



论地学科研信息化环境*

文 / 诸云强¹ 孙九林¹ 冯 敏¹ 宋 佳¹ 苗 茹² 潘 鹏¹ 张金区³ 杨雅萍¹

1 资源与环境信息系统国家重点实验室

中国科学院地理科学与资源研究所 北京 100101

2 河南大学环境与规划学院 开封 475000

3 华南师范大学计算机学院 广州 510631



中国科学院

【摘要】 地学科研信息化环境(e-Geoscience)是在传统地球信息科学基础上,利用新一代信息化基础设施和信息技术,发展起来的支撑地学创新研究及其政府决策服务的科学研究信息化环境。文章提出了e-Geoscience的核心要素及其概念模型,研究了e-Geoscience的典型特征,分析了在Web 2.0、物联网、移动通讯、语义网、云计算等新技术的支撑下,e-Geoscience的技术体系:地球信息资源多源化采集、存储技术、分布式/高效能地学计算技术、地学信息资源共享、精确发现与智能推荐技术、地学数据挖掘与可视化技术以及协同研究技术。最后讨论了地学信息资源持续共享、质量保障、智能发现以及e-Geoscience应用中存在的问题与挑战,并给出积极的应对策略。

【关键词】 地球科学,科研信息化,科研范式,资源共享,协同创新

DOI 10.3969/j.issn.1000-3045.2013.04.011

1 引言

地球科学(简称“地学”)是以地球系统(包括大气圈、水圈、岩石圈、生物圈和日地空间)的过程、变化及其相互作用为研究对象的基础学科,并逐渐朝向以地球系统整体

研究的地球系统科学发展^[1-3]。21世纪地球科学研究具有明显的大科学特征并赋予了更多的社会功能^[4]:

一是不同时空尺度基本地球过程及其相互作用是复杂的,其时间尺度从几秒钟地

* 收稿日期:2013年4月20日

基金项目:中科院地理科学与资源所“一三五”战略科技计划项目(2012ZD010);资源与环境信息系统国家重点实验室自主部署创新研究计划资助项目(O88RA900KA);中科院科研信息化“科技领域云”项目:全球变化生态学“科技领域云”的建设与应用;国家科技基础条件平台:地球系统科学数据共享平台

震到几十亿年的地球环境演化,空间尺度从矿物微区到全球环境变化;二是地球过程的研究依赖于海量的科学数据,是数据密集型的科学,更加重视应用现代观测、探测、实验和信息技术对基本科学数据的系统采集、积累与分析;三是地球系统的整体行为涉及地球各圈层的相互作用,其自然系统中的物理、化学、生物过程和人文因素影响交织在一起,导致地球系统观、整体观的建立;四是空间技术、地球内部探测技术、实验技术以及信息技术的广泛应用,使前沿研究与高新技术发展融为一体,使一系列针对地球科学难题的国际研究计划应运而生;五是自然与人文、社会科学的不断融合,推进“人地关系地域系统”研究,即围绕“人地关系地域系统”探讨人类活动与环境变化的关系,寻求区域人口、资源、环境与发展相协调的途径,为区域发展和国家生态文明建设提供理论支持和决策依据^[5,6]。

中科院地学部“中国地球科学发展战略”研究组^[7]认为:从作为一个整体的、超级学科的发展趋势来看,地球科学研究将朝着“宏观更宏、微观更微、宏观微观相结合”的方向发展,继续保持全球化的研究工作特点。全球化的研究方式不仅指各类国际研究计划的执行,更重要的表现为研究区域性问题和全国性问题离不开对全球性信息和全球其他地区、其他国家研究进展的及时把握。越是前沿性和原创性的研究,这种依赖性越强,因为地球及其各圈层是一个整体。地球科学将更加注重同现代科学技术各有关学科之间的交叉。广泛采用高分辨率的观测系统(地面、水面的直到空间对地观测)、高灵敏度和高准确度的分析测试系统(包括微粒、微量、纳米级和超微量)、不同条件下的实验模拟系统、建立在动力学及高性能计算基础上的数值模拟以及数字化的地球信息系统。

2012 年 国 际 会 议 “Planet under pressure 2012-New Knowledge Towards Solutions”指出^[8]:现有的国际全球变化研究必须进行革命性的变革。应对全球变化风险解决方案的寻求,不但要靠自然对地球系统所发生现象的认识和规律的探

