

文章编号 1001-8166(2008)02-0201-05

美国长期生态学研究网络(LTER)信息化基础设施现状、挑战与未来发展趋势^{*}

——LTER 信息化基础设施战略规划介绍()

牛 栋^{1,2} 杨 萍² 何洪林³

(1.中国科学院生态环境研究中心,北京 100085;2.中国科学院资源环境科学与技术局,北京 100864;
3.中国科学院地理科学与资源研究所,北京 100101)

摘 要:信息化基础设施战略规划是美国长期生态学研究网络(LTER)战略规划的重要组成部分。在美国国家科学基金会资助下,来自多个领域的专家在系统分析美国长期生态学研究网络信息化基础设施的现状及面临挑战的基础上,提出了为支持 LTER 在未来(几)十年的核心科学计划——社会与环境综合科学计划,在站点、网络办公室、网络层次需要开展的一系列信息化基础设施活动及产出,为未来开展 LTER 信息化基础设施建设提供指导。同时也可为中国生态系统研究网络(CERN)及世界其他长期生态学研究网络的规划与发展提供启示。

关 键 词:长期生态学研究网络;信息化基础设施;规划

中图分类号:Q14;TP3 文献标识码:A

1 提出的背景

信息化基础设施(Cyber Infrastructure, CI)是由美国国家科学基金会(National Scientific Fund, NSF)蓝带委员会(Blue Ribbon Committee)提出的术语(Atkins 2003),它描述的是一个全新的研究环境,该环境支持先进的、基于因特网的数据获取、存储、管理、集成、挖掘、可视化,以及其他计算和信息处理服务。在科学应用上,CI是一种有效的技术解决方案,可以将数据、计算机和人三者结合起来,从而获得先进的科学理论和知识。同时,CI也包括计算机操作、设备维护、软件开发、标准和规范制定,以及提供安全和用户支持等其他关键服务的技术人员和组织机构。

随着生态学的综合性、跨学科性和协同性日益增强,生态问题(如全球变化、生物多样性等)也越来越复杂,对生态学的研究逐渐由单个科研人员、单个站点的研究模式向协同、集成和综合研究的模式

转变,网络层面的综合科学研究成为长期生态系统研究网络的优先发展方向。美国长期生态系统研究网络(The Long Term Ecological Research Network, LTER)在美国国家自然科学基金会资助下,于2004年正式开始了制定规划的工作。目的是建成一个将生态学、社会学等学科有机结合起来的高度综合的研究网络,并提高研究人员在不同时空尺度下对社会状况和生态环境的认识程度和对未来发展的预测水平。LTER 对未来(几)十年的网络层面的综合研究进行了近两年的战略规划研究工作,制定出了科学计划的核心部分:即社会与环境综合科学方案。该方案是网络未来(几)十年发展的指导性纲要文件,该方案明确了网络建设初期的重点是研究多重胁迫条件下人类与自然耦合系统的长期影响,要求在多尺度、多学科上开展集成研究和系统教育,同时要求将分布于不同站点的各类资源、数据和信息进行共享,该计划的实施已经在不同程度上给 LTER 的信息化信息基础设施建设带来了重大的挑战。

* 收稿日期 2007-12-15,修回日期 2008-01-24.

作者简介:牛栋(1974-),男,河南开封人,博士,主要从事长期生态学和生态系统管理研究工作.E-mail: niudong@cashq.ac.cn

面对 LTER 中长期战略规划对其生态站的 CI 建设所带来的挑战,由来自多个科学技术中心、大型 IT 开发项目以及 LTER 的 IT 专业人士,与 LTER 信息化基础设施的规划人员组成专家组,对 LTER 战略规划中需要信息化基础设施(CI)辅助制定的科学计划,重点围绕:多站点和联网试验、数据集成、生态建模、网络架构和人力资源等 4 个关键领域,全面展开 LTER 的信息化基础设施建设的规划工作。

