

国家科研信息化 基础环境建设与实践*



廖方宇 汪 洋 马永征 王彦桐 周园春 许海燕

中国科学院计算机网络信息中心 北京 100190

摘要 科研信息化基础环境是在科研信息化劳动工具中，满足共享需求、提供共享服务、支撑科研活动的软硬件系统和信息化环境。它涵盖了以硬件设施为主的“硬”服务环境、以软件和数据为主的“软”服务环境，以及包括协同工作环境、运维平台在内的运行管理和环境。文章简要综述了国外科研信息化基础环境建设的最新进展，阐述了我国国家科研信息化基础环境的发展现状，通过近年来中科院在科研信息化基础环境建设中的实践与经验，分析了我国国家科研信息化基础环境建设中的不足，并针对问题提出了关于我国国家科研信息化基础环境发展的建议。

关键词 科研信息化，科研信息化基础环境，科研网络，高性能计算，科学数据

DOI 10.16418/j.issn.1000-3045.2016.06.006

2000年，英国科学家提出了科研信息化（e-Science^[1]）概念，其内涵是采用互联网络技术和广域分布式的高性能计算环境建立的一种全新科学研究模式。随着信息技术的飞速发展和信息化应用在各个领域的渗透，本质来看，科研信息化是科研生产关系的信息化^[2]，是指采用信息化技术和方法辅助科学研究，目标是依靠信息化手段，以前瞻性、战略性、全局性布局来优化创新要素配置，加快科技生产力的改造和升级，促进创新驱动发展战略的实施。科研信息化在我国信息化体系中处于源头创新的引领地位，是我国创新驱动发展的关键引擎。科研信息化基础环境则是指建立在信息技术和信息装备基础上，满足科研活动需求、提供共享服务、支撑科研活动的软硬件系统和环境。

国家科研信息化基础环境是建立在现代信息技术高速发展基础之上的国家级软硬件融合服务环境，能够为全国各类科技要素机构以及科学家群体的科技创新活动提供专业化、集成化、自助化的科研信息化服务平台，主要包括：（1）以高速科研专用网络、高性能计

*修改稿收到日期：2016年4月22日

算资源、科学数据海量存储资源等为代表的“硬”服务环境；（2）以计算模拟软件、数据分析软件、观测数据和计算数据等为代表的“软”服务环境；（3）以基于社会网络的社交和协同工作环境、硬件运维平台和资源统一调度平台为代表的运行管理和服务环境。

国家科研信息化基础环境的主要作用是基于高速专用科研网络，连接国家重大科技基础设施集群、野外观测台站、各类计算中心、存储中心以及各科研院所、国家实验室、企业研发中心、高校等科技机构和科学家群体，实现全国各类科技要素间的集约集成、互联互通、资源共享，形成支撑新型科研模式以及服务社会经济发展的科研信息化环境，促进科技投入效用最大化，助力科技创新。

1 国外科研信息化基础环境建设进展

欧美发达国家和地区为了提升国家科技竞争力，在国家科研信息化基础设施方面制定了中长期发展规划，并由政府出资长期稳定支持，科研信息化基础设施的服务能力得到了长足的发展。主要体现在：科研信息采集装置全面具备了无线传输能力；高速传输网络的通用带宽已超过 100 Gbps，骨干传输速度向 1 Tbps 发展；数据存储 I/O 吞吐速度已达 100 GB/s 级；高性能数值模拟装置向百亿亿次（E 级）迈进；信息化基础设施连接了各类科技设施和国家级实验室、研究院所和大学。

（1）在科研数据的观测采集方面，数据获取的方式随着技术的发展发生了巨大变化。得益于精密芯片制造和无线传感技术的发展，科研信息采集装置全面具备了无线传输能力，欧美发达国家已将无线传感器网络部署到城市、农田以及各类生态系统等科学研究区域中，全面地搜集物理世界的各种特征信息。如美国国家生态观测站网络^[3]（National Ecological Observatory Network, NEON）在全美范围内部署了约 15 000 个传感器，采用混合组网方式开展长期、大尺度生态变化观测；美国海洋观测计划^[4]（Ocean Observatories Initiative, OOI）通过可

