

生态系统研究与管理简报

立足科学 服务决策 促进人与自然和谐发展

2007 年第 1 期（总第 6 期）

1 月 15 日印发

美国长期生态研究（LTER） 计算机基础设施（CI）建设战略规划

中国生态系统研究网络综合研究中心 编译

【编者按】2006 年 12 月 18~19 日，中国科学院召开中国生态系统研究网络(CERN)科学委员会三届四次会议。本次会议讨论并决定推进 CERN 成果总结与未来发展规划。CERN 信息系统建设是该规划的重要组成部分。为了借鉴美国长期研究网络（ILTER）在信息系统建设规划的经验，CERN 综合中心组织编译了 ILTER 计算机基础设施建设规划，供 CERN 和有关机构参阅。

一、引言

编制该战略计划的目的是，为了说明美国长期生态研究网络（简称为 LTER）为实施目前确定的部分研究计划，在计算机基础设施（Cyberinfrastructure，简称 CI）建设方面应提前做的准备工作。集成性科学新的发展方向与计算机基础设施的发展是密不可分的。按照维基百科（Wikipedia）的定义，计算机基础设施（CI）是由美国国家科学基金会（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、数据存储、数据管理、数据集成，数据挖掘、数据可视化，以及其他计算和信息处理服务。在科学用途上，这种计算机基础设施是一种有效的技术解决方案，可以有效地将数据、计算机和人三者之间结合起来，从而获得先进的科学理论和知识”。同时，计算机基础设施还包括那些操作和维护设备、开发软件并提供支持、建立有关技术标准和规范，以及提供安全服务和用户支持等其他关键服务的个人和组织。

推动计算机基础设施建设的主要目的，是为了促进协同科学的发展。本文描绘了进入 LTER 计算机基础设施的新时代将会出现的情景，并详细阐述了通过最大程度地开展数据共享、信息综合以及知识生成，促进社会生态综合科学在多时空尺度上开展合作。长期生态研究网络（LTER）如要实现其科学任务，就必需拥有这些关键的计算机基础设施。建设这些必要的计算机基础设施，不仅要求在 LTER 内部建立全新的能力（见图 1），同时也要求在人员和技术方面进行新的重大投资。

——在人员方面：配备和培养包括数据管理和数据综合人员（以适应未来数据量的增加和对综合产品的需求）；应用软件开发和服务人员（以便推动数据集成的快速发展）；新的精通信息技术（IT）的生态学者，以及培训复合型信息专家的人员（其职责是制定针对复合型信息专家的培训计划并负责具体实施这些培训计划）在内的各种人

员。

——在技术方面：开发包括用于协作、沟通、数据获取和产生、数据管理和控制（curation）、数据发现、数据综合、知识表示、分析、合成和建模的各项技术。

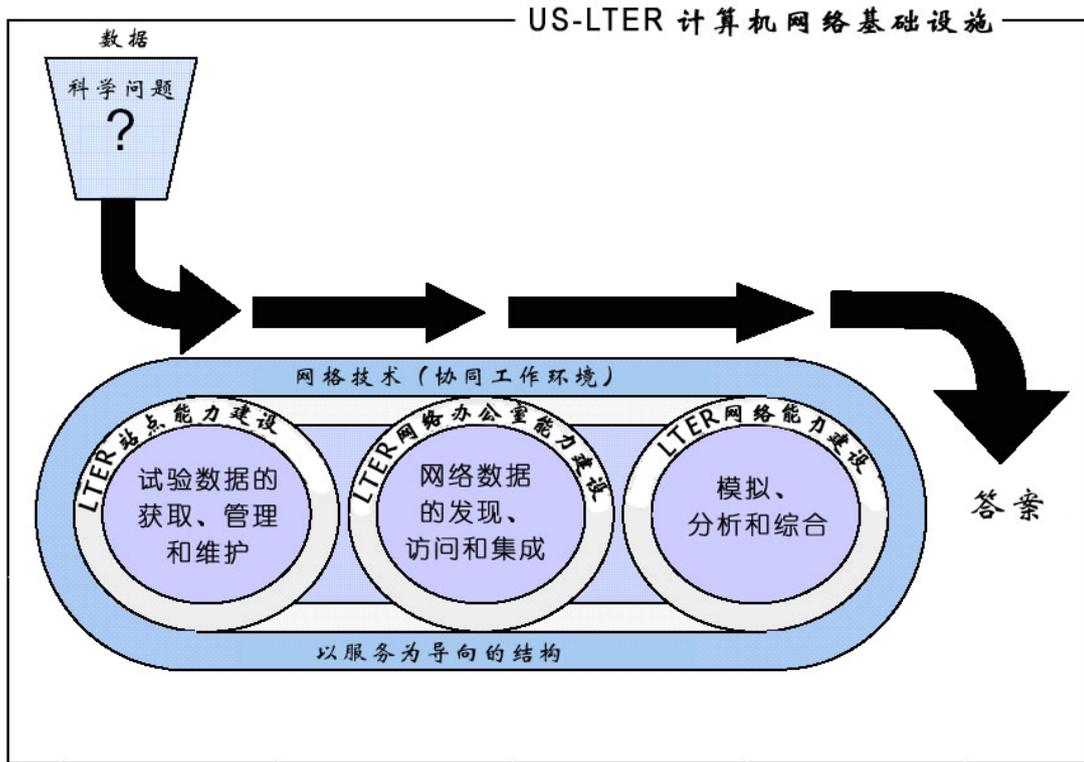


图 1: LTER 新型计算机基础设施示意图

我们（这里指美国 LTER CI 规划小组，下同）的下一个目标是组建一个高度综合的研究网络，来提高我们对多种时空尺度下的社会生态状况的认知和预测水平，并将生态学、社会学和地球物理学等学科有机地结合起来。要实现以上这个目标，前面所提及的在基础设施、人员以及技术方面的投资将是必不可少的。为此，LTER 网络正在制定一个科学计划，用以指导大尺度社会生态综合科学在未来十年（或数十年）间的发展。该计划初期的重点是研究多重胁迫力对人与自然耦合系统的长期影响。今后，LTER 计算机基础设施的建设，不仅将应对在支持研究者从事以上初始研究方面所面临的挑战，而且将应对下一个十年综合科学出现难以预料挑战。由于这些尺度上跨学科、多元以及多站点的研究工作将面临诸多的挑战，因此要求必须对现有

的计算机基础设施进行重大的改进。LTER 计算机基础设施的规范建设及其实施成功与否，将取决于具备专业知识的科学家们和信息专家，同时也将依靠 LTER 外部的信息学方面的专业知识和计算机基础设施系统。另一方面，LTER 内部计算机基础设施建设及其实施也将促进 LTER 外部的信息学专业知识和计算机基础设施建设的发展。

为了确定出我们在计算机基础设施方面所面临的挑战，并为迎接这些挑战更好地做好准备，一个由来自 LTER 科学技术中心、大型 IT 开发项目以及生态观测网络的 IT 专业人士所组成的专家组，与 LTER 计算机基础设施的规划人员共同开展计算机基础设施建设方面的规划工作。这样做的目的，是为了力求扩大规划过程中的专业知识范畴，促进各个项目之间的整合，并鼓励更多的合作者参与进来。这种合作主要集中在规划人员认为必需有计算机基础设施来辅助制定科学计划的四个关键的功能领域，即多站点和网络试验、数据集成、建模，以及架构和人力资源四大领域。项目人员按照这四大领域分成四个小组，每个小组均召开了多次会议，并对各个小组具体讨论的情况进行了汇总。该战略规划正是在这些总结性文件的基础上，并广泛征询了美国 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） 的意见而制定出来的。

在本战略规划中，我们首先（在第二部分）描述了 ILTER 科学规划工作中所提出的科学情景（science scenario），同时对通常的计算机基础设施以及 LTER 计算机基础设施的前景进行了简要的介绍。在第三部分中，我们根据调查和讨论的结果，概述了 LTER 网络计算机基础设施的现状。在第四部分中，我们确定出了计算机基础设施建设在推动整个网络科学的进展方面所面临的主要挑战（这些挑战是我们在整个规划过程中精心提炼出来的）。最后，在第五部分，我们提出了 6 个战略动议，这些动议将充分利用 LTER 网络和 IT 界所拥有的优势，来应对我们所确定的那些主要挑战。同时，我们也对这些动议需要花费的成本进行了估算，并列出了一个初步的优先实施进度表（这部分