究,更需要社会科学共同参与,在全球尺度可持续发展的框架下,由各个利益相关者对所面临问题以协同设计、共同形成应对方案的方式开展研究和应用活动。

综上,现代地学研究正朝向“地球系统科学”的综合集成研究发展,更加强调圈层间的相互作用,学科间的交叉集成,其研究对象是复杂的非线性巨系统,具有时空尺度大等特点,要求在全球变化的背景下开展区域集成的研究,需要不同学科、不同区域科学家之间的协同研究,甚至科学家与政府、企业之间的合作。现代地学研究是典型的数据密集型研究,综合集成及其观测、理解、模拟和预测已经成为其基本的研究方法^[5],对获取科学数据的新型观测技术、数据转化、处理、压缩技术和大量复杂模型计算技术以及科学数据、资料共享的信息系统至关重要。

鉴于此,作者认为:在信息资源日益丰富和信息技术高速发展的今天,研究方法的创新至关重要,是我国地学研究赶超国际先进水平,实现从地学大国走向地学强国的良好机遇。因此,研究构建满足现代地学研究需要的科研信息化环境(简称地学科研信息化环境,e-Geoscience),提供跨区域、跨学科、跨团队地学协同研究及其对科学数据、模型工具、高性能计算能力等的支撑,对于促进地学科技创新和服务社会经济可持续发展非常重要和迫切。

2 地学科研信息化环境的提出

为了支撑地球科学,特别是地球系统科学的研究,20 世纪 90 年代产生了地球信息科学(Geo-informatics 或 Geo-information Science)。地球信息科学是在信息科学和地球系统科学基础上,由卫星遥感、全球定位系统、地理信息系统、计算机制图与电子地图、数字通讯网络、多媒体技术与虚拟技术等高度集成的科学技术体系,是 20 世纪 70 年代发展起来的信息科学和 80 年代兴起的地球系统科学交叉形成的一门新兴科学^[9,10]。

地球信息科学强调地球信息的三元特征(属

性、空间、时间),在对地球系统及各组成部分信息流形成机理研究的基础上,开展地球信息采集、传递、存储、处理、显示与应用的研究,认为地球信息在流动中不断发生转变和增值。地球信息科学的技术体系主要由对地观测技术、地理信息技术、互联网技术综合集成^[10]。

随着 Web 2.0、互联网、智能移动终端、云计算等技术的发展以及科学数据的急剧性增长,海量科学数据对科学研究的影响、新的科研范式等引起了国内外学者的广泛关注。2011年2月 *Science* 刊登了“数据处理 (*Dealing with data*)”专辑,围绕日益增长的研究数据洪流进行研讨,认为^[11]:大部分的学科领域正在面临数据洪流的挑战,如果能更好地组织并访问到数据,对于科学研究来说将是巨大的机会。在地理信息领域,Michael F. G.^[12]在 Web2.0、集体智慧等背景下,提出了自发地理信息(VGI)的概念,认为人人都是地理信息的传感器、使用者和贡献者。VGI将是传统地理信息采集方法非常有效的补充。李德仁等^[13]认为新地理信息时代已经到来,地理信息服务对象扩大到大众用户、用户同时是空间数据和信息的提供者、传感器网络将数据从死变活、提供按需求服务等。

Jim Gray 更是认为新一代的科研范式“数据密集型范式”(The fourth paradigm of scientific research: Data-intensive science)已经出现^[14]。Jim Gray 将科学研究范式总结为四种:第一范式产生于几千年前,是以观察和实验为依据的研究,可称为经验范式;第二范式产生于几百年前,是以建模和归纳为基础的理论学科和分析范式,可称为理论范式;第三范式产生于几十年前,是以模拟复杂现象为基础的计算科学范式,可称为模拟范式;第四范式今天正在出现,是以数据为

基础,联合理论、实验和模拟一体的数据密集计算的范式,数据被一起捕获或者由模拟器生成,由软件处理,信息和知识存储在计算机中,科学家使用数据管理和统计学方法分析数据库和文档,可称为数据密集型范式。

由此可见,相比与地球信息科学强调发展地球信息的采集、获取(如全球定位系统 GPS、对地观测系统 EOS),处理、分析与可视化(如地理信息系统 GIS、虚拟地理环境等)单一技术与应用,数据密集型时代下的科研范式,更加强调网络环境下,科研人员之间的协同交流、科技资源(数据、模型、计算资源等)的开放共享、智能关联与协同应用。因此,数据密集型科研范式下,需要进一步在地球信息科学的基础上,研究和发展地学科研信息化环境,构建科研人员既是地学信息资源(数据、模型、文献、知识等)的使用者,更是信息资源贡献者的氛围,利用这些信息资源、工具软件、信息化基础设施等开展协同的研究,从而提升地学研究的效率和水平。

3 地学科研信息化环境的显著特征

科研信息化环境(e-Science)作为科学研究的下一代基础设施,早在1999年由 John Taylor 提出后,便引起了科技界的广泛关注,发达国家和我国都启动了一系列相关的科技计划^[15-17]。