靠的无线传感器网络获取分布于海洋各处的浮标锚泊观测数据；Argo 全球海洋观测网^[5]计划组织世界各国在全球范围内部署 3 900 个海洋浮标，用于大尺度全球气候变化观测，目前已经有超过 3 900 个浮标正通过卫星将数据源源不断地传回地面，组成一个全球 Argo 海洋观测网。

（2）在高速科研网络方面，为应对大数据应用的需求，在科研网络资源倾斜性支持政策的扶持下，发达国家的高速科研专用网络传输能力和传输速度大幅提升。美国能源科学网（ESnet）^[6]创建于 20 世纪 80 年代中期，是由美国能源部科学办公室资助、劳伦斯伯克利国家实验室运行和管理的一个数据通信网络，提供通用、有效、可靠、高性能的网络基础设施，促进美国能源部的科技人员间的协作。ESnet 以打造世界领先的网络基础设施、能力和工具为目标，2009 年通过骨干网专项投入 6 300 万美元，实现了骨干传输网络的 100 G 传输带宽，2012 年度和 2013 年度分别投入 3 450 万美元和 3 200 万美元的运维经费，保障 ESnet 的安全稳定运行。GÉANT^[7]是欧盟为欧洲地区提供科研及教育信息化方面公共服务的基础平台，正在计划部署先进传输设备和转换平台，全面升级长达 5 万公里的 GÉANT 骨干网，支持 GÉANT 核心网络实现最高达 2 Tbps 的传输能力，使欧洲 4 000 万用户从中受益，并有望在欧盟范围内得到全面部署。在国际互联方面，GÉANT 通过欧盟 FP6、FP7 及 Horizon 2020 计划支持与其他国家科研教育网络实现高速互联，包括与拉丁美洲地区网络互联的 ALICE 项目、与地中海地区互联的 EUMEDCONNECT^[8]项目、与亚太地区互联的 TEIN 项目、与中国互联的 ORIENTplus^[9]项目等。仅中欧合作的 Orient 一条线路 GÉANT 获得欧盟 FP6 与 FP7 计划每年 100 万欧元的稳定支持。GÉANT 内部骨干网建设和国际互联战略的实施，为欧洲各科研领域数据平台建设和国际合作的有效开展提供了良好基础和坚实保障。

（3）在科研数据存储及数据中心方面，数据存储 I/O 吞吐速度已达 100 Gb/s 级，在数据管理方面，朝着模

式自由、易于复制、提供简单 API 以及支持海量数据的发展方向,以支持科学大数据的分析处理、高通量高性能科学计算。以 Google、Facebook 为代表的互联网创新企业在数据中心建设的技术创新上取得了巨大成功。整个数据中心系统的硬件环境、软件系统都应围绕应用需求来设计,数据中心技术创新也带动了应用模式的创新。Google 发明了 MapReduce 处理框架、GFS 存储体系,建立了支持横向扩展的应用架构。与上述应用架构相匹配,采用 DIY、ODM 模式自行设计设备机柜、服务器及存储设备等,在低成本条件下满足应用要求,并实现了高可用、高效的目标。同时,在数据中心选址、建筑设计、内部布局、模块化设计、可再生能源、自然冷却等方面的一系列技术革新,都体现了创新颠覆的思想,极大提升了能效比,为环境保护做出了表率。

