

全球变化科学领域的信息基础设施研究

白玉琪

(清华大学地球系统科学研究中心 地球系统数值模拟教育部重点实验室, 北京 100084)

摘要: 全球变化科学领域面临着体量巨大、增量迅速、种类繁多的观测数据和模拟数据的挑战, 亟需新的研究思路和技术手段, 实现数据的有效管理和分析, 提高科学发现的效率。本文建议围绕科学数据、科研过程、学科交叉和国际国内合作4条主线开展全球变化科学领域的信息基础设施研究。该项研究需要理解该领域不同学科方向的科学数据, 挖掘它们在概念、空间、时间、量纲、格式、编码等方面的联系, 建立统一的数据概念模型; 需要分析该领域不同学科方向的研究全过程, 分析不同过程、阶段之间的共性, 建立统一的研究过程模型; 需要在“数据-方法-模型-系统-标准”5个层面开展学科之间的共享和集成, 建立统一的数据逻辑模型和物理模型、方法库、模型库、集成系统和标准体系; 需要发展国际国内的合作伙伴, 践行开放共享的共建方案。本文最后概要介绍了这一理念的GlobalChangeOne研究项目的总体设计, 以及在遥感科学、生态科学、大气科学、地球系统模式、计算地球系统科学等学科方向的具体研究进展。该项研究有助于促进全球变化科学领域多学科间的交叉研究, 有助于提高该领域的科学研究工作的效率, 有助于建成全球变化与可持续发展研究全球共享的一体化基础平台。

关键词: 全球变化; 可持续发展; 信息基础设施; 学科交叉; 科学研究效率

DOI: 10.3724/SP.J.1047.2013.00809

1 引言

全球变化是指由自然和人文因素引起的地球系统功能的全球尺度的变化, 包括大气与海洋环流、水循环、生物地球化学循环以及资源、土地利用、城市化和经济发展等的变化^[1]。全球气候近百年来呈现出以变暖为主要特征的显著变化, 对生态和环境产生了严重影响, 也对人类社会的生存和发展提出了严峻的挑战。

全球变化领域的科学活动, 主要集中在观测地球系统发展过程、理解地球系统整体行为、模拟未来地球系统状态、评估人类活动影响、制定经济发展对策、开展国际谈判和合作等方面。通过这些科学和管理活动, 人们积累了体量巨大、增量迅速、种类繁多的观测数据和模拟数据。预计这些数据的总量在2013年达到25 Petabytes, 到2015年将达到45 Petabytes, 到2030年更将达到350 Petabytes^[2]。例如, 2005年美国航空航天局的戈达德飞行中心的地球科学数据和信息服务中心的遥感数据存量达到了1.75 Petabytes, 日增数据量为1.7 Terabytes^[3];

2013年国际气候模式诊断和对比计划委员会(PCMDI)的第五阶段工作中, 23个模式研发团队提交了超过50个气候系统或者地球系统的数值模拟结果, 总数据量大约是1.5 Petabytes (<http://cmip-pcmdi.llnl.gov/cmip5/availability.html>)。从全球范围来看, 全球变化科学领域数据“将在物理存储、管理和共享等很多方面带来巨大的挑战, 另外对于非气候科学家而言, 能够方便准确地实现数据查询和获取, 已非易事^[2]”。

从数据层面来看, 全球变化科学领域需要提出可持续可扩展的解决方案, 不仅要满足科学数据的有效管理、研究分析、共享集成的需求, 还要满足不同学科领域和不同类型用户的科学发现、应用研究、学科交叉的需求。

信息基础设施(Cyber-infrastructure)的概念是2003年由美国NSF在一系列研究报告中提出的, 包括《通过Cyber-infrastructure促进科学和工程的革命(Revolutionizing Science and Engineering through Cyber-infrastructure)》^[4-11]。它是利用高速的计算机网络和软件系统, 集成多个物理分散的组织或机