John Taylor^[18]认为 e-Science 是促使全球性的、跨学科的、大规模合作研究和资源共享成为可能的基础设施。Christine L. Borgman^[19]认为 e-Science 的目标是构建一种新型的科学研究模式,该模式具有信息密集、数据密集、分布式、协作和多领域特征。江绵恒^[20]认为 e-Science 的实质就是科学研究的信息化,是信息时代中科学研究环境和科学研究活动的典型体现。桂文庄^[21]认为



中国科学院

e-Science是信息技术渗透到科研活动中,是在各种科研活动中系统地应用最先进的信息技术成果,革命性地发展新的科研手段、科研模式、科研环境,从而实现科学技术新革命的途径。钱德沛^[22]认为计算、数据、协同是科研信息化的三个基本要素。协同是现代科学研究的基本特征,今天的重大科学研究必须在开放的空间,通过交换信息,交流经验,密切协作来解决问题。郭毅可^[23]提出了科学即服务(Science as a Service)的概念,认为科学研究信息化旨在实现信息的共享,而共享信息导致了科学研究从封闭走向开放,从个体走向协同,从少数走向大众。这种开放式、协同性、大众化的科学成为网络的新服务,即科学即服务。

e-Geoscience是数据密集型科研范式下,为了满足现代地学研究的需要而构建的科研信息化环境。e-Geoscience是在传统地球信息科学的基础上,进一步在高速互联网络、高性能计算、海量存储设备以及Web 2.0、云计算、移动通讯等新一代信息化基础设施和信息技术支撑下,支撑现代地学创新研究和服务社会经济可持续发展决策的科研信息化环境。

e-Geoscience的概念模型如图1,包含了4个主客体要素:科学问题、科技资源、科研人员和管理服务人员。即在e-Geoscience中科研人员(相互间可进行交流协作)利用科技资源(相互间可交换互操作)解决科学问题,而科技资源依靠管理人员得到有序管理与集成。科研人员在解决科学问题的同时也将自己的科技资源贡献到科研信息化环境中,从而形成“科研人员既是科技资源的使用者,又是科技资源的贡献者”的可持续发展氛围。

e-Geoscience具有以下显著特征:

(1)e-Geoscience是地球信息科学的继承和发展。e-Geoscience是在地球信息科学基础上发展起来的,地球信息科学中的地球信息机理、图像信

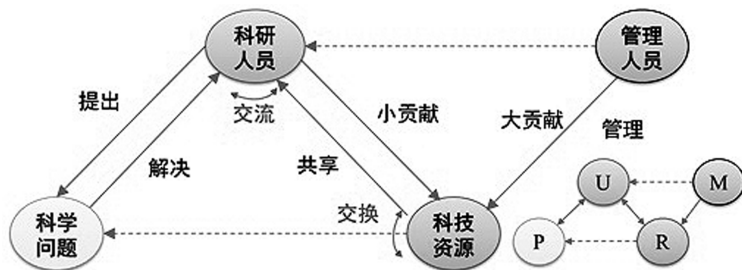


图1 e-Geoscience概念模型

息机理、地学信息图谱、地学本体等理论方法以及所涉及的信息采集、处理与可视化技术体系,如地对地观测遥感技术、全球定位技术、地理信息技术等都应是e-Geoscience的基础。同时,e-Geoscience更加强调以科研人员需求为核心,实现地球信息科学技术方法的集成、信息的创造与共享、协同的研究。

(2)新一代信息化基础设施和技术是e-Geoscience的基础。e-Geoscience必须构建在新一代信息化基础设施和技术之上,包括:高速的网络环境、海量的数据存储设施、高性能的超级计算环境、分布式计算、云计算等等。由于所有的活动和应用都需要在网络上开展,因此高速的网络环境又是e-Geoscience基础的基础。海量数据存储设施和高性能超级计算环境则解决了地学海量数据存储和高精度地学模拟计算的迫切需要。

(3)地学科技资源共享是e-Geoscience的核心。地学科技资源包括科技设施(实物)资源和科技信息资源。科技设施(实物)资源包括:通讯网络、台站/基地、样地/样品、仪器设备、存储、高性能计算等基础设施和硬件设备、实物资源。科技信息资源包括:科技文献、科学数据、模型算法以及软件工具等。各类科技资源之间的关系如图2所示。长期以来,由于体制、观念和技术的原

因,地学科技资源得不到充分的共享,地学资源重复建设现象明显,严重阻碍了地学研究的发展。因此,全球范围内各类地学资源的共享是开展全球性跨区域、跨学科合作研究的核心。

