

卷首语

科学大数据驱动地学学科发展

2013年7月17日,习近平总书记指出:“浩瀚的数据海洋就如同工业社会的石油资源,蕴含着巨大生产力和商机。谁掌握了大数据技术,谁就掌握了发展的资源和主动权。”大数据已成为信息主权的一种表现形式,将是继边防、海防、空防之后大国博弈的另一个空间。

1 大数据发展历程

第二次工业革命的爆发,导致以文字为载体的数据量约每10年翻一番;从工业化时代进入信息化时代,数据量每3年翻一番;现在,随着计算机技术和网络技术的快速发展,半结构化、非结构化数据的大量涌现,数据的产生已不受时间和空间的限制。

2008年9月,《Nature》杂志出版“大数据”专刊。大数据的发展不断得到科技界和国际组织的推动与重视。在政府层面,大数据得到高度重视。如,美国发布了“联邦大数据研发战略计划”,投建4个“大数据区域创新中心”;欧盟推出“欧洲云计划”,确保科学界、产业界和公共服务部门均从大数据革命中获益;英国开展大数据技术在政府、高校和公共领域的拓展与应用等。中国提出“实施国家大数据战略,推进数据资源开放共享”,大数据正式成为国家战略。

中国拥有的数据在国际上举足轻重,截至2012年,中国的数据占全球的13%,预计到2020年,中国将产生全球21%的数据。中国与大数据相关的论文数量列全球第2位,仅次于美国。

2 科学大数据

科学大数据具有数据密集型范式的特点,它具有数据的不可重复性、数据的高度不确定性、数据的高维特性、数据分析的高度计算复杂性等内部特征。利用大量数据的相关性可取代因果关系和理论与模型,基于数据间的相关性能够获得新知识、新发现。比如,早在1609年,第谷·布拉赫的助手约翰尼斯·开普勒从布拉赫对天体运动的系数观察记录中发现了行星运动定律,并发表了伟大的著作《新天文学》;再比如,欧洲大型强子对撞机(LHC)帮助物理学家检验关于不同粒子物理和高能物理理论的猜想,并且确定了希格斯波色子的存在。现在,越



郭华东,江苏徐州人,空间地球信息科学家,中国科学院遥感与数字地球研究所研究员、学术委员会主任。中国科学院院士、俄罗斯科学院外籍院士、发展中国家科学院院士。现任国际数字地球学会主席、联合国教科文组织国际自然与文化遗产空间技术中心主任、《国际数字地球学报》主编、《地球大数据》主编,曾任ICSU国际科技数据委员会主席等职。主要从事遥感科学、雷达对地观测、数字地球研究。

来越多科学上的发现证明,大科学装置产生海量的数据已经成为我们认识世界的手段之一,利用大数据驱动科学发现。

科学大数据正在成为一种新的科研方法论,是科学发现的新引擎,近年来中国提出并重视科学大数据的研究。国务院《促进大数据发展行动纲要》中,对“科学大数据”进行了专题论述:发展科学大数据,积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享,构建科学大数据国家重大基础设施,实现对国家重要科技数据的权威汇集、长期保存、集成管理和全面共享。面向经济社会发展需求,发展科学大数据应用服务中心,支持解决经济社会发展 and 国家安全重大问题。

3 地球大数据

地球大数据是一种典型的科学大数据,是具有空间属性的地球科学领域大数据,它一方面具有海量、多源、多时相、异构、多尺度、非平稳等大数据的一般性质,同时具有很强的时空关联和物理关联。这些特征对地学学科的发展可以起到重大的推动作用,在环境、资源、灾害等领域有重要作用和经济社会价值。

地球大数据为地球科学的深入研究

带来了重要的发展机遇,可推动空间地球信息科学,进一步推动地球科学的发展。地球大数据的研究方向主要有4个方面:空间对地观测大数据,涉及海洋卫星、气象卫星、资源卫星、环境减灾卫星、卫星数据接收等;地球大数据处理方法,涉及云计算、智能处理、数据同化、数值模拟等;地球科学学科大数据,涉及海洋科学、大气科学、地理科学、地质科学、地球物理、地球化学大数据等;地球大数据与地球科学,涉及地球系统科学、地学发现等。

地球大数据具有重要作用。比如,利用地球大数据开展“一带一路”研究。中国科学家于2016年发起的数字丝路(DBAR)国际计划,就是要建立一个地球大数据共享平台,提供大数据汇集、大数据服务、大数据分析和大数据系统,形成“一带一路”地球观测数据集。这个为期10年的科学计划最终将不仅贡献“一带一路”,还可贡献联合国全球可持续发展目标,为粮食安全、生态环境保护及风险评估、气候变化和灾害应对以及文化-自然遗产保护与发展等提供科学的决策。

中国科学院正式设立了“地球大数据科学工程”A类先导专项,它的目标是建成国际地球大数据科学中心,构建全球领先的地球大数据基础设施,形成国际一流的地球大数据学科驱动平台,构建服务政府高层的决策支持平台,它将在地球学科发展、政府决策、全球可持续发展等方面产生重大意义。

大数据是知识经济时代的战略高地,大数据是国家新型战略资源,大数据正在改变人类生活及对世界的深层理解。作为少量依赖因果关系,而主要依靠相关性发现新知识的新模式,大数据已成为继经验、理论和计算模式之后的数据密集型科学范式的典型代表,带来科研方法论的创新。科学大数据正在成为科学发现的新引擎,驱动学科创新跨越,驱动地球学科创新发展。

(中国科学院遥感与数字地球研究所,
北京 100094)