内容缺失——译者注)。

二、科学情景以及支撑性计算机基础设施的前景

在此，我们提出了如下科学情景，来更好地阐释 LTER 计算机基础设施的具体需求。在 LTER 网络研究中，首先是对来自各个站点的基础信息进行综合，这就需要获取来自不同渠道的站点数据并对其进行管理。之后随着数据的范围和复杂性的增加，很快就需要一个集成的计算机基础设施，来完成整个网络的数据管理工作。

目前，美国 LTER 的一些站点已经记录了植物群落组成的长期变化状况。按照先前的假设，这种变化是由氮沉降、夜间变暖以及降水等发生变化所导致的。为了验证该假设的正确性，预测未来大陆尺度的变化格局，人们设计了一个融多站点实验和建模为一体的综合研究项目。在该项研究中，氮沉降和夜晚变暖所出现的变化被作为持续压力 (presses)，而降雨变化则被认为是短暂变化 (pulse)。

为了确定研究中需要处理的变化的量级大小，必需获得有关降雨量和降雨格局、氮沉降以及气温状况的背景信息。在试验样地，技术人员需要收集的数据包括植被群落成分，植物种的氮成分和土壤养分的数据。人们利用传感器网络来收集水热通量、土壤和大气温度、土壤湿度和干湿度有关的数据。新采集到的植被成分数据，将与现有的群落成分长期数据进行集成，用来分析试验样地植物群落的变化轨迹与整个研究站的植物群落变化轨迹有何不同。因此，科学情景中的第一个层次上的综合，就是对来自多个站点的新采集到的数据和现有的长期数据进行集成。

接下来，将用群落动态变化模拟模型，来整合从试验中获取到的各种知识。该模型可以预测群落组分，并可模拟个体植物在小型样地 (1-5m²) 中从每日到年度不同时间梯度内的分布、生长和死亡状况。该模型同时可以对不同深度的土壤水分含量逐日进行模拟，其对植物过程的影响按年度进行累加。目前在建立模型的过程中，也考虑到了

植被、土壤水和土壤结构的反馈作用。在所有站点对该模型进行验证，就是对气候和氮沉降在控制植物群落组成变化状况中所起作用的总体假设所进行的验证。

为了扩大研究结果的适用范围，预测未来气候和氮沉降变化对整个北美地区所产生的影响，就需要建立一个空间确定的植物群落组分模型（10*10km）。这就要求尺度的转换可从单个植物尺度上推到整个群落尺度。同时在重建各个 LTER 站点观测的时空格局以及植物群落变化的区域历史格局的基础上，对该模型进行校正和确认。如果将该模型与全球气候模型（GCM）以及对未来氮沉降率的预测相结合，将提高人们对植被群落组成未来潜在格局的预测水平。

1. 计算机基础设施

在该情景中，推动计算机基础设施建设最明显的驱动力主要有两个：一是从事科学研究必须要进行协调和合作；二是必须对大量的历史数据，以及来自新的数据采集方式、传感器和各个分布式数据源中所新获得的数据进行整合。此外，必须提供相应的支持，以便从多个站点获取实验数据，并对一系列复杂的传感器进行安装和管理。在第二阶段，必需对来自集成模型中的分布式数据源进行例示（instantiation）和确定参数。对模型的验证要求我们具备对多站点的数据采集和集成能力。最后一个阶段是开发新的模型，使研究尺度能够扩大到所预期的结果。同时，必需开发新的运算方法，以便从卫星数据中抽取有关植被群落组成的数据。此外，还必需开发新的工具，来确认植被群落的空间变化格局。

要实现以上这一点，就需要有一个真正综合性的计算机基础设施来支持它。而应用目前的计算机基础设施，要实现以上这一点将进展十分缓慢，即便可以的话，也需要数月或数年才来完成。

2. 计算机基础设施的前景概述

要对前面描述的计算机基础设施的情形进行概括，我们必需首先简要了解一下美国长期生态研究计算机基础设施的整体前景，它将包括下面几方面的开发工作：1）提供快速传输海量实验数据产品的服

务；2) 开发建立在可集成大量历史数据和实验数据，以服务为导向构架基础上的计算机虚拟环境；3) 开发虚拟合作环境，在该环境下拥有可用于知识发现和数据挖掘的全面广泛的工具和算法，同时提供用户友好的界面，方便用户登录、浏览生物信息，并实现这些信息的可视化以及评注等功能（见图 2）。要搭建该框架的各个组成部分，要求 LTER 的各个站点、网络办公室以及整个 LTER 网络在计算机基础设施建设方面采取统一的行动。

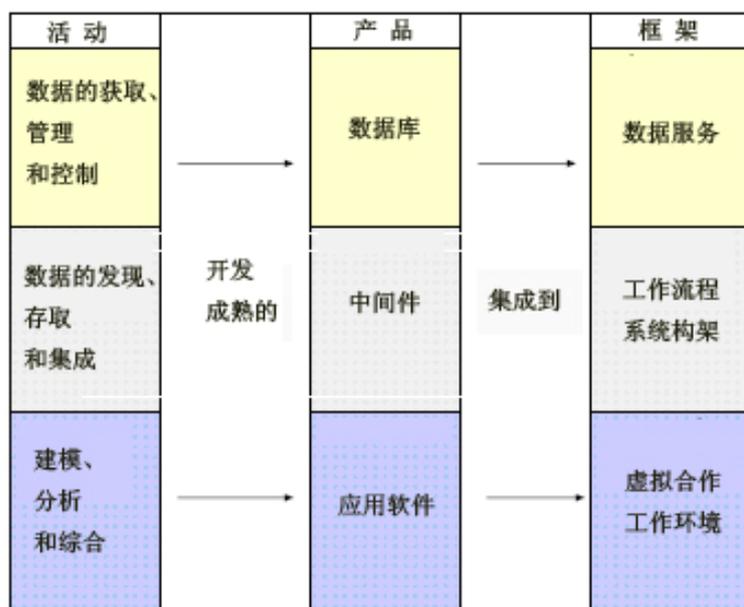


图2：拟建计算机基础设施的活动、产品及框架图

图 2 拟建计算机基础设施的活动、产品及框架图

3. 支持数据获取、管理和控制 (curate) 的计算机基础设施框架

一个由数据服务、相关工具和专业知识的网络框架，将有望向所有的联网研究和试验提供支持。该框架除了能产生规模效益外，还能激励研究者遵循标准协议，为他们提供试验元数据，从而开发出强大的分析工具，实现安全的数据存储。网络人员可以为跨站联网试验提供设计和开发方面的支持，比如开发登录定制数据的软件，设计和维护数据库，开发数据质量控制和数据查询的工具等。计算机基础设施在各个站点的组件和工作人员将提高他们采集高质量数据，以及参与整个网络自动和半自动化信息处理、集成和综合的能力。各

个台站的信息管理专业人员实际上不仅将参与特定站点的研究项目，同时也将参与整个 LTER 网络、全美乃至全球的信息系统建设工作。

4. 支持数据发现、存取和综合的计算机基础设施框架

我们所设想的对数据综合的支持，不仅是利用传统的数据仓储（data warehousing）方法，同时也是通过基于知识表示（knowledge representation）和语义媒介（semantic mediation）等新兴技术来实现的。一个典型的项目，如前面所提到的科学情景，一般会涉及分布在多个不同机构中的数据源和用户。那么，此类项目就需要一个成熟的信息基础设施，来实现信息的无缝连接、分析和存储，并将信息传输到分布式用户群那里。

5. 支持建模、分析和综合的计算机基础设施框架

我们认为，建立一个支持建模和分析的计算机基础设施，将达到以下几个目的：1）推动整个 LTER 网络的数据综合工作；2）促进可用于回答对于完成 LTER 任务非常重要的那些科学问题的分析工具和模型的改进及开发；3）对数据、模型及其成果进行存档，以服务于更广泛的用户群体。

6. 支持大尺度合作的计算机基础设施框架

要使此类大型项目取得成功，最根本的一点，就是要创造一个虚拟的合作环境。在该环境下，居住在不同地方甚至不同大陆的科学家可以共同分析、讨论、评注和查看数据。通过利用远程视频会议系统、共享界面、网页服务以及其它合作工具，合作团队就可以有效地确定、讨论和解决共同面临的科学问题。

除了虚拟合作环境的开发，影响 LTER 计算机基础设施开发的其他一些重要问题还包括：建立一个将虚拟合作环境搭载在上面的，以服务为导向的体系结构；建立计算机基础设施建设战略伙伴关系，共同开发、使用和安装应用软件和中间件；制定人员培训和教育计划，更好地将计算机基础设施纳入 LTER 的研究工作之中。

