

引言

地理国情监测的属性数据是信息化社会国家各部门建设以及开发的重要参考数据,地理国情普查是为了建设完善的地理信息系统而进行的一项全国性的普查性活动,地理国情普查获得的数据质量的水平高低将直接影响到地理信息系统的建立,进而直接影响到我国各部门的建设和发展,因此,对地理国情普查中获得的属性数据进行质量评价和控制是非常有必要的^[1]。

1 地理国情普查的质量控制

对于地理国情普查的数据质量管理来说,数据的质量控制对于地理国情普查的质量来说最为重要,该质量控制活动应该贯穿地理国情普查的全部过程,因此和每个部门的工作都息息相关,质量控制的内容见表1。

表1 地理国情普查的质量控制内容要点表

阶段	内容要点
原始资料	应对收集的基础测量信息和专题信息的完备性、精确性进行核查。
多源遥感影像	特别是影像的纠正、镶嵌、融合、分类等方面应加强内外业的核查力度。
外业调查	应对外业调查方法、外业调查底图等方面进行质量控制,采用电子调绘时应保留其记录。
普查数据库	根据项目自身的技术规范和质量标准,建立相应的数据库质量控制办法,例如建立统一的数据库设计规范等。
统计分析	专家应审查论证统计分析中采用的数学模型,质量管理部门应鉴定各单位自行开发的统计分析软件。
信息系统建设	信息系统的建设应按照国家相关标准,对系统设计进行验证,对开发过程应有相应的质量控制办法。
成果发布前	果管理部门应同质量管理部门在成果发布前,做好质量审核工作。主要是对已发布的数据进行对比分析,当数据间出现非时间因素导致的重大差异时,应进行质量核查并提请专家进行数据质量审查论证。

2 属性数据质量的定义和误差类型

2.1 地理信息系统的属性数据质量定义

地理信息系统是通过地理国情普查所得的数据所建立的,它的数据质量,就是这些数据产品是否能够满足某些需求,例如属性精度、位置精度、现势性和逻辑一致性等。

2.2 地理信息系统的属性数据误差类型

地理信息系统的属性数据误差有三种类型,分别是数据获取误差,数据处理误差,显著误差等,其中最容易被检测出来的误差是显著误差,而最难被检测出来的误差是数据处理时产生的误差,因为数据处理的误差不但和被处理的数据有关系,使用不同的数据算法,也会出现不同的误差。虽然显著误差的内容也非常复杂,但是较前两种误差来说,更容易被检测出来。三种误差的内容详见表2。

表2 属性数据调查表

误差类型	内容
显著误差	现势性, 大部分地理信息系统都有现实数据可以作为参考, 但是因为采用的数据的现势性都不好, 现势性和误差的概率成正比
	地图比例尺, 只有采用大比例尺的地图, 内容才会更加全面详细, 由此采集的数据的精确度也会越高
数据处理误差	数据处理误差是指数据在处理过程中产生的误差, 一般数据处理软件会产生一些误差, 另外对数据的分类处理过程中也会产生一些误差。
数据获取误差	仪器误差是指采用的仪器的信号强弱或者仪器受到天气、湿度、温度影响而导致的误差
	采样误差, 一般地理国情监测采用的是人工获取数据的手段, 而人工的读数不可避免的回产生一些误差, 另外天气、环境也会对人工读数产生一些影响而产生误差

3 属性数据质量控制

3.1 属性数据的质量控制目标

表3 地理信息系统的数据质量控制目标

内容	概述
精度	精度是属性值与真值接近的程度, 包含精度和准确度和精确度三个方面, 如地表分类的程度, 地名的详细程度等。精度包括有位精度, 属性精度和时间精度。数据现势性的优劣将直接影响到属性数据的时间精度, 从而影响到属性数据的质量。属性元数据中包含着由纯数据剖析概括得到的信息, 所以属性元数据的质量也可以影响到属性数据的质量。
数据完备性	指数据库中的全部数据值都应该是准确无误的, 如果存在着不正确的数据值, 则该数据库称为已丧失数据完备性。数据完备性是指数据的精确性和可靠性, 它是防止数据库中存在不符合语义规定的数据和防止因错误信息的输入输出造成无效操作或错误信息。数据完整性分为四类: 实体完整性、域完整性、参照完整性、用户自定义完整性。
逻辑一致性	指数据在数据结构、数据格式和属性编码的正确性方面, 尤其是拓扑关系上的一致性。属性逻辑一致性则是侧重于属性值与真实值之间的精确性, 除了上述方面的正确性还应该包含地物的属性、数量、名称等注记的正确性等。在包含时间维度的数据库中, 数据现势性的优劣也将直接影响到属性数据的逻辑一致性。