(4)科研人员主动参与及开放性是e-Geosci-

4 地学科研信息化环境技术体系

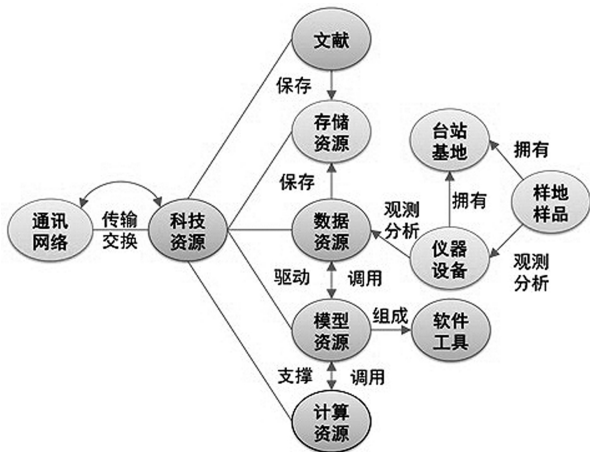


图2 地学科技资源实体及其关系

ence 的生命力。e-Geoscience 的服务对象是科研人员,因此科研人员的参与是 e-Geoscience 发展的生命力。一方面,基于 e-Geoscience,科研人员要能够获取到开展科学研究所需要的各类资源,包括:数据资源、模拟模型、文献知识以及强大的计算、存储资源;更重要的是要按照 Web 2.0 的理念,构建一个开放的 e-Geoscience,使得科研人员能够将自己的数据、模型等资源不断地整合到 e-Geoscience 中,营造科研人员既是地学资源的使用者,又是资源的贡献者的氛围和环境,只有这样才能促进 e-Geoscience 的持续发展。

(5) 科研创新驱动是 e-Geoscience 的终极目标。e-Geoscience 基本功能是为全球性、跨区域、跨学科的协作研究提供信息资源和信息技术的支撑,包含:分布式地学资源的协同调度和组合应用以及科研人员之间的协同交流等等。更重要的是 e-Geoscience 不断积累和丰富各类地学信息资源与知识以及先进、智能的数据挖掘分析工具、模拟模型等,基于这些信息资源和技术工具可以更好地发现地学数据背后隐含的地学规律、知识甚至是新的科学问题,从而驱动和引领地学研究的发展。

e-Geoscience 技术体系如图 3 所示,分为广义和狭义两个层面。广义的 e-Geoscience 技术体系涵盖整个地球信息的采集、获取、存储、处理、模拟分析与可视化等,由地球信息科学传统技术体系和现代计算机信息技术体系组成,并在不同地学研究环节中进行有机的融合。传统地球信息科学技术体系主要包括:全球定位系统(GPS)、遥感技术(RS)、地理信息系统(GIS)、虚拟地理环境、地理信息共享

技术等。现代计算机信息技术体系主要包括:Web 2.0、物联网、移动通讯、语义网、云计算等。狭义的 e-Geoscience 技术体系主要指在数据密集型时代下,地球信息资源多源化采集、海量地球信息资源存储、分布式计算、地学信息资源共享、智能关联与协同应用、地学资源可视化与虚拟现实以及跨区域/跨学科地学科研人员协同交流技术等。e-Geoscience 技术体系的一个重要特点就是动态扩展性:一方面,地学科学问题的复杂性,要求 e-Geoscience 技术体系根据应用需求不断地扩展和增强;另一方面,信息技术的高速发展,也在不断补充完善 e-Geoscience 技术体系。

传统的地球信息科学技术体系在本文中不再赘述,下面重点对数据密集型科研范式下 e-Geoscience 的技术体系进行阐述。

4.1 泛在地球信息资源多源化采集技术

e-Geoscience 的信息资源不仅包含传统的野外考察、调查、监测数据,室内处理、测试分析与模拟数据,也包含文献资料、互联网中的网页、微博客、社交网络等新媒体中的信息资源,体现出泛在特征。因此, e-Geoscience 中的地球信息资源采集、获取在传统方法(如野外测量、自动监测、考察调