我们必须做出决定，决定上述资源在 6 个已得到公认的组织（生态研究站，跨站研究项目，提供规模效益的网络中心，资源开放的团

体，学术计划以及以盈利为目的的企业）中如何最有效地进行分配，并决定应采用什么样的形式（人力、资金，还是技术）来提供这些资源。不过在我们做出决定前，最重要的还是要清楚地认识到 LTER 现有计算机基础设施的实际状况以及我们将面临哪些具体的挑战。

三、LTER 计算机基础设施的现状

在制订新的 LTER 计算机基础设施建设战略规划前，我们必须首先了解一下美国 LTER 网络的现状及其优势，了解现有规划的内容及其制定的过程。由于美国 LTER 网络现有的计算机基础设施所具有的众多优势在其它类似的网络中都没有或刚有，因此要迎接这个雄心勃勃的新科学议程所带来的挑战，由美国 LTER 网络来承担这个任务，也许是再合适不过的了。美国 LTER 网络所具备的这些优势包括：

（1）长期数据：长期数据丰富、全面、存档完好并在线提供，是 LTER 网络的主要优势之一。

（2）行业认可的标准：制定并采纳了用于元数据的行业标准——即元数据语言（EML），以及用于站点信息管理系统中的行业标准。在全世界的生态网络中，美国 LTER 网络是第一个，也是最大规模地采用元数据标准的一个网络。由于现有的标准化元数据将极大地推动计算机基础设施建设的实施，现在已是这项工作产生回报的时候了。美国 LTER 网络所制定的有关站点信息管理系统的标准，已通过了整个生态界的同行评议，并得到了广泛的认可。

（3）开放数据政策：制定并采纳了有关发布、获取和使用 LTER 数据的政策，该政策明确地提出了对数据用户和数据提供者的要求。

（4）网络办公室：资助成立了网络办公室，其职责是为信息学、计算和通讯设施提供支持和指导。

（5）多样化的知识和技术：美国 LTER 信息技术团队所具备的多样化的知识和技术，已产生了众多创新的信息学解决方案。

（6）信息学研究：美国 LTER 学者们的任务之一就是推动生态

学研究的发展，他们正积极地参与信息学的研究工作。

美国 LTER 内部长期的合作，使得 LTER 信息技术领域始终能够顺应潮流的发展，并掌握最新的动向。这种优势主要得益于与美国几个国家级生态或计算中心的战略伙伴关系以及与生态信息合作伙伴的合作。

1. 美国国家生态分析和综合中心 (NCEAS)

此前，该中心在推动整个生态界的信息学发展方面做出了重要贡献。今后，它将在开发和支持计算机基础设施的综合集成方面发挥至关重要的作用。

2. 美国国家超级计算应用中心 (NCSA)

美国 LTER 网络一直在开发一个网络信息系统 (NIS)，来加快对数据以及综合产品的生成和应用，这些数据及其综合产品是从 LTER 各个站点所组成的网络及其合作伙伴的研究活动中获取的。网络信息系统建设战略规划中提供了大量与我们所讨论的计算机基础设施建设战略规划中所列的策略保持一致的信息管理战略。现有的网络信息系统建设战略规划主要关注于现有数据的应用，也就是说，通过制定和采用生态元数据语言来改善数据质量和数据发现，通过联合构架来增加现存数据的数量，并通过利用这些数据和基础设施的应用程序来推动数据综合的发展。网络信息系统将在各站点信息内容的开发和管理过程中，通过提供资源、指导和培训，为信息的标准化提供强有力的支持。目前，网络信息系统正在开发和配置可容纳 LTER 信息内容的应用程序，包括一个在线数据目录以及可利用这些数据来发现信息的应用软件。网络信息系统将通过应用关系数据库技术、共享中间件，通过广泛应用以及科学合作，来促进网络综合信息产品的形成。该系统将主要为 LTER 的科学界及其合作伙伴服务，但同时也将为 LTER 以外的广大的科学界、自然资源管理者、政策制定者以及大众提供一个进入 LTER 网络的“门户”。在此方面持续不断地开展下去，是 LTER 新的科学议程取得成功的重要保证。

3. LTER 各站点的计算机基础设施现状

在整个 LTER 网络计算机基础设施规划过程中，我们必须首先对 LTER 站点的计算机基础设施现状进行评估。在 2005 年 6 月进行的关于 LTER 站点计算机基础设施状况的调查显示，LTER 的各个站点均拥有十分强大的计算机基础设施能力。现将其中一些重要的特点归纳如下：

- LTER 的各个站点所能获得的（或长期拥有的）信息管理（或信息技术）专业人才存在很大的差异，LTER 的资金一般资助 0.5~3 个全职人员（FTE），某些站点可以从 LTER 以外获得资助，另外再得到 2.5 个全职人员的资助。LTER 大多数站点可得到一些机构的支持，用于其计算基础设施建设，而“数据中心”通常只能获得一个行政助理的运行费用，这一点有所不同。这些信息管理人员的主要工作是软件开发和数据录入等，其大部分时间用于站点的日常数据管理。
- LTER 站点的数据形式多种多样，从传感器流数据，到自动上传系统以及人工记录的野外数据都有。尽管所有的站点都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，但其元数据的质量、覆盖范围以及用于管理元数据的系统类型却存在很大的差异。此外，尽管 LTER 的大部分数据都已上线，但是用于数据传输的系统及其限制条件也迥然不同。
- 总体而言，美国 LTER 的各个依托机构都具有良好的联网设施，然而涉及到各个野外站，尽管这些野外站都能连通因特网，但其带宽及其服务水平则存在很大的差异。目前，美国 LTER 一半以上的野外站都有足够的带宽和无线基础设施来支持传感器网络，但是为传感器网络提供支持的人员和技术能力则显不足。
- 在 LTER 台站工作的科学家与该科学领域的中心就有关信息技术问题开展合作，部分站点则与计算机学者就信息技术问题开展内部或外部合作。

- LTER 的部分站点有商用远程视频会议的设施（比如 50% 的站能访问 Polycom-宝利通视频会议系统），但能访问个人台式电脑、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 各个野外站的人员一般很少使用其它的网络沟通工具，而电子邮件（E-mail）几乎成了专用的电子沟通工具。

从以上结果可以看出，如果要用一个词语来概括 LTER 各个站所具备的功能和条件的话，那就是“参差不齐”。这些结果充分说明了，LTER 网络必需切实扩大 LTER 各个站的技术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是放在计算机和通信基础设施以及信息管理方面，并聘用了几个计算机专家，但要应对大型项目的海量数据及其集成，并为未来十年的科学问题提供支持的话，其人员和装备力量都还不足。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 就必须将足够的资源投入到计算机基础设施的建设之中。

四、计算机基础设施在推动网络科学的发展方面所面临的挑战

在利用社会生态数据的巨大科学价值来认识和预测生命系统的反馈作用这方面，长期生态研究（LTER）和社会生态学界面临着众多挑战。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们所面临的问题（Maltev 2006）。比如，我们很难将分布在全美各地的众多科学家组织起来进行合作。此外，各地都缺乏训练有素的学者和技术专家。因此，在为综合科学提供集成的高质量数据集方面，我们将长期面临巨大的挑战（Stevens 2006）。

尽管美国 LTER 具备某些优势，可以使在网络工作的学者们在迎接以上这些挑战方面占得部分先机，不过在规划 LTER 计算机基础设施建设方案的过程中，我们也发现整个 LTER 网络面临一些特殊的挑

战，这些挑战包括：

(1) 各个站的计算机基础设施条件存在极大的差异性，并要获取、管理和控制来自网络科学研究中日益增加的数据量所带来的挑战。

(2) 对来自于网络科学研究，涉及多学科、多元和多站点的日益增多的数据量的集成和传输提供支持，以及对基于站点的生态研究中产生的标准不同的原始数据（这些数据在格式、精度、尺度、语义和质量保证/控制方面均有所不同）进行统一所带来的挑战（包括在使用历史性社会科学数据，以及在使用和综合来自网络外部的数据，如大容量的地球物理数据时出现的问题）。

(3) 推动在多重地理尺度上组织、研究团队分散在不同地方的日益增多的科学合作所带来的挑战（这种合作行为常常不是自发组织的，需要精心制定规划，将各种技术，如虚拟的合作环境，共享软件工具，远程会议技术等运用到科学活动中）。