为了深刻地理解信息化基础设施建设所面临的挑战,扩大规划过程中的影响范围,促进各项目之间的整合,参加人员按照四大领域分成 4 个小组,在系统地分析 LTER 信息化基础设施的现状、面临的挑战的基础上,描述了未来开展 LTER 信息化基础设施建设的前景,研究了与之相关的活动、成果产出和产品开发,提出了发展 LTER 的 CI 的 6 项战略思想,广泛咨询美国 LTER 国家顾问委员会(National Advisory Board)和 LTER 网络信息系统顾问委员会(Network Information System Advisory Committee)的意见,完成了 LTER 信息化基础设施战略规划的制定,该项战略规划将成为美国长期生态系统研究网络战略规划的重要部分。本文主要介绍 LTER 的 CI 的现状、面临的挑战及未来 LTER 在站点、网络办公室和网络层面开展 CI 建设的活动及产出。

2 LTER 信息化基础设施的现状

LTER 经过 20 年的发展,信息化基础设施在全球生态系统研究网络中具有一定的优势和特点,主要包括网络 and 站点两个层次。

2.1 LTER 信息化基础设施的现状

在网络层次上,积累了丰富、全面、存档完好并可在线提供的不同生态系统的长期数据;开发了数据目录、水文和气象网络数据产品、LTER 站点基本信息以及研究人员数据库、LTER 站点遥感图像数据产品,制定了生态元数据语言(Ecological Metadata Language, EML)行业标准,并在 LTER 网络 and 站点两个层面推广使用。在全球生态系统研究网络中, LTER 是第一个最大规模地采用了元数据标准的网络,现有的标准化元数据是未来信息化基础设施建设的基础。同时, LTER 网络所制定的有关站点信息管理的标准,已获得整个生态界广泛认可。

LTER 网络开发出了网络信息系统(NIS),推动从各站点的研究活动中获取数据,并开发数据产品;关注现有数据的应用,通过制定和采用生态元数据语言(EML)来保证数据质量和发现数据问题,通过

联邦构架来增加数据资源,通过利用已有数据和基础设施应用程序(系统、软件)来推动数据产品的开发,通过提供资源、指导和培训,支持标准化的站点信息内容开发和管理。该系统将主要服务于 LTER 网络与站点的科研及其合作伙伴,同时也成为 LTER 以外的科研人员、自然资源管理人员、政策制定者以及普通大众一个进入 LTER 网络的“门户”。

2.2 LTER 信息化基础设施的优势和特点

2.2.1 LTER 网络层次

(1) 数据的长期性与数据产品的完整性和系统性:在网络层次上,已经积累不同生态系统的长期数据,开发了数据产品。

(2) 具有自己的行业标准:制定了生态元数据语言 EML 行业标准,并在 LTER 网络和站点两个层面推广使用。第一个最大规模地采用元数据标准的网络,同时,网络也制定了有关站点信息管理的标准。

(3) 开放的数据政策:该政策明确地提出了如何发布、获取和使用 LTER 数据,以及对数据用户和数据提供者的要求。

(4) 完善的管理体制:成立了网络办公室(LTER Network Office, LNO)明确其职责是为生态信息学、计算和通讯设施提供支持和指导。

(5) 生态信息研究与技术团队:积极参与信息学的研究工作,推动生态信息学的发展,为网络的生态学问题提供新的信息学解决方案。

(6) 数据管理网络信息系统:主要为 LTER 的科研及其合作伙伴服务,同时也是 LTER 以外的科研人员、自然资源管理人员、政策制定者以及普通大众一个进入 LTER 网络的“门户”。

(7) 稳定的战略合作伙伴:在网络层面上, LTER 与美国几个国家级生态或计算中心,如美国国家生态分析和综合中心(NCEAS)、圣地亚哥超级计算中心(SDSC)、美国国家超级计算应用中心(NCSA)等建立了战略合作伙伴关系,积极开展生态信息学研究,如 NSF 资助的 KNB 项目、SEEK 项目、GRID 前沿项目等,为 LTER 内部能够长期合作, LTER 信息技术领域始终能够顺应潮流的发展,并掌握最新动向,奠定了坚实的基础。