中国科学院

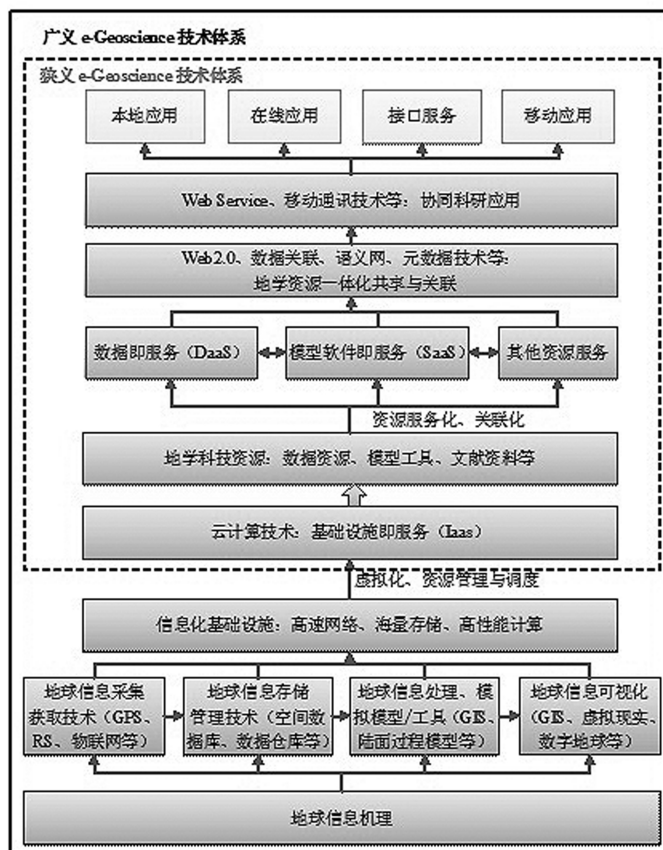


图3 e-Geoscience技术体系

查、统计、测试分析、对地观测、深部探测等)的基础上,在采集方式上进一步发展了物联网感知、互联网数据挖掘、移动智能终端采集等方式;在内容上进一步丰富为数据、照片、音视频等的集成采集;在传输方式上,通过 Zigbee 等无线通信技术、无线电波、GPRS 通讯网络以及卫星、互联网等的综合应用,实现偏远野外观测站点到数据中心的即时、高保障的传输等。

4.2 海量地球信息资源存储技术

在依靠存储设备容量扩展以及数据仓库技术等的基础上,e-Geoscience 应该面向大量用户并发访问的需求,更加关注数据的高吞吐率、高传输率和安全可靠性,提出大规模数据的分布式存储技术方案。如 Google 文件系统(GFS)、Apache Hadoop 的 HDFS,就是可扩展的分布式文件系统,用于大型的、分布式的对大量数据进行访问的应用,

它运行于廉价的普通硬件上,但可提供容错功能,可以给大量的用户提供总体性能较高的服务,采用冗余存储的方式来保证数据的可靠性。e-Geoscience 中应该重点解决海量地理空间数据的分布式存储与高效检索等问题。与此同时,在云计算时代,拥有海量数据存储设备的数据中心,可以通过虚拟化技术,为普通科研用户提供数据存储即服务(IaaS)。

4.3 分布式、高效能地学计算技术

分布式计算技术是研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分,然后分配给许多计算机进行处理,最后把这些计算结果综合起来得到最终的结果。如 Apache Hadoop 提供的 MapReduce 分布式计算框架,通过 Map 将一个任务分解成为多个任务,利用 Reduce 将分解后多任务处理的结果汇总起来,得出最后的分析结果。图4显示的是作者基于 Apache Hadoop 的分布式数据存储(HDFS)和计算框架(MapReduce)开发的分布式地形山影计算系统^[24]。测试

结果表明,分布式计算能够显著提高地形数据计算的性能,特别是在计算节点超过3个以上时,计算效率会大幅度提升。同时,为了提高地学计算效率,需要进一步将地学计算模式与并行计算(多线程、多进程、OpenMI 等方法)、GPU 等技术进行紧密结合。

4.4 地学信息资源共享技术

通过国家科技基础条件平台、国家电子政务工程、国家地理信息公共服务平台等各类重大科学计划和工程的推动,我国在地学数据资源共享方面取得了长足的进步。e-Geoscience 更加关注数据共享到模型共享,甚至是知识共享的发展,提供一站式的地学资源共享。如图5所示,作者研究构建了以元数据和文献服务、数据服务(OGC 的 WMS、WFS 服务)、模型服务(OGC 的 WPS 服

