

建设地理信息大数据的思考

■ 翟永 刘津

“地理信息大数据的采集，一方面要坚持传统的数据采集的优点，适当扩大数据采集接收面（如地市级测绘单位），另一方面要积极引入VGI（Volunteered Geographic Information，志愿者地理信息）数据。”

大数据作为一种新型战略资源，提供了一个在虚拟信息世界中了解和掌握客观现实世界的前所未有的机会。地理信息行业作为信息产业的重要分支，应抓住大数据技术和应用所带来的新机遇，面向最广大的公众在地理信息大数据采集、分析与服务方面进行机制创新，推动地理信息行业的快速发展。

一、以数据为核心的地理信息行业

从20世纪50年代GIS（地理信息系统）诞生开始，地理信息行业就以数据为核心逐步发展壮大，其主要业务如采集、检查、处理、建库和分发服务无一不是紧紧围绕数据进行的。数据是地理信息行业的灵魂。

目前，测绘地理信息系统内已经积累了大量的地理信息数据，如全国重力、三角、水准及GPS大地测量成果，各省（自治区、直辖市）覆盖重点区域的1:500—1:10000大比例尺地形图数据，超过500万张的航空影像数据，覆盖陆地国土超过7000万km²（含重复覆盖面积）的多种分辨率卫星影像数据，覆盖全部国土的10~30m分辨率卫星影像，对重点地区基本完成必要覆盖的优于5m分辨率的卫星影像数据，以及分辨率优于2.1m（截至2012年）的资源三号卫星影像数据等。

以上地理信息数据成果（含历史数据）在线存储量超过560TB，离线存储量超过1.3PB，并且在线数据成果以每年不低于300TB的增量扩展。这些内容丰富、类型繁多的地理信息数据是地理信息大数据的基础，可以更好地促进地理信息工作服务大局，服务社会与民生。

客观来说，没有地理信息数据，地理信息服务无从谈起；没有海量、规模化、多类型的地理信息大数据，地理信息智能服务将举步维艰。在即将来临的“智慧中国”时代，拥有海量数据级别、高价值含量的地理信息大数据是推进测绘地理信息服务走向智能

化、智慧化转型升级的必经之路。大数据的发展是大势所趋，是开启智慧时代的必然要求。同时，测绘地理信息是国家重要战略信息资源，建设地理信息大数据具有深远的意义和强烈的现实需要。

二、地理信息大数据发展中的制约因素

目前，地理信息系统已积累了大数据量级别和多种类的地理信息数据，但应客观看到地理信息大数据发展中的制约因素。

1 传统数据采集机制的制约

地理信息数据的采集是地理信息行业的基础环节，是大数据的孕育环节。传统的地理信息数据采集机制是项目驱动下的计划模式，即中央级测绘地理信息单位申请国家立项后，将任务分包到各个省级测绘地理信息单位，分头组织数据采集和质检，再集中汇交进行数据建库、分析与服务。这种采集模式在测绘信息化工作中发挥了重要作用，具有数据标准统一和管理高效等特点。同时，应该看到，这种模式存在数据获取面不广，未能纳入市县级测绘单位和数量众多的公众采集的地理信息；另一方面，数据的类型固化，更新周期长，现势性不高，以1:50000基础地理数据库为例，全国范围的更新周期一般需要3~5年；第三方面，传统数据采集机制灵活度不高，反应时间长，难以满足应急测绘等实时或准实时服务要求。

2 传统数据分析架构的制约

地理信息行业传统上以数据密集型、IO密集型处理为主，随着大数据技术的进步和新时期用户的需求变化，计算密集型处理逐渐占有一席之地。在相关计算机装备配置方面，大型高性能计算服务器主要配置在国家级数据中心的涉密局域网内，其他地理信息单位难以使用；同时，传统的地理信息算法受制于早期计算机技术的制约，以串行计算为主，未能及时改进为并行算法，难以满足大数据环境下分布式并行处理的架构要求，不得不在昂贵的SMP（对称多处理）服