2.2.2 LTER 站点层次

LTER 的各个站点均拥有十分强大的信息化基础设施能力,但各不相同,归纳起来有以下几个特点。

(1) LTER 各站点信息管理(或信息技术)人员

存在较大的差异,既有兼职也有全职的信息管理人员,其职责也有一定的差异,但这些管理人员的时间大部分用于站点的日常数据管理,包括系统维护、用户支持服务、软件开发和网页设计等。定期的信息管理培训活动是站点参与到网络层次的科学研究活动的重要内容。

(2) LTER 站点的数据形式、传输方式多种多样,包括传感器数据、自动观测系统采集数据及人工记录的野外数据。各站遵循统一的信息管理标准,特别是以生态元数据(EML)形式存在的结构化元数据标准,但在采集和使用过程中,元数据的质量、覆盖范围、元数据的管理系统等仍存在较大差异,此外,不同站点之间的数据传输方式也存在迥然不同的问题。

(3) LTER 各野外站目前已经全部联网,都能通过因特网传输信息,但各站的带宽及服务水平存在差异。美国 LTER 50%以上的站都有足够的因特网带宽和无线基础设施支持传感器网络,但为传感器网络提供支持的人员和技术力量明显不足,造成网络水平的差异。

(4) LTER 各野外站的生态信息学研究已经具备一定的基础。部分站的生态学研究人员和网络办公室的信息人员共同开展生态信息学的研究,同时生态学科研究人员与 IT 人员也逐步加强生态信息学广泛的合作及研究。

(5) 站点已具备了初步的网络协同工作能力。LTER 的部分站点已经具备商用远程视频会议的设施,50%的站能访问 Polycom——宝利通视频会议系统,但能够访问个人台式电脑、众多会议室和野外站点的还是很少。除了使用 wiki, Groove 和某些网络会议软件外, LTER 各野外站的人员一般很少使用其它的网络沟通工具,电子邮件(E-mail)是各站科研、技术人员的主要电子沟通工具。

(6) 各站点的科学家已经开始使用传统的统计和分析软件(如 SPSS, MATLAB),但站点内能够使用遥感分析、可视化和项目管理等软件工具的人仍然很少。

总之, LTER 各站的信息化基础设施水平“参差不齐”。LTER 网络必须扩大 LTER 各站的信息化基础设施和人员力量,才能为整个网络层次的科学研究提供支持。同样, LTER 网络办公室当前的工作重心主要在信息化和通信基础设施以及信息管理方面,但应对大型项目的海量数据及其集成,并为未来(几)十年的科学问题提供支持,人员和装备力量远

远不足。而整个 LTER 网络在推动生态科学的发展方面将面临众多重大挑战,要成功地迎接这些挑战, LTER 必须将足够的资源投入到信息化基础设施的建设之中。

3 LTER 信息化基础设施所面临的挑战

利用社会生态数据的巨大的科学价值来认识世界,并预测生命系统的反馈作用,长期生态研究(LTER)和社会生态学界目前正面临着众多挑战。而这些挑战,同样也是其他从事高性能计算、大尺度和集成科学研究的学者所面临的重要问题(Maltev 2006)。如我们很难将分布在全球各地的科学家组织起来进行合作,面对为综合科学提供集成的高质量数据集等方面, LTER 长期面临巨大的挑战(Stevens 2006)。尽管 LTER 具备某些优势,可以使网络研究人员在迎接上述挑战中占有先机,但整个 LTER 网络仍然面临一些特殊的挑战,主要包括:

(1) 各站的信息化基础设施条件存在极大的差异,如何快速获取数据,科学地管理来自网络科学研究中日益增加的海量数据所带来的挑战。

(2) 对涉及多学科、多变量和多站点的日益增多的数据开展集成,提高数据传输能力,以及对基于台站产生的标准不同的原始数据,如格式、精度、尺度、语义和质量保证/控制等,对这些数据的统一及规范所带来的挑战。