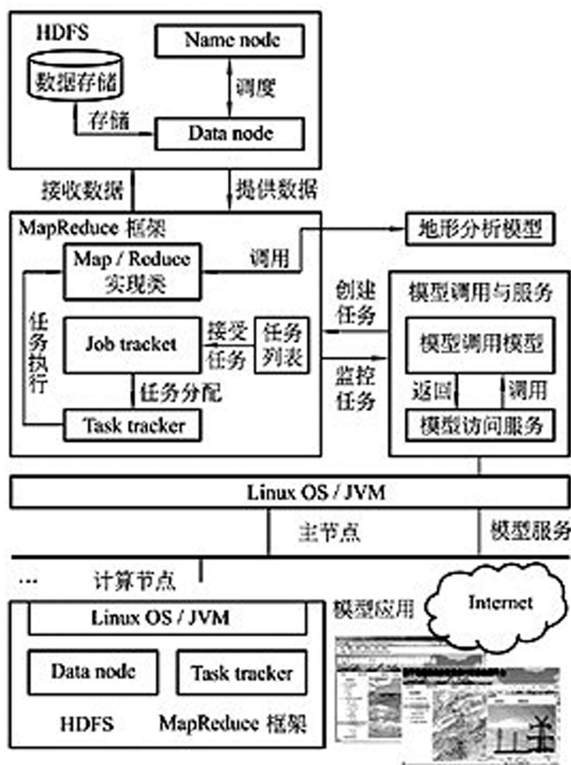


图4 分布式地形数据存储和山影计算系统

务)等为核心的e-Geoscience原型系统^[15],实现了文献资料、地学数据、模型工具的集成共享;更加关注利用Web2.0技术,充分发挥科研人员的积极性,激励科研人员自发在e-Geoscience中发布地学资源,营造“人人都是科技资源的使用者,也是贡献者”的地学科技资源共享氛围,构建更加高效、更加开放、更加智能、更加社会化、更加简便的资源共享软件^[25]。

4.5 地学资源精确发现与智能推荐技术

如何精确发现用户需要的地学信息资源,并把相关的资源智能推荐给用户,是地学信息资源在爆炸式增长、大数据时代下,e-Geoscience必须考虑和解决的问题。地学信息资源精确

发现就需要利用本体技术,对各类地学信息资源进行语义标识和推理;而智能推荐则需要利用关联数据(Linked data)技术对各类地学资源进行明确的语义表达和相互链接,同时通过用户行为挖掘分析,将相关的数据自动推介给用户。

4.6 地学数据挖掘与可视化技术

“数据海量、信息缺乏”是当前大数据时代地学研究面临的尴尬问题。因此,需要利用数据挖掘技术从大量、不完全、有噪声、模糊的数据中挖掘出隐含的时空分异规律、格局演变趋势等。如利用任意形状空间丛聚区域挖掘方法揭示疾病和犯罪空间丛聚区域^[26],利用轨迹数据挖掘方法提取城市交通时空变化模式^[27]等。同时,为了快速捕获数据背景的知识,需要采用直观、形象、多维的可视化技术来

表征地学数据和信息。如利用GIS可视化技术发现义乌对全国各地流动人口吸引形成的V字型空间格局^[28],中国畜禽养殖存在明显的东西疏密分界线等^[29]。

4.7 地学科研人员协同研究技术

协同研究是指在虚拟e-Geoscience社区中,每个科研人员都是虚拟成员,虚拟成



图5 e-Geoscience原型系统

员之间能利用 e-Geoscience 中的资源协作开展合作研究,包括:利用即时通讯技术进行实时的文字、语音、视频等的交流;进行各类地学科技资源的共享,更重要的是资源间的协同利用,如通过 e-Geoscience 在线模型的执行,驱动后台高性能计算资源,调用匹配的数据资源等;面向复杂应用,利用服务链技术,进行模型服务的建模集成等等。

5 地学科研信息化环境面临的问题与对策

e-Geoscience 在支撑促进现代地学创新研究和服务社会经济可持续发展决策的同时,也面临着一系列的问题和挑战。我们必须正视这些问题和挑战,寻求积极的政策机制、标准规范、技术方法的对策,抓住机遇,切实推动和发展 e-Geoscience,助力我国从地学大国向地学强国的发展。

5.1 地学信息资源持续共享

地学信息资源共享是 e-Geoscience 的核心。尽管我国科技资源共享取得显著的进步,然而如何确保持续良性循环的地学信息资源共享还存在很大的挑战。一方面要持续加强国家层面的政策引导,约束和规范各类科技计划产生的信息资源共享;另一方面要尊重知识产权,采用 DOI(Digital object identifier)和 DCI(Data citation index)对信息资源进行标识引用和评价,要求用户在成果中必须明确标注信息来源,引用提供者要求的参考文献,保障信息资源提供者的知情权、决定权和被引用权,形成信息资源拥有者自愿参与地学信息资源共享的激励机制,发挥群体智慧才能促进地学信息资源的持续共享。

5.2 地学信息资源质量保障

传统信息资源共享模式中,信息资源通常按照规定的标准进行汇交、集成和发布,由共享管理中心负责信息资源的质量。在 e-Geoscience 中,实行“人人都是科技资源的使用者和贡献者”的机制,大量科研人员的参与有利于地学信息资源的持续汇集和更新,然而如何保障科研人员共享的地学信息资源质量是关键的问题。一方面要倡导