服务器上运行，本单位或有业务关联的其他单位的数目众多的计算机设备无法有效利用。

3 传统数据服务模式的制约

大数据应用的根本目的是将挖掘、分析得出的数据价值，快速全面向用户服务，服务对象是政府、企事业单位、国防部门以及公众的最大集合。否则，大数据就失去了全部或部分存在价值。传统的地理信息服务主要面向涉密部门，在国民经济和国防建设以及应急救援中发挥了重要作用，例如2013年，国家级中心向用户提供数据超过5TB，效果显著。但是，应该看到，在面向非涉密的企事业单位、社会公众服务方面，存在不足：一是数据内容较少，仅限于可公开的低分辨率、低精度的地理信息数据；二是现势性不高，“旧数据”多，最新数据时效超过1年以上，其他数据多超过3年，这类用户（数目更多）对此是不满意的；三是功能不强，面向涉密单位的数据提供以人工为主、网络手段为辅，面向非涉密单位的数据提供集中在“天地图”网站，由于装备能力不足，面向海量用户服务存在等待时间较长、服务波动等不足。

三、地理信息大数据建设的主要改进措施

通过以上分析，地理信息大数据的采集、分析、服务需要不断挖掘新思路，突破原有测绘专业部门进行数据采集、质检、处理、分析和建库的闭合环路，突破主要面向专业用户服务的制约。

1 地理信息大数据采集

目前，大多数地理信息系统是基于自上而下的方式建立的，只有官方提供者可以采集、处理、发布和维护数据资源，这种机制导致了资源维护技术比较复杂，限制了用户的参与，造成贡献资源的不足。地理信息大数据的采集，一方面要坚持传统的数据采集的优点，适当扩大数据采集接收面（如地市级测绘单位），另一方面积极引入VGI（Volunteered Geographic Information，志愿者地理信息）数据。

VGI数据主要来自于公众的自发行为，每一个人都可以成为地理信息数据的采集者，人人都是数据的生产者。长期居住的居民会对周边的地理环境（地名、交通路网）有着较为详尽的了解，其中的部分信息是难以通过自动化的手段快速获取的。例如，专业部门通过航空摄影获得一个区域清晰的像片后，需在后期投入大量的人力、物力和财力才可获取其详细的地名注记，而发动公众，以“人人都是传感器”的思路，打破地理信息专业人员和公众之间的界限，有效整合人们掌握的自己周边的各种信息，降低成本，提高效率，可实现地理信息大数据快速分享和传播。

在地理信息行业内，将传统模式采集的地理信息数据与VGI数据相融合，会创造出具有活力的满足移动互联网时代用户需求的大数据成果。

2 地理信息大数据分析

对于传统的串行处理的大数据处理，基于涉密广域网，面向省级测绘单位开放国家级中心的高性能计算服务器，充分利用宝贵的计算资源；对于可以分布式并行计算的大数据处理，基于涉密广域网，充分利用省级测绘单位的中低端服务器资源，统筹协调，构建基于云计算的大数据分析环境。

对于VGI数据，由于志愿者是自发贡献数据，会存在数据分布不均匀、连续性不一致甚至错误等问题，给VGI数据的处理带来挑战。VGI数据必须经过处理和质量检查以保证数据的形式有效和内容合法合规。在此基础上，基于Hadoop等云计算架构，利用MapReduce技术开发分布式并行算法，面向涉密广域网调用计算资源，进行地理信息大数据分析。

3 地理信息大数据服务

公众是地理信息大数据的最终使用者，是地理信息大数据服务的主要对象。在国家保密法规许可的前提下，提高可公开大数据的分辨率和精度。同时，基于VGI数据，动态更新传统地理信息数据库，增强现势性，经济发达地区和灾害多发区数据时效应在1年之内；丰富内容，数据层和属性信息不应少于Google Map。另外，提升“天地图”网站计算机和网络装备水平，采用性价比较高的Hadoop云计算架构和非关系型数据库，在全国布设多个云计算分中心，全面提升网站服务质量。

地理信息大数据分发服务，要快速将大数据及其分析成果与相关应用发布到最广大的用户群之中，到公众中去，为公众服务。

四、结束语

将大数据与VGI理念相结合，立足传统地理信息数据成果，构建地理信息大数据，优化分析和服机制，发挥地理信息大数据的核心价值，即地理信息大数据来自公众，地理信息大数据服务公众。

（作者单位：国家基础地理信息中心）

深入阅读

Goodchild M F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4): 211-221.

李德仁, 钱新林. 2010. 浅论自发地理信息的数据管理. *武汉大学学报: 信息科学版*, 35(4): 379-383.

周顺平, 徐枫. 2014. 大数据环境下地理信息行业发展的几点思考. *地理信息世界*, 21(1):45-50.