(4) 在高性能计算方面,研发具备百亿亿次(E级)计算能力的高性能计算机正成为各国及其科研机构的目标,美国、欧盟和日本均已启动了面向E级超级计算机的研究计划。据估计,E级超级计算机极有可能在2020年研制成功。美国橡树岭国家实验室、阿贡国家实验室、劳伦斯·利弗莫尔国家实验室于2014年2月宣布,将联合研制峰值性能达20亿亿次的下一代超级计算机,其速度将是目前最强超级计算机的10倍,预计2017—2018年完成。“欧洲百亿亿次级软件计划”将联合产业界和政府机构,帮助用户在未来10年内从千万亿次(P级)超级计算提升至E级超级计算。“欧洲先进计算合作伙伴(PRACE)”^[10]计划部署一个泛欧Peta-Scale生态系统,并计划在2020年前达到E级的运算性能。2013年,日本文部科学省表示计划从2014年春季开始,将着手研发拥有超级计算机“京”100倍性能的“E级超级计算机”,力争在2020年前后完成研发任务并投入使用,日本政府拟为此项目投入总额1000亿日元的研发经费。

(5) 基于国家科研信息化基础设施提供的服务,欧美发达国家通过连接和融合各类科技设施和国家级实验室等科技创新机构,逐步打造完善的、全链条的国家科

技创新生态协作系统。美国通过ESnet,在能源部所有的实验室以及由能源部支持的大学之间建立高速连接,并与100多个其他网络进行互联,以便科学家们摆脱时间和地理位置的困扰,有效地利用能源部的研究设备和计算资源开展研究工作。ESnet直接接入了美国能源部下属的十大国家实验室,连接泰坦(Titan,橡树岭国家实验室)、米拉(Mira,阿贡国家实验室)、NERSC中心(劳伦斯伯克利国家实验室)、Cascade(西北太平洋国家实验室)、红杉(Sequoia,劳伦斯·利弗莫尔国家实验室)和Vulcan(劳伦斯·利弗莫尔国家实验室)等数据中心和计算系统。美国能源部依托ESnet实现了计算资源和存储资源的聚合,构建了面向国家重大任务的一流科研信息化基础设施。欧盟的GÉANT已连接了欧盟40个国家的8000个研究机构,支持的应用包括从医学、气象到行为艺术等广泛的应用领域。

2 我国国家科研信息化基础环境建设的现状与实践

2.1 我国国家科研信息化基础环境的发展与现状

过去20多年来,我国科研工作者逐步建设了一批科研信息化基础环境,不断尝试将信息技术和信息化手段应用到科学研究和工程建设中,极大改善了我国科技创新的信息化环境,促进了科研模式的变革。

近年来,我国在超级计算机方面投入较大,突破了异构融合体系结构、高性能处理器、高速互连、高密度组装、高效冷却、系统可靠性、并行系统软件等关键技术,成功研制了天河、神威、曙光、深腾等系列高性能计算机^[11]。特别是自2010年以来,“天河1号”和“天河2号”6次在世界超级计算机排行榜“TOP500”中占据第1,在国际上引起高度关注。

在高速网络方面,我国的科研机构 and 高校主要接入中国科技网、中国教育网和其他商业网络。中国科技网是中科院领导下的全国性计算机互联网络,是学术性、非赢利性的信息化基础设施。中国科技网(图1)已形成

在对网络动态调度、分布式高性能计算、学科数据存储及容灾、各类模拟计算软件、数据分析处理等软硬一体的整体信息化基础环境的管理、运行和服务平台系统才刚刚开始规划和设计。

《国家“十二五”科学和技术发展规划》^[13]中明确指出了“重视建设信息化基础设施”，但国家层面尚未出台与科研信息化相关的具体计划。总体来看，我国还未形成国家层面统一的科研信息化基础环境，主要以各部委的发展规划和重大专项的形式来牵引科研信息化的发展，但各规划和专项之间缺乏统筹协调，客观上形成了资源投入分散、政策滞后等问题，制约了国家科研和创新能力的提升。