收稿日期: 2013-11-21; 修回日期: 2013-12-10。

基金项目: 卫星测绘技术与应用国家测绘地理信息局重点实验室经费资助项目(KLMSTA-201307)。

作者简介: 白玉琪(1976-), 男, 安徽人, 博士, 研究方向为地球信息科学。E-mail: yuqibai@tsinghua.edu.cn

构,形成逻辑集中的科研环境,提供先进高效的计算和处理服务,支持科研数据的获取、存储、管理、集成、挖掘、可视化、分析等研究任务。信息基础设施能够将硬件设备、软件资源、数据资源和研究人员四者有效地联系起来,实现资源的互联互通,提供协同的工作环境,从而促进科学家之间跨地域、跨学科的对话与合作,促进科研活动中的信息共享与交流,促进学科交叉,提高科学发现的效率和决策支持的能力。从全球范围来看,欧盟、英国和日本先后支持了不同学科领域的信息基础设施研究。例如,英国从2000年开始大规模地资助地球科学、生物信息学、粒子物理、社会模拟等领域的e-Science研究^[12]。美国国家自然科学基金委则先后支持了GEONGrid(地球科学网络网格)、CIG(地球动力学信息基础设施)、LTER(http://www.nsf.gov/pubs/2000/nsf0030/nsf0030html/longterm_eco_research.htm)(长期生态学研究网络)、LEAD(<http://d2i.indiana.edu/leadii-home>)(大气科学研究网络)、EarthCube(<http://earthcube.ning.com>)(全美新一代地学信息基础设施)等国家级规模的大型科研项目。

从以上分析可以看出,信息基础设施是一个新的研究思路,顺应了科学研究走向数据密集型的发展趋势^[13-14]。另外,需要对特定领域开展信息基础设施研究,以解决领域特定的科学问题和应用问题。

本文针对全球变化科学领域的研究工作,提出该领域的信息基础设施研究主线,并介绍实践这一设想的GlobalChangeOne研究项目的具体进展。

2 全球变化科学领域的信息基础设施的研究主线

针对全球变化研究的特点,本文提出全球变化科学领域的信息基础设施研究应围绕科学数据、科研过程、学科交叉和国际国内合作4条主线展开。

第1条主线,是从“科学数据”入手。理解该领域不同学科方向的科学数据,分析出它们在概念、空间、时间、量纲、格式、编码等方面的关联关系,建立全球变化科学领域的统一数据概念模型。

第2条主线,是着眼于“研究的全过程”。分析该领域不同学科方向的研究全过程,明确各自在数据获取、存储、管理、集成、挖掘、可视化、分析、发表等不同方面存在的功能需求;分析不同过程阶段之间的

共性,建立全球变化科学领域的统一研究过程模型。

第3条主线,是注重“多学科间的数据共享和集成”,这涉及到“数据-方法-模型-系统-标准”5个层次的递进。在提供解决方案,满足不同学科方向的功能需求的同时,可以进一步把这些解决方案在5个层次上进行分解,而后在同一层次上进行横向集成,逐步建立统一的数据逻辑模型和物理模型、方法库、模型库、集成系统和标准体系。

第4条主线,是注重“国际国内合作”。需要明确研究可以依赖的科研资源(包括开放性数据、系统和研究力量),明确和其之间的合作方式,并确定其通讯和集成的技术规范,逐步实现滚动式的发展和完善。

图1以层次递进的方式,总结了上述建议方案。其中,“科学数据”主线部分,用方块表示不同的学科方向,用相同的颜色表示在不同学科方向的观测数据或模拟数据在概念、空间、时间、量纲、格式、编码等方面具有的某种关联性。在“科研过程”主线中,引用了两幅图(<http://mc.libguides.com/content.php?pid=47444&sid=349712>; <http://www.archimuse.com/papers/ukoln98paper/index.html>),用以表明研究过程的一般环节,并示意了6个共性的研究步骤:从数据的获取、存储,到分析、集成、可视化等处理过程,再到研究数据的发表和研究结论的公布。在“学科交叉”主线中,显示了“数据-方法-模型-系统-标准”5个递进层次。在“国际国内合作”主线中,本文选取了国内外一些典型的开放数据源用示例:不同领域的数据共享网(比如地球系统科学数据共享平台(<http://www.geodata.cn>)、中国气象科学数据共享服务网(<http://cdc.cma.gov.cn>)、基础科学数据共享网(<http://www.nsd.cn>)、交通科学数据共享网(<http://www.transdata.cn>)、北京市政务数据资源网(<http://www.bjdata.gov.cn>)、国家科技基础条件平台(<http://nstic.gov.cn>)、政府间对地观测(CEOS)卫星数据主目录CWIC(www.ceos.org/wadc)、全球综合对地观测系统GEOSS(<http://www.earthobservations.org>)^[15]、国际气候模式诊断和对比计划委员会PCMDI(<http://www.pcmdi.llnl.gov>)、美国国家气候资料数据中心NCDC(<http://www.ncdc.noaa.gov>)等。图1最下层是计算存储的硬件资源层,表明全球变化科学研究需要强大的超算设备和存储设备的支持。上述5个方面共同支撑了全球变化领域的科学研究和应用研究。