地理信息系统的属性数据质量标准一般包含两个层面, 例如数据的质量要求, 还有数据的视觉性能标准等。另外, 属性数据的质量包含元数据的质量。元数据实际上也是一种视觉性能, 而数据质

量的要求还包含数据的完整性、精度还有逻辑一致性, 因此这两个方面都是地理信息系统的质量控制目标^[2]。地理信息系统的数据质量控制目标详见表3。

3.2 属性数据质量控制过程

质量是任何一个产品的命脉, 质量应该是一个产品生产全过程的质量的优劣程度。为了让地理信息系统的属性数据质量更优, 要对地理信息系统的进行数据生产前期和数据生产后期的全过程控制。

3.2.1 数据质量数据的前期质量控制
在地理国情数据录入地理信息系统之前, 如果能够对收集来的原始数据进行一些筛查, 就可以发现一些明显的错误, 如果对这些明显错误进行纠正或者直接删除这些数据, 或者在建立数据库时采用一些方法来减少误差的发生, 就可以明显提高前期数据的质量^[3]。

3.2.2 数据质量数据的后期质量控制

表4 属性数据的质量控制过程表

过程	内容
基础属性数据资料的质量控制	基础属性数据原始资料的正确合理的处理与利用, 一方面可以减少误差的产生, 另一方面利用已有的数据资料可以大大提高工作效率, 但利用已有的数据资料时, 要先查看该数据的误差限度是否超过本次项目质量要求的限度, 如超过则一律不予使用。由此可见, 属性数据初始资料的优劣将会影响到属性数据获取的质量好坏, 所以一定要对属性数据初始资料进行质量控制。
属性数据的获取方法	应根据产品的应用范围、客户的需要、经济方面的考量及对产品精度要求的高低, 合理有效地选择最适合的数据获取方法, 尽量达到既经济实惠又具有高质量水平的产品。
属性数据获取前的预先准备	想要达到预期的产品质量, 就要先做好准备, 如了解和掌握与产品相关的技术要求、质量标准、技术规范等。
数据获取的质量控制	指在属性数据获取时, 利用一些方法, 实时监测和改正错误。利用进程表和跟踪记录表是控制数据在获得过程中其质量的主要办法。进程表对数据的获取过程有着很严格的要求, 从而使得全部工作人员可以遵循同样的步骤进行作业; 跟踪记录表是在数据获取过程中利用编好的程序, 通过属性数据间具有的相连关系来控制可能产生的质量问题。

所谓地理信息系统的数据处理的后期, 就是数据已经录入地理信息系统了。然后对这些数据进行数据处理环

节。在这一过程中, 设计一定的检验模板, 或者设计一定的检验程序, 就可以对这些数据进行一定的精度分析, 通过精度分析, 可以让后期的数据质量信息更加可靠精确。

3.3 属性数据质量控制实施

属性数据的质量控制实施过程详见表4。

3.4 属性数据质量控制检查与方法

一般地理信息系统软件是通过 ARCGIS 软件把所有录入的地理信息数据联系起来的, 因此属性数据的质量控制检查也是通过以 ARCGIS 软件依据开发的软件而进行的。例如线状地物的连续性可以通过库查图的方式进行, 面状地物的闭合性也可以通过这一方法完成, 数据录入完成后, 还要通过抽样检查, 来确保各属性代码的准确性, 还要检查各个图形元素和属性要素的正确性。

4 结论

随着科学技术的提高, 地理信息系统的运用范围将更加广泛, 因此对地理信息数据进行质量控制就显得尤为重要, 利用各种数学方法结合现有的评价方法, 将会让评价结果更为全面和正确。

[参考文献]

[1] 赵桂华. 基础性地理国情监测内业成果质量控制方法[J]. 地理空间信息, 2021, 19(03): 124-127+8.

[2] 陈海鹏, 张莉, 郭婧, 等. 地表覆盖分类数据成果质量检查技术与方法[J]. 测绘与空间地理信息, 2018, 41(7): 72-74.

[3] 姚炳全, 赵海涛, 毛文娟, 等. 地理国情监测质量抽查常见问题及关键点分析[J]. 北京测绘, 2021, 35(2): 193-197.

作者简介:

赵晶晶(1979--), 女, 汉族, 新疆乌鲁木齐人, 大学本科, 高级工程师, 在新疆维吾尔自治区第一测绘院工作, 研究方向: 地理国情监测、基础测绘等。