2.2 中科院科研信息化基础环境的建设和实践

在过去的20余年里，中科院通过持续实施院级信息化专项，逐步形成了中科院信息化基础环境。建设了覆盖我国主要地区、连接国内主要科研院所的中国科技网，通过与欧美主要科研学术网络的高速互联，为科研机构 and 科学家提供了高速、安全和可靠的优质网络服务；建设了服务全院的超级计算环境，并作为科技部指定的中国国家网络北方主节点承担中国国家网络的运行与管理工 作，聚合通用计算能力超过2 PFLOPS，GPU计算能力超过3 PFLOPS，支持了2 000余个国家自然科学基金、“973”计划、“863”计划、中科院战略性先导科技专项等重大项目，累计提供超过2.5亿CPU机时；自1986年启动“科学数据库”工作以来，积累和整合了一批宝贵的科学数据资源，建设了包括北京中关村数据中心、怀柔备份中心和12个分布全国的区域分中心，形成达50 Pb的分布式海量存储与处理环境，面向全院提供云存储、云归档和云处理等服务，支撑重要数据资产的容灾备份、长期保存、共享服务与增值应用（图3）。

“十二五”期间，中科院初步建成科研信息化云服务体系，有效支撑中科院乃至全国科研创新活动。为院内各类用户提供公共的基础设施服务、平台服务、软件服务，重点建设了“科学计算子云、科研软件服务子云、科学数据子云”等云服务，以重大科研活动需求为

牵引，以面向科技创新的需求为核心，构建了一批跨机构、跨地域和跨领域的科研信息化应用平台，强化中科院信息化基础设施之间的相互操作性，实现信息化环境与重大科研项目（工程）有机融合。结合有关学科领域及重大科研项目的需求，定制和扩展出一批面向重大领域科技整体创新活动的云服务平台，通过整合计算、数据、模型等学科领域的科研资源，直接服务于某种类型或某一领域科技创新活动。

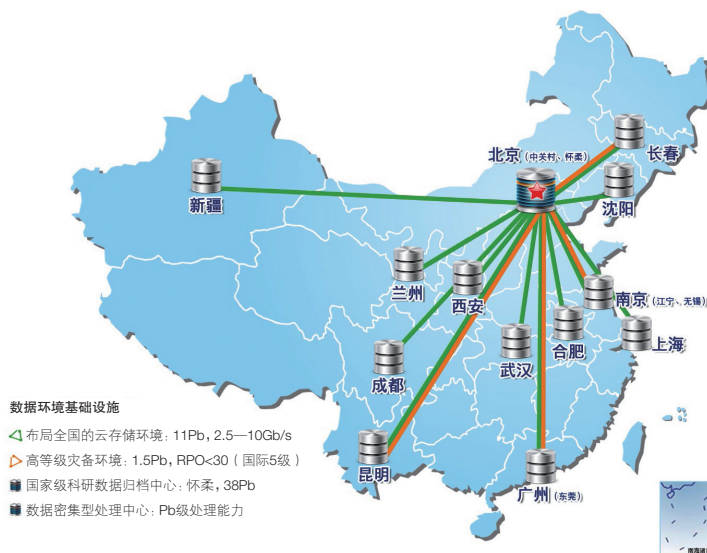


图3 中科院数据存储环境基础设施

中科院科研信息化基础环境在互联网、计算、数据硬件能力3个方面均基本形成了多级架构和云化环境建设，初步实现了全院的资源整合与优化调度，并培养了一支经验丰富的信息化运维与服务团队，不仅能为中科院的科学家群体提供信息化基础环境服务，而且向全国科技界辐射了科研信息化在当今大数据驱动科技创新中的巨大作用和影响。

2.3 国家科研信息化基础环境建设中的不足

过去20多年，以中科院为代表的国家科研机构紧跟国际科研信息化的步伐，边建设边探索，建设了一批科研信息化基础环境，取得了一系列进展。但仍然在诸多方面存在问题。

(1) 科研信息化是信息化的先导力量。当前,我国经济正处于从要素驱动、投资规模驱动为主向以创新驱动为主的发展转型期,信息化已成为驱动国家经济社会发展的重要先导力量,科研信息化更是整个信息化的先导力量。然而,通观我国信息化发展现状,科研信息化却成为我国唯一没被“叫响”的信息化领域。科研信息化的重要作用未能在全社会得到共识,其公益性定位也没有得到国家层面的认同,导致许多科研信息化基础环境的依托单位仍然需要通过承担其他项目为信息化基础环境的运维筹措资金,国家科研骨干网络每年仍需要向商业运营商缴纳数千万元的信道费。