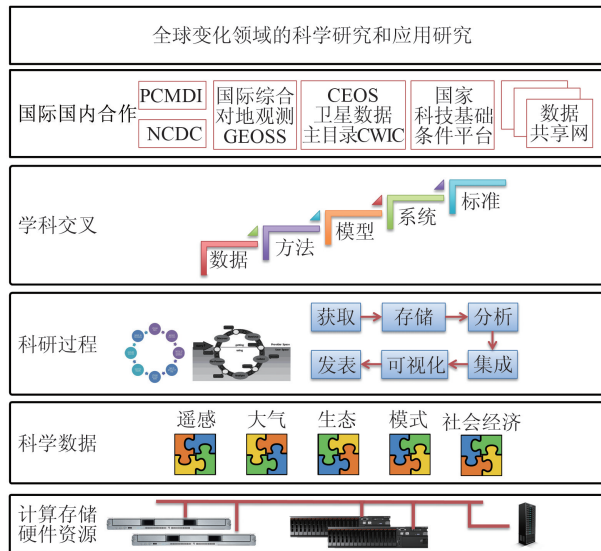


图1 全球变化科学领域的信息基础设施研究主线图

Fig.1 Roadmap for research cyberinfrastructure for global change studies

3 GlobalChangeOne 研究案例介绍

GlobalChangeOne 是清华大学地球系统科学研究中心(以下简称为地学中心)的一项科研计划,旨在遵循上述的研究主线图,依托地学中心在遥感科学、生态科学、大气科学、地球系统模式、全球变化经济学、计算地球系统科学的多学科研究力量,逐步建成示范性的全球变化科学领域的信息基础设施,在满足这些学科方向的数据获取、存储、管理、集成、挖掘、可视化、分析、发表等功能需求的同时,有效地促进学科之间在科学数据、研究方法、分析模型、应用系统、标准规范等5个层面的共享和集成。

3.1 GlobalChangeOne 系统结构

遵循上述的全球变化科学领域的信息基础设施研究主线图,结合地学中心的学科发展现状和规划等实际情况,GlobalChangeOne 具体设计了6个层次的系统结构(图2):最下方是计算和存储硬件资源层,往上依次是数据存储和管理、数据处理、数据分析、数据可视化和系统互操作。其中,数据存储和管理层针对这些学科方向中的主要数据类型开展数据模型和管理方法的研究;数据处理模块从操作系统、支撑软件、处理代码3个层面提供对这些数据的处理功能;数据分析服务则进一步把这些功

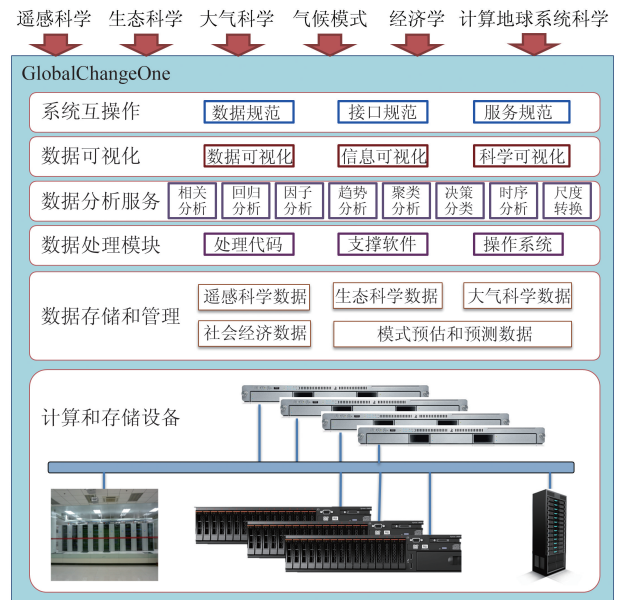


图2 GlobalChangeOne 的系统架构图

Fig.2 System srchitecture of GlobalChangeOne system

能模块封装成相对独立而完善的分析包,以 Web 服务等方式提供出来;数据可视化层需要提出有效的数据可视化、信息可视化和科学可视化的方法,满足不同类型用户对全球变化领域观测、模拟、分析结果等数据的可视化需求,以利于提高决策支持的能力;系统互操作需要研究该领域的规范、接口规范、服务规范、为大范围、多层次的合作研究提供有效的、可持续的支持。