(3) 日益增多的分散在不同地方的研究团队,合作开展跨尺度综合科学研究所带来的挑战,这种合作往往不是自发组织,需要精心规划,同时充分应用各种先进的信息技术,如虚拟合作环境、共享软件工具、远程会议技术等。

(4) 整个 LTER 网络层次的建模和研究活动服务功能的信息化基础设施建设所带来的挑战。应对这个挑战,要求提高认识和预测区域、大陆和全球生态系统变化的能力,要求信息化基础设施能够提供支持的最佳途径、层次以及类型。

(5) 为获取、处理和分析来自于多站联网实验、传感器网络、遥感数据以及建模过程中的网络层次的海量数据,需要提供共享计算和数据服务所带来的挑战,通常不可能为每个新布设的试验单独提供人力以及信息化基础设施。

(6) 各站受过专门训练的人员的需求所带来的挑战。包括对同时掌握其它专业学科知识的信息学专家以及精通 IT 技术的生态学研究生和研究人员的的需求所带来的挑战。

4 未来 LTER 信息化基础设施建设的主要活动及产出

为了适应未来 LTER 的科学计划发展的需要,各站点、网络层次现有的信息化基础设施远远不能满足需要,必须建设 LTER 的信息化基础设施在站点层次、LTER 网络办公室(LTER Network Office)层次、网络层面等所须具备的全新能力,同时也需要在精通 IT 技术的生态学人员中培养和发展用于数据采集、管理、发现、集成、知识表达、分析和建模技术等方面的能力和力量,对此必须进行重大投资。

未来 LTER 信息化基础设施具有以下能力:快速传输海量数据产品的数据服务能力;提供可以集成大量历史数据和实验数据,面向服务构架基础上的计算机虚拟环境;构建虚拟协同工作环境,使其具有服务于生态学知识、并能发现和挖掘数据的计算机工具和算法,同时做到界面友好、方便用户登录和浏览信息、实现信息的可视化以及标注等功能。未来 LTER 信息化基础设施的架构如图 1 所表示,其中包含有在站点、网络办公室(LNO)、网络层面开展的一系列信息化基础设施建设的活动及产出。