(4) 推动和协调为整个 LTER 网络层次的建模和研究活动服务的计算机基础设施建设所带来的挑战。要应对这个挑战，需要我们大大提高认识和预测地区、大陆和全球生态系统变化的能力。同时要求我们必须做出决策，决定计算机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、处理和分析来自于跨站联网实验、传感器网络、卫星传感器以及建模过程中的网络尺度的大量数据，在提供共享的计算和数据服务方面所带来的挑战（在上述这些方面，是根本不可能为每个新布设的试验单独提供人力以及计算机基础设施的）。

(6) 在满足各个站点对受过专门训练的人员的需求，包括对受过其它学科知识培训的信息学专家以及精通信息学的研究生和学者的需求方面所带来的挑战。

五. 开发 LTER 计算机基础设施的战略动议

为了支持 LTER 所开展的主要科学活动，我们提出了以下 6 个战略动议，这些动议主要是为了建立整个网络在数据管理、数据集成和数据分析三个关键功能领域的的能力，以及建立整个网络在其它方面的能力，从而增强网络的整体能力，促进合作研究的发展，补充新的复合型人员，并建立以服务为导向的构架，使其成为一个服务于更大尺度、提供共享服务的综合系统。一些战略动议（如人员培训计划）充分体现了如何利用这些动议来应对 LTER 科学规划过程中所提出的挑战，而其它一些动议则利用整个网络的组织管理优势，将不同的实施内容集中在同一主题之下（如数据集成能力），以便统一满足不同领域的科学活动的需求。

1. 建立 LTER 计算机基础设施获取、管理和控制日益增长的数据的能力

随着来自网络科学研究的数据的不断递增，我们在获取、管理和控制这些数据方面也面临挑战。要应对这些挑战，必需加强各个站的能力建设，其主要方法是为这些站点配备相应的人员、提供站点信息管理系统，以便确保数据的质量，最大限度地保持数据的吞吐量，并为网络科学数据（包括来自站点的实验数据）的综合提供支持。LTER 的各个站点必需增强其采集高质量数据的能力，以及参与整个网络层次的自动和半自动信息处理、集成和综合的能力。同时，它们还必需确保在台站工作的信息管理专家既能真正参与针对某个站点的计划，也能参与整个 LTER 网络、全美乃至全球的信息系统建设。这些能力建设包括以下内容：

——为台站配备支持网络信息系统和维持高质量数据最大吞吐量的人员，这些人员可能包括网络管理员、信息管理员、程序员、传感器技术人员，以及受过有关卫星、传感器和空间数据综合知识培训的专家

——为台站提供开展长期数据服务的技术设施，这些技术设施可能包括硬件、大容量存储器、软件和传感器

——为网络办公室配备开发数据传输标准和提供网页服务的人

员，纳入整个网络信息系统。同时为台站配备具体执行这些标准和服务的人员。

——对台站人员进行新技术的培训

(1) 基本原理

LTER 各个站采集和管理的数据，是 LTER 在各个站、跨站以及整个网络层次开展的科学研究的基础。因此，要很好地应对计算机基础设施建设所面临的 6 大挑战，很大程度上将取决于 LTER 各个站自身的能力以及各站之间的协同能力。新的综合科学研究的发展，要求能很方便地在线获取各个站的完整数据。尽管试验对研究人员来说是一个常见的研究手段，但当研究者试图进行跨站联网试验时，由于很多常用的工具和方法的适用性大大降低，因此网络层次的试验对研究者来说仍将构成很大的挑战。有足够的证据表明，数据管理是许多跨站联网试验中最薄弱的环节，这往往是由于缺乏这方面的资源，以及试验人员所具备的专业知识有限并且很不稳定所致。要为综合科学研究提供支持，就必需最大程度地保证从野外采集到的高质量数据的吞吐量，并确保这些数据能存储到集中式的门户中。应用无线技术的嵌入式传感器网络，可以提供新的时空尺度上的数据，从而使产生多站联网的标准化数据成为可能。维护这些传感器网络，以及对这些更大尺度上的大量数据进行处理，包括数据的自动筛选，都需要增加在台站的工作人员。即便数据上线且保存完好，对各个站点的数据进行集成也是一项非常艰巨的任务。鉴于各个站的数据管理系统存在较大的差异性，开发和实现此项功能要求网络中心和各个站点均做出重大的努力。

(2) 方法

LTER 网络要建立此项能力，必须配备充足的硬件和人力资源，并开发出功能强大的站点信息系统（这些系统将用来采集、管理和控制数据的质量）。为了充分发挥网络的功能，各个站点必须实施整个网络层次的标准，并与 LTER 网络信息系统进行对接。要实现这一点，要求各个站点必须制定相应的解决方案，使站点的信息系统能够协同

工作 (interoperability)。利用规模效益, 可以开发出各个站都可通用的工具, 用来支持数据的获取、发现、存取和集成。同时, 对台站人员进行有关数据采集、信息管理和数据分析的培训, 将从根本上增强 LTER 站点的能力。此外, 为台站提供充足的经费, 使其能够配备足够数量的信息管理人员, 确保台站的数据能够满足需求, 使科学活动不至于受到数据和元数据 (元数据是对数据进行解释的数据) 存取的限制。

(3) 实施

要实现增强台站充分参与网络科学动议能力的近期目标, 要求 LTER 网络必须配备相应的人员和技术力量来完成以下这些工作:

- 改善数据采集的方法 (野外数据录入系统, 传感器网络, 自动质量保证/控制, 元数据的自动产生), 为站点人员培训新的技术。
- 提高由站点生成的元数据语言 (EML) 文档的质量 (如提供更多的内容; 提高各个站点的元数据的标准化程度, 如关键字和计量单位保持一致; 限制关键字的词汇总量, 以便可以直接访问元数据的浏览界面), 使站点的信息系统之间能够协同工作。
- 开发可自动获取站点数据、为站点提供所需信息的系统 (带有各个站点认证的安全网页服务界面; 网络服务标准, 为站点信息管理员进行提供站点网页服务的培训)。
- 开发 LTER 站点数据的单点访问系统, 以便为从事多站点综合联网研究的科学家提供特定的查询服务。
- 建立相关的工具和技术转让机制, 使站点提供的数据能够满足个别的全局计划的要求。

要改变站点现有计算机基础设施参差不齐的状况, 就必须开发一套所有 LTER 站点都需要的最基本的功能, 来实施这个战略计划, 同时制定一个各站点为提供这些最基本的功能必备的软硬件和人员所需要的大致预算 (通过从台站收集需求信息, 进行审核和汇总, 然后

再次征询台站反馈意见的一个多次反复的过程来实现)。

2. 建立 LTER 计算机基础设施发现、存取和集成日益增多的数据的能力

预计今后 LTER 的数据量及其类型将不断增加, 为了从这些数据中提炼出相关的生态知识, 必须首先对这些数据的质量进行控制(评估数据的质量, 将数据与本体(ontologies)和元数据相耦合)、集成(找到和建立各个组成部分之间的一致性)和传输(以科学研究可用的一种形式进行提供)。要为提供多种基于多学科、多元和多站点的集成网络数据产品提供支持, 是我们面临的一大挑战, 要迎接这个挑战, 要求我们在软件开发、技术和分析支持以及长期基础设施建设方面, 必须采取既有针对性又涉及面广的对策, 这些对策包括:

- 为 LTER 网络办公室配备设计和维护网络信息系统(来统一站点数据服务)的人员, 如程序员、软件开发人员和数据集成专家。
- 为 LTER 网络办公室配备(为站点执行网络标准, 以及为整个网络利用网络信息系统开展综合研究而提供)分析和技术支持人员, 如数据和系统分析人员。
- 为合作研究, 以及通过知识表示和本体开发技术解决数据异质性问题的工作组提供资助。
- 为 LTER 网络办公室配备开发和布设网络信息系统的基础设施(可能需要增加大容量存储器和计算资源等持久性计算基础设施)。

(1) 基本原理

按照生态学者最通常的理解, 数据集成就是发现、获取、解释和综合数据的过程。通常情况下, 这个过程是因某个科学问题无法用我们现在掌握的可分析数据对象来直接解决而引发的。我们认为, 只有当相关数据可以被组合成一个可分析对象时, 才能形成我们研究某个科学问题所必需的数据和信息, 因而就必需进行数据集成。而在生态学者看来, 数据集成更是一个涉及方方面面(现在看来主要是人力)