“谁发布、谁负责”的信息资源质量观;另一方面要加强信息资源质量标准的宣传和贯彻及在线自动检查工具的应用;尤为重要是在 Web2.0 环境下,要建立网络信息资源质量同行评审、管理中心鉴定和用户评论制度。对于通过评审和鉴定的信息资源,或者用户好评的信息资源,进行高信誉度的标签,以使用户的使用。

5.3 地学信息资源智能发现

不同类型、海量的地学信息资源在 e-Geoscience 中不断积累、增长时,采用传统的关键词匹配或分词检索技术,由于信息资源语义异构等问题,会造成难以精确、快速发现用户需要的信息资源问题。为此,必须加强地学信息资源语义互操作研究,建立地学时空本体库、信息资源内容本体库、信息形态结构本体库等,利用关联数据(Linked data)和语义推理技术,进行各类信息资源的自动链接和智能搜索,提升地学信息资源发现的查全率和查准率。

5.4 地学科研信息化环境应用

数据密集型时代下,大科学特征的现代地学研究,尤其国家生态文明建设的新要求迫切需要一个“资源丰富、功能强大、开放共享、按需服务、协同应用、稳定运行”的地学科研信息化环境的支撑。当前,应该在国家层面制定或启动“地学科研信息化环境发展规划或科技计划”,在国家科技基础条件平台或国家科技计划数据资源汇交的同时,推动地学科研信息化环境的发展。当然,我们必须清楚地认识到:e-Geoscience 是一个复杂的、渐进的系统工程,不可能一蹴而就,需要地学、数学、计算机等不同领域科学家和工程师的长期努力,需要深入分析科学研究对象,建立科学问题概念模型及其对应的信息流模型,并在应用中不断完善发展。

参考文献

- 1 孙枢. 对我国全球变化与地球系统科学研究的若干思考. 地球科学进展, 2005, 20(1): 6-10.
- 2 汪品先. 我国的地球系统科学研究向何处去. 地球科学进展,

- 2003, 18(6): 837-850.
- 3 周秀骥. 对地球系统科学的几点认识. 地球科学进展, 2004, 19(4): 513-515.
 - 4 国家自然科学基金委员会地球科学部. 地球科学的战略地位及其发展趋势和特点. <http://www.nsfc.gov.cn/nsfc/cen/00/kxb/dqback/yxzzly/daoyan1.htm>. 2000.
 - 5 孙九林, 林海. 地球系统研究与科学数据. 北京: 科学出版社, 2009.
 - 6 樊杰. 主体功能区战略与优化国土空间开发格局. 中国科学院院刊, 2013, 28(2): 193-205.
 - 7 中国科学院地学部“中国地球科学发展战略”研究组. 地球科学: 世纪之交的回顾与展望. 中国科学院院刊, 2001, (2): 101-105.
 - 8 国际地圈生物圈计划中国全国委员会秘书处, 国际全球环境变化人文因素计划中国国家委员会秘书处, 中国科协联合国环境咨商委员会. 从伦敦“重负下的星球”大会看国际全球变化研究的动向—兼论中国全球变化研究的未来发展方向. <http://wenku.baidu.com/view/9ee111d13186bceb19e8bb29.html>, 2012.
 - 9 陈述彭主编. 地球信息科学. 北京: 高等教育出版社, 2007.
 - 10 廖克等著. 地球信息科学导论. 北京: 科学出版社, 2007.
 - 11 Science Staff. Challenges and Opportunities. Science Special Issue: Dealing with Data, 2011, (331): 692-693.
 - 12 Goodchild M F. Citizens as sensors: the world of volunteered geography. GeoJournal, 2007, 69(4): 211-221.
 - 13 李德仁, 邵振峰. 论新地理信息时代. 中国科学(F辑: 信息科学), 2009, 39(6): 579-587.
 - 14 Tony Hey, Stewart Tansley and Kristin Tolle. The Fourth Paradigm Data-Intensive Scientific Discovery. Microsoft Research, 2009.
 - 15 诸云强, 孙九林, 宋佳等. 地学 e-Science 研究与实践—以东北亚联合科学考察与合作研究平台构建为例. 地球科学进展, 2011, 26(1): 66-74.
 - 16 孙坦主编. 数字化科研—e-Science 研究. 北京: 电子工业出版社, 2009.
 - 17 中国科学院、中华人民共和国教育部、国家自然科学基金委员会. 中国科研信息化蓝皮书 2011. 北京: 科学出版社, 2011.
 - 18 Hey T, Trefethen A E. The UK e-science core programme and the grid. Future Generation Computer Systems, 2002, 18(8): 1017-1031.
 - 19 Borgman C L. Data, disciplines, and scholarly publishing. Learned Publishing, 2008, 21(1): 29-38.
 - 20 江绵恒. 科学研究的信息化: e-Science. 科研信息化技术与应用, 2008, (1): 8-13.
 - 21 桂文庄. 什么是 e-Science? 科研信息化技术与应用, 2008, (1): 1-7.
 - 22 钱德沛. 科研信息化的若干思考. 中国科学院科研信息化实践与探索. 北京: 科学出版社, 2011.
 - 23 郭毅可. 科研信息化 2.0: 从构造为科学研究的服务到服务化的科学. 中国科研信息化蓝皮书 2011. 北京: 科学出版社, 2011.
 - 24 冯敏, 尹芳, 诸云强等. 基于 MapReduce 的分布式地形数据计算研究. 华中科技大学学报(自然科学版), 2011, 39(Sup.I): 24-27.
 - 25 诸云强, 宋佳, 冯敏等. 地球系统科学数据共享软件研究与发展. 中国科技资源导刊, 2012, 44(6): 11-16.
 - 26 Pei T, Wan Y, Jiang Y et al. Detecting arbitrarily shaped clusters using ant colony optimization. International Journal of Geographical Information Science, 2011, 25(10): 1575-1595.
 - 27 Chen J, Shaw S L, Yu H B et al. Exploratory data analysis of activity diary data: a space-time GIS approach. Journal of Transport Geography, 2011, 19(3): 394-404.
 - 28 Hongsheng Li, Yingjie Wang, Jiafu Han et al. Origin Distribution Visualization of Floating Population and Determinants Analysis: A Case study of Yiwu City. Proceedings Environmental Sciences, 2011, (7): 116-121.
 - 29 Fu Q, Zhu Y Q, Kong Y F et al. Spatial analysis and districting of the livestock and poultry breeding in China. Journal of Geographical Sciences, 2012, 22(6): 1079-1100.