(2) 由于历史原因以及科技管理体制的特点,我国缺乏科研信息化的顶层设计。各主管部门主要在各自工作范畴内,部署并开展了科研信息化的一些应用,且倾向于将科研信息化视为一般信息化工作进行管理,客观上造成了投入资源分散、共享不足、数据封闭、政策滞后等问题,导致科研信息化基础环境各元素之间没有完全打通,而且服务能力不相匹配。如目前我国高性能计算能力已达世界领先水平,但在计算算法、模拟软件、科学数据等方面远落后于国际先进水平,削弱了整个信息化基础环境应有的服务交付能力和对科技创新的实际支撑作用。

(3) 科技创新活动和科研信息化未能协同发展。由于缺乏国家层面的统筹考虑与顶层设计,各个分散的科研信息化基础环境缺乏与国家的重大科技布局相对接,形成了科技创新活动和科研信息化基础环境支撑“两张皮”的现象。现有不多的科研信息化基础环境,没有能与一些重大科技基础设施和科技布局形成匹配的网络环境、数据存储环境以及高性能计算环境,缺乏学科领域的自主的高性能计算模拟软件、大数据分析处理软件对学科领域科技创新活动的支持。

(4) 尚未形成健全的科研信息化基础环境建设与运行维护的投入机制,对于信息化运维服务以及服务团队的评估、评价体系不明确。

3 我国国家科研信息化基础环境建设建议

习近平总书记指出:“没有信息化就没有现代化”。当今科技创新必然站在信息化的肩膀上来发展,能够适应科研范式的变革才能赢得竞争的优势。世界各发达国家在实施科技发展战略中,无一例外都将国家科研信息化基础环境作为国家科技创新布局中重要的、不可或缺的基础设施。我国应更加充分重视科研信息化基础环境的战略意义,尽快出台能够实施的建设计划,建设国家级的科研信息化基础环境,促进科技创新,提升科技竞争力。

国家科研信息化基础环境应从国家层面进行顶层设计、统筹布局,在建设过程中充分考虑国家重大科技布局的地域和学科上的分布特点,合理规划、集约资源,加强基础性、战略性、前沿性科学研究和共性技术信息化基础环境的持续投入,建设与国家重大科技布局相匹配的国家科研信息化基础环境。

国家科研信息化基础环境作为国家公益性的战略平台,应纳入国家科技专项中予以考虑,建立一套可评估、可考核的评价机制,保障国家科研信息化基础环境的长期稳定运行。

国家科研信息化基础环境一方面结合国家重大科技布局,面向国家重大战略和重大科学问题需求,支撑国家科技创新。另一方面,通过信息化基础环境汇聚科技界的资源,面向企业乃至全社会提供科技知识服务,降低企业创新和大众创新所必须的资源获取、加工处理和创新应用等门槛,推动信息化与经济社会各领域深度融合,服务“双创”和经济社会发展。

参考文献

- 桂文庄. 什么是e-Science?. 科研信息化技术与应用, 2008, 1(1): 1-7.
- 张亚平, 谭铁牛. 国家科研信息化战略研究咨询报告. 2015. 4-13.