的过程，包括在数据发现、获取和解释等各个领域内尚未解决的问题，特别是在连续性方面的挑战。

要使数据集成过程中各个方面，包括采集、存档和传输的标准，以及用于数据发现和解释的工具的能力得到增强，需要我们继续努力，解决两个层次上的问题，一个是与数据产品提供者相关的问题（如为科学产品开发数据仓储服务、分布式查询服务和知识服务），另一个是与那些与数据使用者相关的问题（提供便于分发、获取和解释现有数据集的服务）。

（2）方法

这里的数据集成是指生成一个相容、一致、可用的科学资源（即一个网络信息系统）的整个过程，而不是简单地将多个独立的数据资源合并在一起的过程。在此，我们将探讨在数据集成的过程中，关键应该如何使整个过程保持完全的一致。

● 确立样板（prototyping）

对于数据服务提供者层次上的数据集成来说，人们普遍认为是将多个不同种类的数据源合并到一个统一的数据源中，以便向用户提供完全一致的数据服务。由于数据源及其相互之间的关系各不相同，也给数据提供者层次上的数据集成增加了难度。这个层次上的数据集成，是数据仓储（data warehousing）过程或分布式查询系统的一个功能组成部分。数据仓储涉及数据的获取、提取、转换以及将数据上传至可访问系统的整个过程。但必须首先确定数据源、集成方法、属性映射(attribute mapping)以及数据质量评估方法之间是否存在关联。尽管数据仓储是一个极有价值且在当前不可替代的工具，但它还是无法充分满足未来生态学发展的需要。与数据仓储方法所不同的是，分布式查询系统不集中存储数据，只是提供一个数据的总体框架(federated view)，并且总是返回数据源。由于分布式查询系统要求自动进行映射，因此很大程度上依赖于知识管理系统来实施查询。我们认为，在数据集成、数据仓储和分布式查询系统的建设方面建立一个原型，对于该项战略规划的实施，将起着至关重要的作用。

● 研究

要改变数据的异质性，我们就必须在知识表示（knowledge representation）和语义媒介(semantic mediation)技术的应用研究和实际应用方面进行大力投资，包括集中那些关注知识管理的战略专家小组，增加开发本体论（ontologies）和提供知识服务的软件开发和数据集成专业技术人员等。然而，本体论的开发工作，目前仍取决于科学家之间能否达成广泛的共识——这是一个涉及众多社会和科学问题的复杂问题。开发用于挖掘和汇总现有的科学理念和术语的新工具和新算法，将为以上进程提供十分重要的帮助。在探索如何才能发现、存取、解释和集成异质性数据的可能性时，整个 LTER 数据标准化方法的开发和推广工作，具有十分重要的价值。同时，在制定标准时必须加强沟通，以便大幅度地减少因随意造成的异质性数据（例如，在多站点的联网试验中，必须确保对采样尺度、处理方法以及有关试验设计其它方面的方法进行了充分细致的讨论）。此外，讨论结果应当形成统一的数据存储方法、统一的数据类型（data typing）、统一的变量命名和语义，以及最大程度上减小由于时空尺度不同所带来的数据不兼容性。

（3）实施

为了在实施新的科学议程中取得规模效益，LTER 网络办公室必需提供包括数据质量保证、分析和控制在内的数据管理服务，在需要网络层次管理的研究数据集成中发挥协调作用。这项协调工作包括与各个站点，也可能与服务于多站点联网试验的数据中心之间进行沟通，其目的是通过减少数据采集方法的差异性，并在一个联网系统中统一处理数据服务，达成数据集成方法的标准。LTER 网络办公室尤其适合做这项工作，但它在计算机基础设施方面的投资还必须大幅度地增加。

以上这项工作的实施，将由网络办公室的一个开发小组负责领导，并与 LTER 的信息管理专家以及主要的信息建设伙伴进行合作。目前，这种实施方法已经成功地应用于网络综合元数据目录和工具的

开发工作中，同时也成功地应用在了支持整个网络实施生态元数据语言（EML）标准的工作中。近来，该小组与美国国家超级计算应用中心(NCSA)成功地完成了一个试点项目，验证了网格技术在解决网络信息建设所面临的特定挑战中的有效性。网络办公室的这些实施工作将由一个网络层次的咨询委员会负责监管，该委员会将由 LTER 网络办公室的人员、LTER 的学者以及信息管理专家组成。

我们将建立一个基于“**数据空间**”（dataspace）(Franklin等.2005)概念的数据集成战略框架，这是建立大型综合性数据库的另一种途径。在该框架内，某个组织中的参与者负责管理一个包含了该组织中的各种格式和来自于不同地点的所有数据和信息的数据空间。利用数据空间概念框架，LTER网络将可以把所有的数据毫无例外地汇集在一起，用于模型的开发，同时也能集中做好特定需求领域内的数据集成工作。

在 LTER“数据空间”为我们提供了一个总体概念框架的同时，我们还必须采用一整套各不相同但又互为补充的方法，针对以下优先研究领域开展数据集成：

- 在试验设计处于优先考虑对象的地方所获得的试验数据，如果能建立在一个全球计划的基础上，将对其更为有利。
- 如果可行的话，数据采集后可望形成不断增值的数据产品的数据集成工作，应当纳入数据仓储的工作流程之中。
- 数据采集后可望形成一次性增值数据产品的数据集成工作，可利用人工数据仓储技术来完成。
- 对于所有存储的数据来说，都应当继续确定和开发结构化（structural）和本体化（ontological）元数据，以使为某些专门的分析工作提供半自动数据集成。
- 应当利用 GEON（地球科学网络）和 SEEK（生态知识科学环境）项目的方法，来开发现有数据库的注册和集成工具。

3. 建立 LTER 计算机基础设施处理日益增多的建模和分析活动的的能力

为了推动和协调 LTER 在网络层次的分析 and 建模活动，显著提高我们认识和预测区域、大陆和全球生态系统变化的水平，要求我们必需在计算服务、软件开发和人员配备方面进行大力投资：

- 为各个站点以及致力于网络层次的分析 and 建模活动的网络中心配备相应的人员（程序员和软件开发人员），增加对研究人员的资助；
- 获得包括新的硬件技术、高性能计算机、并行处理器、大容量存储器和高吞吐量在内的计算服务；
- 投资相关软件的合作开发，包括开发可视化工具，将不同的程序语言的模型和多元控制结合在一起的软件，开发基于数据和模型的管理工具以及网络层次的站点协议；
- 为 LTER 网络办公室开发和安装一个持久性的数据和模型存储器（可能要求增加大容量存储器和计算资源等持久性计算基础设施）。

该项动议的内容还包括为研究人员提供必需的资源，为分析活动和合作建模提供计算支持，以及为多个模型的存储提供支持等。

（1）基本原理

在认识生态系统的结构、功能和动态变化，生态系统对气候变化的响应，生物地球化学循环，外来物种的引入，以及生态系统服务的变化给人类行为带来的变化等问题时，建模和先进的数据分析起着至关重要的作用。基于过程的模拟是整个网络层次的科学活动中一个不可或缺的组成部分。要解决包括 LTER 网络各种生态系统类型所涉及的问题，要求我们必须将模型与分析应用软件、试验数据、观测、遥感影像和空间数据库集成在一起。

● 先进的分析应用软件

新的科学议程要求我们必须对今后 10~20 年的分析方法做出重大的转变。要做出这些转变，要求我们必须首先改进分析工具，这些工具包括可视化软件、利用分布式计算资源的软件、将不同程序语言的模型与多元控制结合起来的软件，以及基于数据和模型的管理工具。

由于在改进研究软件使其可用于批量生产和使用方面的投入非常高昂（**prohibitive costs**），因此要求我们必须开发出可快速重新配置和重复利用、功能强大、可升级的软件工具。

- 建模

当美国 LTER 计划着手解决有关网络层次的科学问题时，模型的开发在未来网络的成功中将发挥日益重要的作用。尤其是可以利用模型来进一步分析持续压力（**press**）和短暂变化（**pulse**）这两种不同的驱动力（**dichotomy of drivers**），从而将相互作用的不同时间尺度连接起来。同样，空间变异性涉及从站到地区再到跨大陆等各个空间尺度的变化，如果仅仅是利用试验来进行研究，成本将会十分高昂，而通过模拟则能得到很好的解决。在中间尺度上的运行过程非常重要的条件下，只能通过基于过程的、空间尺度明确的模型，才能将小尺度的动态变化外推至更大尺度的空间范畴和更长的时间段上。另外，包括生态系统、水文、气候和社会科学的综合模型，对于归纳试验的结果、分析人与自然系统间的相互作用将起着至关重要的作用。尽管模型在提高我们的认识水平和进行预测方面具有非常强大的功能，但由于受到各地的技术和资源的限制，加之缺乏集中建模的人员和资源，开发和应用服务于网络层次科学研究（**network-level science**）的模型的这项工作目前做得仍然不够。