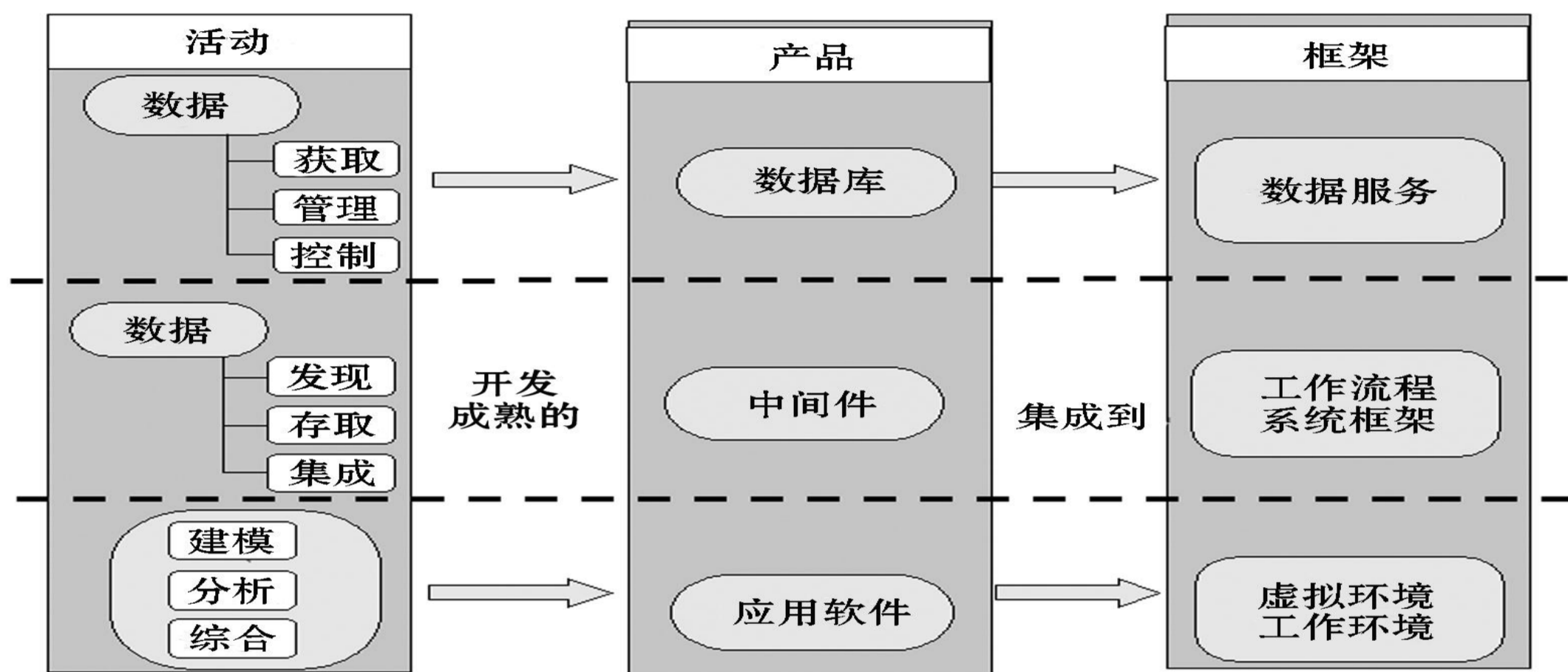


图 1 未来 LTER 信息化基础设施的活动、产品及框架图

Fig.1 Diagram of CI envisioned CI activities, products, frameworks

(1) 台站层次的信息化基础设施——支持数据获取、管理和质量控制的 CI 框架。通过开展 LTER 网络的数据获取、数据管理和数据质量控制的技术研究,建立 LTER 基础数据库,并应用信息技术构建面向服务架构的数据服务 CI 框架。该框架由数据服务、相关数据获取、数据管理和数据质量控制工具和生态学专业组成,为所有的联网研究及试验提供数据支持。该框架的不断完善,能够激励研究人员遵循标准协议,提供联网研究及试验的元数据,从而以此为根本可开发出强大的数据分析工具,并实现数据的安全存储。该基础设施的建设能够使信息管理人员为跨站联网试验提供系统设计思路 and 开发研究等方面的主要支持,如设计和维护数据库,开发数据质量控制工具和数据查询工具等。各站的信息化基础设施(包括信息管理人员)的建设将提高台站数据的采集质量,台站参与整个网络的自动和半自动化信息处理能力,台站的数据集成和综合分

析能力。各站信息管理人员不仅将参与台站的研究项目,同时也将参与整个 LTER 网络、全美乃至全球的信息系统建设。

(2) 网络办公室层次的信息化基础设施建设——支持数据发现、存取和集成 CI 框架。通过开展数据发现、获取和数据集成技术的研究,在网络层面不仅可以利用传统的数据仓库(Data warehousing)技术,同时也可充分利用知识表达(Knowledge representation)、语义解释(Semantic mediation)、数据溯源(Data provenance)等信息技术领域中的新技术新方法,开发出一系列用于数据集成的工具(中间件),从而构建一个基于科学工作流技术的数据集成框架。该框架将使分布在不同学科的异构分布式数据资源成为一个服务于专业主题的统一的数据产品,从而进一步提供数据的服务。

(3) 网络层次的信息化基础设施——支持建模、分析和综合的 CI 框架。通过一系列建模、分析

和综合处理等应用软件的开发,建立一个拥有分析、综合能力的工具及高性能计算能力的信息化基础设施,推动整个 LTER 网络的数据综合应用工作,促进相关生态过程模型和跨学科模型的开发与改进。未来的社会和环境综合科学需要集成模型、可视化工具以及科学工作流环境,建立包括不同类型应用软件和模型的复杂工作流(Jones 等 2006)。同时,也需要利用新的资源,共同开发和支持复杂的科学工作流,建立模型库,实现模型的源程序共享和模型验证。