中国科学院

Study on e-Science for Geosciences

Zhu Yunqiang¹ Sun Jiulin¹ Feng Min¹ Song Jia¹ Miao Ru² Pan Peng¹ Zhang Jinqu³ Yang Yaping¹

(1 State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Beijing 100101, China

2 The College of Environment and Planning of Henan University, Kaifeng 475004, China

3 South China Normal University, Guangzhou 510631, China)

Abstract Scientific research paradigm has been developed from experience, theory, and computing paradigm to the Fourth Paradigm: data-intensive science. Modern geoscience research is a typical data-intensive research with obvious characteristics of big science and more social functions. So it is urgent to promote geo-information resources sharing and collaboration of cross-disciplinary, cross-regional, and cross-team in global range. To meet this need, it is the necessary choice to set up electronic supporting environment for geosciences (e-Geoscience). e-Geoscience is an electronic environment for supporting geosciences innovative research and government decision-making services that is to inherit and carry forward traditional geo-informatics science. In this paper, firstly we put forward the key elements and conceptual model of e-Geoscience, then study the typical characteristics of e-Geoscience and analyze new technologies such as Web 2.0, internet of things, mobile communication, semantic net, and cloud computing. Furthermore, the technical architecture of e-Geoscience is studied in detail which includes: multi-collection of Earth information resources, storage technology, distributed/ high performance geoscience computing technology, resources sharing of geoscience information, accurate findings and intelligent recommendation technology, geoscience data mining, visualization technology, and collaborative research technology. Finally, we discuss the problems and challenges in the sustaining sharing of geoscience information, quality assurance, intelligent findings, and application of e-Geoscience. The positive coping strategies are presented in the last.

Keywords geoscience, e-Science, science paradigm, resource sharing, collaborative innovation

诸云强 中科院地理科学与资源所副研究员、博士,资源与环境信息系统国家重点实验室副主任,地球系统科学信息共享中心副主任。主要从事地学数据共享关键技术、地学科研信息化环境、资源环境信息系统研究。主持研发了分布式科学数据共享软件和地学科研信息化环境平台软件等,研究成果直接应用到国家地球系统科学数据共享平台中,并成功推广到生态环境、地质环境、环保档案、人口健康等领域。E-mail: zhuyq@igsrr.ac.cn

孙九林 男,中国工程院院士,中科院地理科学与资源所研究员、博士生导师。兼任中国科学院信息化专家咨询委员会主任、国家环境信息化第一届顾问专家委员会主任、国家农村农业信息化示范省建设专家委员会、国家国土资源信息化咨询委员会名誉主任等。长期从事信息科学与资源及农业科学交叉领域的研究和实践,取得多项开拓性的理论与应用成果,为创立“国土资源及农业信息”工程科学和国家建设做出了突出的贡献。在国家层面积极倡导、推动科学数据共享和科研信息化环境。出版专著11部,发表论文80余篇。获省部级以上18种奖励。E-mail: sunjl@igsrr.ac.cn