- (2) 方法

该项动议将采用组织和指导与分析模拟活动相关的计算支持服务，以及合作开发和集成新的分析工具的方法来进行。为了鼓励和持久支持相关的建模和分析活动，必须建立一个计算框架和合作基础设施。为了满足这些需求，我们拟定了一个模拟和分析活动动议，来开发和实施以下这些内容：

- 可升级的计算资源

该项动议要求能获得更多新的硬件设施和技术，包括高性能计算机、处理某些应用程序的并行处理器、网格技术、大容量存储器以及高吞吐量的能力（**high throughput capacity**）。

- 先进的分析环境

利用科学工作流 (scientific workflow) 系统作为分析工具和框架, 是该领域内最有前途的新兴技术, 采用该技术可以进行应用软件和模型的开发和集成工作。所谓科学工作流, 是指分析步骤的流程或网络, 它既可能包括数据库的存取和查询步骤, 也可能包括数据分析和挖掘步骤, 以及其它一些包括在高性能计算集群上的高强度计算工作步骤 (Ludaesher 等.2005)。目前, SEEK 项目正在开发一个可以让生态学者设计和实施科学工作流 (Jones) 的分析和模拟系统 (AMS), 该系统可以连续不断地获取数据资源和包括模型在内的各种服务, 并将其并入一个可重复利用的工作流程中。该系统是在开普勒 (Kepler) 系统——一个基于整个生态界和多个工程的科学工作流系统的基础上建立的。同时, 它将在某些方面进行革新, 以满足 LTER 新的科学议程的需要。

- 面向整个社会的公用知识库 (community-based repository)

对有关环境的数据产品进行保存, 已被公认为对科学研究工作具有重大的促进作用: 它在节省大量由于不必要的重复性数据采集活动所导致的成本的同时, 有助于提高我们开展新的分析和产生新的成果的能力。同样的原理也适用于对数值模型 (numerical models) 的保存 (Thornton 等.2005)。保存完好的数据集和模型, 将提供形成新的研究成果所必需的数据来源和方法的具体内容, 从而可以对不同的研究结果进行综合并对新的假设进行验证。此外, 利用这些保存下来的模型 (archived models), 还可以与来自评估或者政策研究中的其它模型的结果进行比较, 从而确定出存在哪些不确定性。模型的源代码也可以让其他人了解这些模型是如何处理单个过程的。

(3) 实施 (本项内容无, 译者注)

- 建立 LTER 开展日益增多的协同工作的能力

当前, 研究活动的开展不仅涉及多个地理分布尺度, 同时也涉及多个传统科学领域。利用一个统一协作的计算机基础设施, 将有助于减小这些差异。为这种合作以及最终为整个研究界提供支持是一个重

大的挑战，要迎接这个挑战，要求我们必须建立一个高效、友好、持久的计算机基础设施。因此，我们要求尽快并长期持续地获得以下这些支持：

- 配备协同虚拟工作环境的程序和软件开发人员
- 提供购买视频会议技术所需的资金
- 配备开发集成分析工具软件的人员
- 提供购买增强型计算机基础设施所需的资金

(4) 基本原理

与计算机基础设施相集成的研究活动，将使研究人员可以与分布在不同地点的同行共同开展日常性的工作 (Atkins 2003)。为了实现这一点，我们必需首先知道应用基于网络的协作方式与传统的工作方式到底有何不同。有关研究表明，如果研究人员不能很好地掌握网络协作技术，地理分布就可能成为影响研究工作开展的因素。因此，那些具有成熟的社会结构和管理方式以及联系紧密的协作更容易取得成功 (Cummings and Kiesler 2003)。此外，除非可能具有非常显著的效益，否则人们通常不愿意选择过于复杂的设计和难度太大的学习方式。因此，用户界面的质量好坏、网络响应的时间长短以及能否提供各种所需的工具，对于协作研究的成功与否将起到至关重要的作用。

(5) 方法

我们将采用多种途径，来进一步满足协作研究的需要，这些途径包括：购买和安装可直接使用的视频会议和网络技术设备；合作开发和布设协同工作环境框架；开发和安装框架内部的分析工具；与社会学家开展合作。将社会学家纳入信息系统的建设工作之中，对于我们成功地迎接这一挑战将是必不可少的。

- 虚拟协作环境

开发虚拟协作环境，可以使居住在不同地点乃至不同大陆的科学家能够共同分析、讨论、解释和查看有关的数据。登录视频会议系统、共享界面，获得公用服务和其它协作工具，可以使研究团队有效地确定、讨论和解决相关的科学问题。利用门户技术，可以让数据采集者、

管理员、分析师和研究人员互相获取数据、文本和图像，促进彼此对项目任务和目标的沟通和共识，同时这种技术还可以提供一个网页内容管理(包括参考书目在内)解决方案及其并行开发文档(合作编写)。在开发过程中，有很多好的兼顾社会和技术方面的合作方式，能帮助我们成功地应对这些挑战。为了将这些合作的激励因素应用到开发和实施基于计算机基础设施的生态科学合作环境中，要求我们必须将生态学与信息技术和组织学习(organizational learning)的专业知识结合起来。为了创建这个协作环境，必须建立相应的战略合作关系，并配备相关的软件设计和编程专家。

- 分析工具

在将虚拟协作环境尽可能地纳入 LTER 日常活动中的同时，我们必须提供相应的分析工具来产生直接的效益。研究人员必须拥有有效的工具，哪怕是低水平的可视化和分析工具，才能掌握该系统的使用方法及其性能。为了推动这项工作的进展，我们必须为研究人员提供与软件开发人员合作开发这些分析工具的机会，并留出足够的程序开发时间。

- 视频会议

公用视频会议(VTC)技术有助于不同站点间开展合作以及共享信息。目前，LTER 网络办公室已安装了一个 48 频道的共享视频会议设备，为事先预定好的或者临时决定的会议，从一对一的交流到大型会议提供技术支持。今后，LTER 网络必须为各个站点及其依托机构增加新的硬件和网络连接设备，以便它们能够利用这个设施以及相似的基础设施来加强与整个网络的交流。

- 网络的接入

LTER 站点与第二代因特网(Internet2/NLR)的接入，将增加 LTER 网络的数据吞吐量，并使站点可以获取到基于 LTER 网络和其它基于 GRID 的资源。建立一个从野外传感器网络连接到研究人员的台式电脑，再到本地和远程数据中心的高性能终端到终端的网络连接，可以最大程度地保持数据和信息的吞吐量。目前，美国 LTER 站点的绝大

部分依托机构，特别是大学都已与第二代因特网相连了。不过，整个 LTER 网络连接因特网的状况并不一致。对于一些依托机构来说，需要帮助它们与本地的千兆电话接入网点（gigaPOP）最后连接起来；对于其它的依托机构来说，则需要帮助它们与城市中心的商业千兆电话接入网点（gigaPOP）或类似的网络系统连接起来。在 LTER 野外站中，只有一半的站点目前拥有 T1 水平（传输速率可达 1.544Mb/s 的通讯线路）或更大的数据吞吐量。与 LTER 站点的依托机构的需求类似，也必须为 LTER 站点提供增加新的硬件和网络接入设施所需要的经费，才能提高整个 LTER 网络的网络接入水平。

4. 将计算机基础设施用于社会生态学研究、教育和培训活动中

要将包括先进的分析和集成工具在内的新的计算机基础设施用于研究过程之中，必需对研究生和学者们进行培训，使其从事的研究活动能够充分从新技术中受益。同时，也必须对技术人员进行培训，使其能够掌握最新的技术及其应用。要应对以上挑战，通过制定一个人员培训教育计划就可以实现。该计划要实现的目标有以下几点：

- 为在 LTER 台站从事数据采集和数据管理的信息管理人员和技术专家提供有关新技术和新方法的培训；
- 向从事长期生态研究的自然和社会学者提供应用先进的信息工具的培训；
- 建立一个受过多方面知识培训的信息管理人员队伍，这些信息管理人员能很快组织一套针对 LTER 的同行和合作伙伴的标准化课程和培训材料；
- 开发适用于远程视频会议、网络研讨会、远程学习，以及利用其它方式可以进行远程信息技术培训的培训材料。