(4) 网络层次的信息化基础设施——支持跨尺度协同工作的 CI 框架。随着生态学的综合、跨学科等学科交叉的日趋加强,跨领域和跨尺度的生态学研究项目取得成功关键之一,就是要在上述活动和产出的基础上建立一个虚拟协同复合工作环境,在该环境下,能够实现不同区域的科学家同时浏览数据、共同分析问题、探讨交流观点和意见。还可通过远程视频会议系统,共享界面、Web 服务以及其它工具,联合研究团队有效地明确未来发展的目标,共同探讨和解决所面临的科学问题。

此外,LTER 信息化基础设施建设战略合作伙伴关系的建立以及人员培训和教育计划也是未来 LTER 信息化基础设施建设的重要活动。

5 启 示

信息化基础设施建设是未来开展长期生态学研究网络层面科学的重要支撑。LTER 信息化基础设施规划是美国 LTER 战略规划的重要组成部分,该规划系统地分析了 LTER 信息化基础设施的现状、

面临的挑战及未来 LTER 信息化基础设施建设的活动和产出,对未来(几)十年的科学计划——社会与环境综合科学计划的实施也具有重要意义。

中国生态系统研究网络(CERN)及其他一些长期生态学研究网络的信息化基础设施的战略规划与发展方向,LTER 的 CI 建设提供了可以借鉴的样板,得到了启示。

(1) CERN 已经开展了与信息化基础设施建设相关的一系列研究活动,如长期生态元数据标准的制定、综合中心数据共享系统的开发和运行、台站数据信息系统的初步建立等,各台站也开展了不同传感器技术与数据传输技术研究,但目前的建设大多数比较零散,尚未进行系统的整体规划,美国 LTER 的信息化基础设施的部署方案和发展方向为我们的未来工作提供了一定的借鉴模式。

(2) 如何应用新的信息技术,开展 CERN 数据管理、数据集成研究、数据产品开发,CERN 模型数据融合平台的建立以及面向 e-science 的中国生态系统研究网络的建设,建立 CERN 的战略合作伙伴关系,形成一个系统的、充分支持和适应未来 LTER 发展和科学研究的中国生态系统研究网络的信息化基础设施框架,LTER 的 CI 建设经验将对 CERN 具有重要的指导作用。

参考文献(References):

- [1] James Brunt, Barbara Benson, John Vande Castle, et al. [EB/OL]. VLTER Network Cyberinfrastructure Strategic Plan-Version 4.2 <http://internet.lter.net/pdf> 2007.

An Introduction to LTER Network Cyberinfrastructure Strategic Plan(I)

NIU Dong^{1,2}, YANG Ping², HE Honglin³

(1. Research Center for Eco-Environmental Sciences, CAS, Beijing 100085, China;

2. Bureau of Science and Technology for Resources and Environment, Beijing 100864, China;

3. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: Cyberinfrastructure (CI) strategic plan is the important component of the Long-Term Ecological Research (LTER) program. Under the support of National Science Foundation (NSF) and based on the Core Scientific Program - Integrated Scientific Program for Society and Environment for LTER in the four key fields of the program, i.e. (1) Multi-site/Network Experiments, exports from multiple applications (2) Data Integration, (3) Modeling, and (4) System Architecture and Human Resources, exports from multiple fields made systematic analysis of the current situation and challenges concerning LTER Cyberinfrastructure. They advanced a series of Infrastructure activities and output in sites, network offices and networks for the CI construction in the future. This also provides inspiration for the program and development of Chinese Ecosystem Research Network (CERN) and other Long-Term Ecological Research Network in the world.

Key words: Long-Term Ecological Research (LTER) Network; Cyberinfrastructure (CI); Plan.