- 3 National Ecological Observatory Network. [2016-02-01]. <http://www.neonscience.org/>.
- 4 Ocean Observatories Initiative. [2016-02-01]. <http://oceanobservatories.org/>.
- 5 Argo. [2016-02-01]. <http://www.argo.net/>.
- 6 Strategic Plan FY2014-FY2023 ESnet, 2013, 1-3.
- 7 GÉANT pan-European network. [2016-02-01]. <http://www.geant.org/Networks>.
- 8 EUMEDCONNECT. [2016-02-01]. <http://eumedconnect1.archive.dante.net/>.
- 9 ORIENTplus project. [2016-02-01]. <http://www.orientplus.eu/>.
- 10 Partnership For Advanced Computing in Europe. [2016-02-01]. <http://www.prace-ri.eu/>.
- 11 谢向辉, 胡苏太. 中国“863计划”高性能计算的发展. 科研信息化技术与应用, 2015, 6 (4): 3-10.
- 12 叶玉江. 国家科技基础条件共享平台现状与发展. 中国科研信息化蓝皮书. 北京: 科学出版社, 2016: 325-331.
- 13 科技部. 国家“十二五”科学和技术发展规划. 2011.

Construction and Practice of National Cyberinfrastructure for Scientific Research

Liao Fangyu Wang Yang Ma Yongzheng Wang Yangang Zhou Yuanchun Xu Haiyan

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

Abstract The cyberinfrastructure for scientific research is a suit of hardware and software systems which can provide general information services and sharing services for scientific research activities. It mainly includes three parts, (1) hardware services, such as high speed scientific research network, high performance computing resources, and mass storage resources for scientific data; (2) software services, for example, calculation and simulation software, data analysis software, and scientific data services; (3) collaboration and resource management platforms, for instance, social network environment, collaboration working environment, operation and maintenance of the hardware platform, and resource unified platform. This article briefly introduces the important status of cyberinfrastructure in national strategic infrastructure around the world, and reviews the latest progress and state-of-the-art of National Cyberinfrastructure for Scientific Research (NCSR). In China, scientific researchers have constructed several cyberinfrastructures for scientific research over the past twenty years. This article concludes the practice and experiences of establishing cyberinfrastructure for scientific research in Chinese Academy of Sciences, and discusses the major problems faced in the development of NCSR: (1) NCSR has not been considered as the public welfare at the national level and reflected its strategic importance; (2) Compared with other fields in our country, NCSR lacks of overall planning; (3) NCSR's distribution and ability do not match with the national key science and technology infrastructures; (4) The investment mechanism of infrastructure construction, the mechanism of operation and maintenance, and the evaluation system of service and career development have not been formed. This article also suggests for the development of NCSR: (1) NCSR is the key infrastructure for promoting scientific and technological innovation-driven development strategy. Our country should invest and construct NCSR as soon as possible; (2) It is important to consider the geographical and disciplinary distribution of national key science and technology infrastructures when constructing NCSR; (3) As a public welfare platform, NCSR should be considered in national science and technology projects; (4) NCSR can also provide scientific and technological knowledge service for companies and public, reducing the threshold of scientific and technological innovation, and serving mass entrepreneurship and innovation strategy.

Keywords e-Science, cyberinfrastructure, scientific research network, high performance computing, scientific data

廖方宇 中科院计算机网络信息中心主任，研究员，中国计算机学会计算机安全专业委员会副主任委员，中科院计算科学应用研究中心理事长。曾担任中科院信息办副主任，负责中科院信息化工作的组织协调，组织编制了《中科院“十二五”信息化规划》；曾长期从事空间科学与探测的工程任务和型号管理，担任国家“863”计划“深空探测”重大项目专家组成员等。E-mail: fyliao@cnic.cn

Liao Fangyu The professor and the director of Computer Network Information Center, Chinese Academy of Sciences (CAS). He is also the vice chairman of Professional Committee of Computer Security of China Computer Federation, and the director of Center of Scientific Computing Applications & Research, CAS. As the former deputy director of the Informatization Office of CAS, he was in charge of informatization affairs of CAS. He organized the Twelfth Five-year Plan of Informatization Development of CAS. He was engaged in space science and exploration project and project management in long term. He was also a member of expert panel of deep space exploration project of the National “863” Program. E-mail: fyliao@cnic.cn