（1）基本原理

当 LTER 网络参与计算机基础设施的实施工作时，只有当整个组织及其内部文化发生重大转变，有利于推动新的科学研究方法的发展时，LTER 的众多研究人员才能通过轻松地获取和分析来自各个方面的数据，从而有效地参与多学科的综合研究工作。这些新的方法中

有理论性的（即我们如何有效地参与跨学科的研究），也有技术性的（即我们应用何种工具来完成这些研究）和社会性的（即我们如何以准确、可信和恰当的方式将研究结果传达给其它领域、决策者和公众）。信息技术的发展，使得信息的获取、集成、传输、分析和交流更为有效。但是，用户必须树立清晰的目标，并且知道利用哪些技术来很好地实现这些目标，才能使信息技术发挥真正的效应。

在不远的将来，新的 LTER 科学议程将产出大量的数据和信息。公众期望我们能充分利用这些投资，在他们看来，持续数十年的人力开发过程是不合理的。目前，各个机构还没有制定对科学家进行信息学知识的培训计划。通过为研究生和研究人员设计和教授有关信息学知识的培训课程，将有效地解决以上这个问题。这将使来自多个学科、从事复杂环境问题研究的研究生和研究人员能够运用最新的技术和计算机基础设施来解决跨学科研究的设计、实施和交流问题。

要有效利用新技术，并开发出整个 LTER 网络层次的计算机基础设施创新解决方案，就必须为技术人员和信息专家提供培训，使其能够掌握当前及最新的技术。利用这个培训计划，可以加强与从事跨领域开发的计算机专家的合作，促进技术的转让。LTER 网络必须在新的层次上开展协调，以便为更广领域的科学议程提供计算机基础设施方面的支持。同时，台站的信息管理员也必须参与到协调工作和利用新技术所必需的培训中。随着越来越多的 LTER 站点开发了传感器网络，来应对数据量大量增加所带来的挑战，其它的技术人员也可能接受统一培训并从中受益。

（2）方法

制定统一的培训计划，可以满足 LTER 网络对受过综合知识培训的信息专家以及对精通信息学的研究生和研究人员的需要。这些培训计划包括在具备良好的现场实习条件的集中设施内举办培训班，也有其它一些可以在本地或远程获得的培训手段。为了满足某些特定群体的需求，一个受过综合培训的信息管理人员队伍将很快组织相关的培训材料，来对有关人员进行培训。针对某些特殊的需求，可利用远程

视频会议、网上研讨会以及其它远程学习方法，来构建一个远程学习的环境。此外，将制定出有关的评估方法，来评价培训班和其它资源所取得的效果。在确定培训需求及开发课程的过程中，需要有目标用户群体的参与。

5. 针对专业研究人员和研究生的培训计划

参与培训的对象为研究生、大学教师和研究人員，以及专门从事环境问题研究的专业研究人员。通过以下方法，可以建立一支应用相关技术开展生态研究的科学家队伍：

- (1) 提供强调信息和知识的管理、集成、分析、综合和传播的有关方法和技术的培训。
- (2) 组织学习那些合理有效地实施的应用范例。
- (3) 指导个人将新的方法应用到实践中去。

培训计划还将提供在元数据和数据库设计等传统信息学领域，以及在嵌入式传感器网络、科学工作流软件、分布式计算和知识表达等前沿技术领域内的指导。在介绍前沿技术前，培训班将向学员们提供有关基础研究领域的背景知识。

6. 针对技术人员的培训计划

针对技术人员的培训内容是由 LTER 所有的信息管理人员和信息技术合作伙伴确定的有关信息学的知识和技术（有关培训主题见下表）。每个培训课程均包括课堂讲授、案例分析（将这些方法应用于环境问题研究的案例）以及现场实习等多种形式。

表 1 LTER CI 培训计划

网络服务	构成网络SOA（服务型系统结构）的工具和技术
空间数据系统	遥感 / GIS / 空间数据引擎
生态信息学	数据管理，数据存档和数据质量控制；包括 EML（生态元数据语言）
LINUX	系统管理培训和认证
RDBMS (关系型数据库管理系统)	MySQL, Oracle
数据质量	质量保证，质量控制和质量管埋

7. 建立面向整个社会（community-based）的服务型系统结构

一个可升级的基于整个社会的服务型系统结构（如下图所示）可以提供有效的数据服务，确保研究人员能够安全高效地获得在站点数据库中存储的数据，并安全高效地获得数值分析和建模所需要的服务，以及大尺度多站点联网试验所需要的服务（这些服务包括传感器网络、卫星传感器和高性能计算服务）。在开发和安装应用软件和中间件时，ILTER网络必须建立一个计算机基础设施的战略合作关系，这些合作关系包括：

- 支持主要的合作伙伴（如美国国家生态分析和综合中心 NCEAS）与 LTER 各站点和网络办公室合作，共同开发和提供面向整个社会的服务
- 向 LTER 网络办公室从事集成软件开发的人员和程序员提供支持
- 支持 LTER 站点参与开发、布设和利用公共服务的工作。

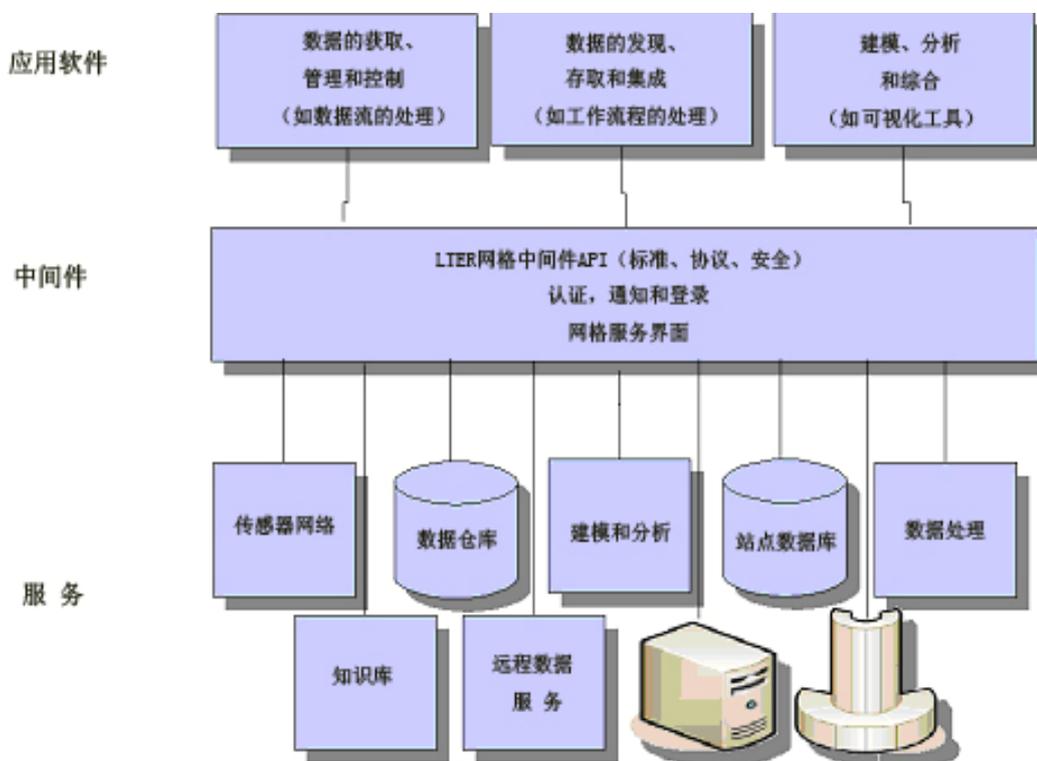


图 3：公共服务型系统结构的示意图

(它对传输和安全系统进行了标准化,以便向所有的数据和服务提供者和使用者提供统一的界面。)

要建立一个能使研究人员连续不断地分享和利用现有的和历史性观测资料的计算机基础设施,必须首先建立一个基础性的系统,用户通过一个安全连续的过程,可以从这里发现和获取以下这些数据和资源: 1) 本地和远程数据; 2) 包括存储器和高性能计算机系统在内的分布式计算资源; 3) 开展同类研究的合作者或机构。可以利用包括 Globus (一种网格软件) 在内的众多网格软件包所提供的框架,来实现以上这一目标。