3.2 GlobalChangeOne 研究进展

GlobalChangeOne 项目已经落实 160 核小型计算集群、64 核 Web 服务集群和 100TB 磁盘存储等计算和存储设备。GlobalChangeOne 在地学中心主要学科方向上的研究进展概述如下:

(1) 遥感科学方向

针对遥感卫星观测数据产品(data collection),建立了语义描述本体库。建立了专用的地表覆盖数据产品的数据分发网站(<http://data.ess.tsinghua.edu.cn>)。分析了微软的 Azure 公有云平台的软件架构,正在建立基于云计算平台的遥感影像数据处理和分析系统原型。

(2) 气候模式方向

梳理了 PCMDI 第五阶段的实验设计文档,了解了命名规则、观测变量等约定;完成了对中科院大气物理所和清华地学中心联合研制的 FGOALS-g2 大

气模式输出数据的分析,明确了所包含的实验信息,明确了每种输出数据所含变量的含义,及其对应的时间分辨率和空间分辨率。以PCMDI第五阶段的海冰数据(Sea Ice Thickness)为例,建立了基于微软公有云平台 Azure 的海冰模式数据的可视化和对比分析平台(http://cmip5thu.cloudapp.net/CESS_CMIP5/)。

(3) 大气科学方向

选取了 WMO 的 Global Climate Normals 1961–1990 年数据集,分析了 WMO 气象数据的语法格式和语义定义;设计并提供了一个完整的在线分析系统原型^[16],完成了气象数据记录的处理、气象数据库的建立、气象数据分析方法的实现。

(4) 生态科学方向

围绕着 Eddy Flux 通量观测数据,制定了通量数据的交换标准^[17],制定了基于 ISO 19115、19115-2、19139 的元数据规范建议^[18],并和中美碳联盟(USCCC)团队合作,选取了湛江和高桥两个通量观测站的实例数据,展示了一个完整的通量数据转换、存储、管理、查询、下载、在线分析的解决方案^[19]。

4 讨论与结论

全球变化研究所面临的挑战既体现在对复杂地球运动现象和规律的科学认知,也体现在需要有效地管理和分析海量、多样、动态的科学数据,还体现在“全球气候变化研究自身就具有多学科交叉的特点”(<http://news.sciencenet.cn/htmlnews/2011/6/248148.shtm> (徐冠华院士:中国应大力加强全球变化研究))。中国的全球变化研究在许多方面具有自己的特色,但在一些方面和世界强国的差距很大。因此,开展全球变化研究,既要考虑到质量,也要考虑到效率。这正是本文建议开展全球变化科学领域的信息基础设施研究的目的。

从前期的分析报告^[20]可以看出,领域的针对性越强,功能需求越具体,能够实现的数据管理、处理和分析的功能就越丰富,对研究全过程的支持力度就越强。因此,针对全球变化科学领域,通过对每个学科的数据需求的深入分析,推动多学科间在数据、方法、模型、系统、标准 5 个层面的深入交叉和无缝集成,沿着科学数据、科研过程、学科交叉和国际国内合作 4 条主线并进,逐步建成该领域的信息基础设施的可行性非常高。这是通往建成“全球变化

与可持续发展研究全球共享的一体化基础平台”(<http://www.dost.moe.gov.cn/dostplan/xwxx/20120820132553> (徐冠华主任在“全球变化与可持续发展协同创新中心”高层论坛上的讲话))的重要一步。

参考文献:

- [1] 徐冠华,葛全胜,宫鹏,等.全球变化和人类可持续发展:挑战与对策[J].科学通报,2013,58(21):2100-2106.
- [2] Overpeck J T, Meehl G A, Bony S, *et al.* Climate data challenges in the 21st century[J]. Science, 2011,331(6018): 700-702.
- [3] Leptoukh G. NASA remote sensing data in earth sciences: Processing, archiving, distribution, applications at the GES DISC[C]. Proceedings of the 31st International Symposium on Remote Sensing of Environment, June 20-24, Saint Petersburg, Russian Federation, 2005.
- [4] Atkins D E, Droegemeier K K, Feldman S I, *et al.* Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on cyberinfrastructure[M]. Washington DC: National Science Foundation, 2003.
- [5] Crawford D, Jameson L, Leinen M, *et al.* Cyberinfrastructure vision for 21st century discovery[R]. Washington DC: National Science Foundation, 2007.
- [6] Atkins D E, Thomas D, Hey T, *et al.* Advisory committee for cyberinfrastructure task force on data and visualization: Final report[R]. Washington DC: National Science Foundation, 2011.
- [7] Stewart C, Giraud G, Long L, *et al.* Advisory committee for cyberinfrastructure task force on campus bridging: Final report[R]. Washington DC: National Science Foundation, 2011.
- [8] Oden J T, Ghattas O, King J L, *et al.* Advisory committee for cyberinfrastructure task force on grand challenges: Final report[R]. Washington DC: National Science Foundation, 2011.
- [9] Zacharia T, Kinter J, Pennington R, *et al.* Advisory committee for cyberinfrastructure task force on high performance computing: Final report[R]. Washington DC: National Science Foundation, 2011.
- [10] Keyes D, Taylor V, Hey T, *et al.* Advisory committee for cyberinfrastructure task force on software for science and engineering: Final report[R]. Washington DC: National Science Foundation, 2011.
- [11] Ramirez A, Fox G C, Bower J M, *et al.* Advisory committee for cyberinfrastructure task force on cyberlearning and workforce development: Final report[R]. Washington DC: National Science Foundation, 2011.

- [12] Hey T, Trefethen A E. Cyberinfrastructure for e-Science [J]. Science, 2005, 817(2005):817-821.
- [13] Hey T, Tansley S, Tolle K. The fourth paradigm: Data-intensive scientific discovery[M]. Redmond : Microsoft Research, 2009.
- [14] Wright D J, Wang S. The emergence of spatial Cyberinfrastructure[J]. PNAS, 2011,108(14):5488-5491.
- [15] Bai Y, Di L, Chen A, *et al.* GEOSS component and service registry: Design, implementation and lessons learned [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012,5(6):1678-1686.
- [16] 徐灏.气象气候数据在线分析系统的设计与实现[D].武汉:武汉大学,2013.
- [17] Lv B, Yan A, Bai Y, *et al.* A new generation of data sharing platform for eddy covariance flux data[C]. Proceedings of the Agro-geoinformatics Conference, Fairfax, VA, USA, 2013.
- [18] Yan A, Lv B, Bai Y, *et al.* Metadata standard proposal for eddy covariance flux data[C]. Proceedings of the Agro-geoinformatics Conference, Fairfax,VA,USA, 2013.
- [19] Yan A, Lv B, Bai Y. FluxDataONE: Integrated flux data management and analysis for agricultural and ecological studies[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (Under Review).
- [20] Bai Y, Di L. Review of geospatial data systems' support of global change studies[J]. British Journal of Environment and Climate Change, 2012,2(4):421-436.

Research Cyberinfrastructure for Global Change Studies

BAI Yuqi*

(Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing 100084, China)

Abstract: In the global change area, the volume of and the variety of accumulated observation and simulation data are expanding rapidly, creating challenges for the archiving, analyzing and sharing. To address these issues, this paper suggests four research strengths towards a research cyberinfrastructure in this area: scientific data, research process, interdisciplinary study, and national and international collaboration. Research on these aspects will lead to an integrated conceptual data model, an integrated research process model, a data processing module library, a data analysis service library, and a tree of standards and specifications. A research pilot project of GlobalChangeOne is highlighted to demonstrate the feasibility and the advantages of the proposed research roadmap. This research is very promising in facilitating the interdisciplinary studies and accelerating the scientific insights in the global change area, and in promoting the establishment of an integrated international platform for global change studies and sustainable development research.

Key words: global change; sustainable development; cyberinfrastructure; interdisciplinary study; research efficiency

*Corresponding author: BAI Yuqi, E-mail: yuqibai@tsinghua.edu.cn