(1) 基本原理

生态研究人员和网格软件开发者之间的需求常常存在矛盾,前者习惯于“在某一段时间段进行某一项试验 (one experiment at a time)”的方法,而后者期望建立多个能同时处理大量工作的系统。另外,正如 Hunter 和 Nielson (2005) 所指出的,目前的网格计算系统虽然性能相对稳定,但仍然很难掌握,并且受到现有硬件、软件和程序开发语言的限制。这对设计更优化的软件开发人员以及有效利用这些软件的研究人员来说,都是一个很大的挑战。带着这些问题,我们可以通过与所有相关组件的开发者密切合作,从现有的工具中选择一些有用的组件,构建一个可经常性利用网格数据的功能强大的软件平台。通过在不同的条件下进行应用,同时满足常规需求和特殊需求,可确定出哪些是可重复利用的组件,而不会对用户要求定制的环境造成损坏。正如前面所指出的,开发这个平台要求各个项目之间相互合作,来指导这些工具的开发工作。

(2) 方法

● 系统构架概述

LTER 计算机基础设施所设想的基于网格的服务系统,将向 LTER 分布在全美各地的研究站点、传感器网络、合作研究以及 LTER 网络其它的公共服务提供支持服务。为了实现这一点,需要通过一个网格 POP(接入点)模型来进行集成(如图 4 所示)。每个接入点 (POP) 都

将提供一个 LTER 资源与其它资源的接口（interface），使其经由第二代因特网（internet2/NLR）与 LTER 网络系统进行连接。通过站点的接入点（Site POP）接口，可以将网络硬件（服务器，本地硬盘以及千兆网络接口）与包括业内标准协议和应用软件在内的软件包连接起来，而这些协议和应用软件可将该站点与其它站点和外部资源进行安全的无缝连接。

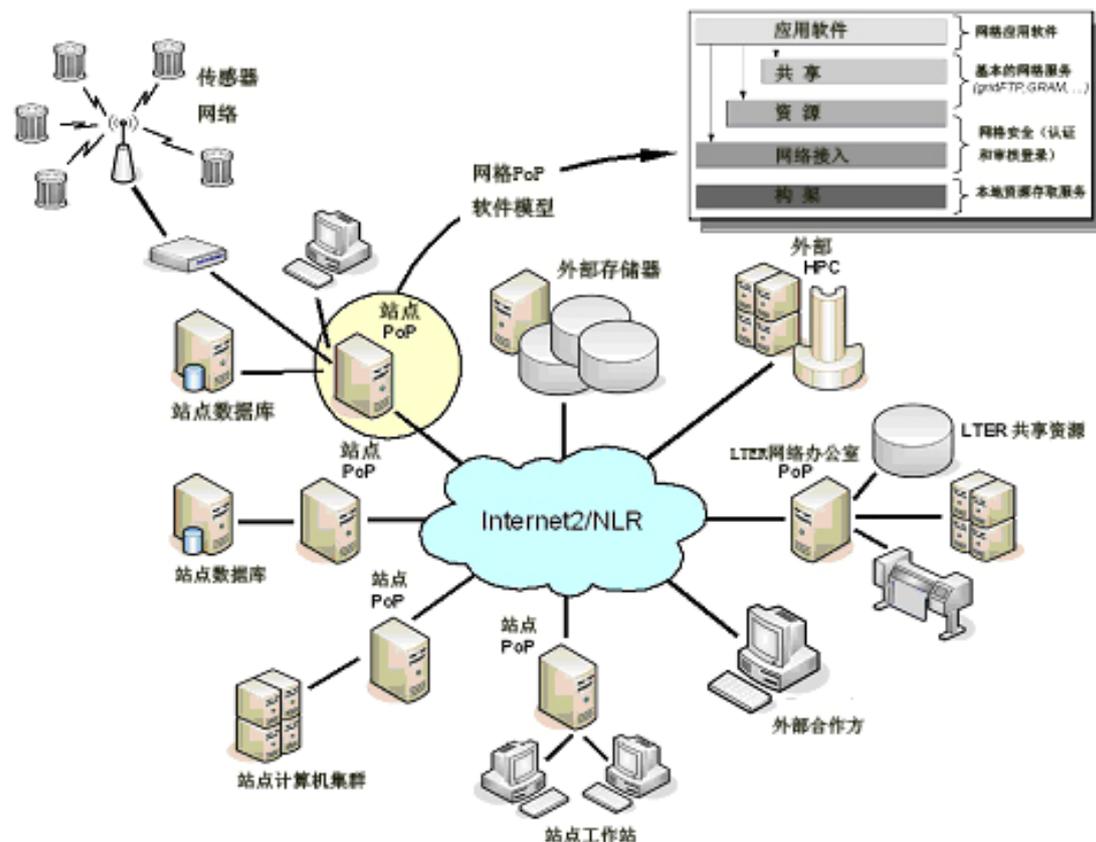


图4 LTER网络网络POP模型结构示意图

- 软件包

站点 POP 软件包（见图 4）必须提供一整套的服务，允许该站点与其它站点或外部资源进行双向连接，但同时必须确保其安全性和合理性。LTER 研究人员所要求的一项重要服务就是安全服务，即要求所有获取站点数据的请求必须遵守本地认证，以及获取 LTER 网络数据政策的规定。因此，POP 软件服务必须收集与资源利用和数据传输相关的审查信息。

(3) 实施

首先，我们将继续与超级计算机机构和组织（如美国国家超级计算机应用中心，圣地亚哥超级计算机中心等）建立战略伙伴关系，以便充分利用它们在设计 and 实施面向整个社会的应用程序和服务方面的专业知识和技能。作为这个战略伙伴关系的一部分，LTER 网络办公室信息建设小组将密切合作，开发出专门定制的应用软件，来满足 LTER 网络和各个研究站点的需要。我们将把已得到实际验证和普及的网格应用软件，作为我们设计模型的基础。此外，我们还能找到其它一些合作伙伴，例如 NEESGrid, BIRN(生物医学信息学研究网络), TeraGRID 等。我们将与这些有经验的合作伙伴，以及那些期望成为先期使用软件模型的 LTER 站点一道，推出软件模型的原型，开始尝试着连接和共享 LTER 的分布式资源。

其次，一旦这些先期使用模型的站点验证和接受了这一原型，我们将把它作为一个模板应用到其它的站点。按照这个过程，在保证站点有合理的集成时间的同时，我们可以对模板不断进行改进。此外，我们还将努力把其它的网格连接资源引入到这个网络中。这些资源在 LTER 的站点的外围，包括高性能计算组，可视化工具和其它的站外存储器等。

为了支持该项计算机基础设施的战略动议，必需为 LTER 网络办公室增加进行软件开发、集成、安装和维护的人员。预计 LTER 的每个站点在对必要的软件进行初始安装时，会要求得到相应的技术支持和资源，同时也要求得到定期对软件进行维护以及对系统进行升级的技术支持，此外还有某些站点要求配备为本地建立模板的人员。

编 译：赖鹏飞、李俊杰
审 阅：何洪林
责任编辑：于秀波

生态系统研究与管理简报

立足科学 服务决策 促进人与自然和谐发展

(2006年目录)

- 9月10日 第1期 国家生态环境科学观测试点站发展的回顾与展望
国家生态环境科学观测研究站专家组
- 10月18日 第2期 生态系统观测与研究应关注的25个科学问题
中科院生态系统研究网络领导小组办公室
- 11月2日 第3期 景观方法在湿地保护与合理利用中的应用
Peter Bridgewater (《湿地公约》秘书长)
- 12月15日 第4期 生态系统服务的集成模拟和评价
Robert Costanza (美国佛蒙特大学教授)
- 12月18日 第5期 美国长期生态学研究网络的战略规划
中国生态系统研究网络综合研究中心

(2007年目录)

- 1月15日 第1期 美国长期生态研究(LTER)计算机基础设施(CI)
建设战略规划
中国生态系统研究网络综合研究中心

关注中国生态系统监测、研究、评估、管理与政策进展

主办单位:

国家生态系统观测研究网络综合研究中心
中科院生态系统研究网络综合研究中心
中科院生态网络观测与模拟重点实验室

编辑部:

于贵瑞、欧阳竹、于秀波(常务)

通信地址: 北京市朝阳区大屯路甲11号
中科院地理科学与资源研究所
CERN综合研究中心

邮政编码: 100101

传 真: 010-6486 8962

电子邮件: yuxb@ignsnrr.ac.cn

网 页: <http://www.cern.ac.cn>