



利用科研信息化手段 推动微生物研究与应用*

文 / 马俊才¹ 刘 斌² 吴林寰¹ 杜晓萌¹ 孙清岚¹ 王 楠¹

1 中国科学院微生物研究所 北京 100101

2 中国科学院科技促进发展局 北京 100864

【摘要】 信息技术对科学研究活动产生了巨大的影响。中科院微生物所开展了一系列科研信息化的有益尝试。承担了世界微生物数据中心(WDCM)的建设工作,开发了微生物菌种信息挖掘与服务平台,并实施了包括工业生物技术知识环境建设、应用微生物研究网络信息中心建设等在内的一系列科研信息化项目,力求将微生物学研究与应用学科应用紧密结合起来,以学术研究服务产业应用。文章通过具体案例的实践探索,针对如何利用科研信息化手段推动微生物学研究和应用,提出了若干意见。

【关键词】 科研信息化,世界微生物数据中心,微生物菌种信息挖掘与服务平台,工业生物技术知识环境建设,应用微生物研究网络

DOI 10.3969/j.issn.1000-3045.2013.04.013

现代网络信息技术使当今社会发生了革命性的改变,对科学研究活动也产生了巨大的影响。海量科研数据的产生,科研管理手段的更新,全球科技合作的诉求,地理国界樊篱的突破,这些都意味着科学研究急需飞速发展的信息化手段的支撑。而作为已在“第四次科技革命”中大放异彩的生物学的重要分支,微生物学研究亦融入科技发展的奔涌浪潮中,搭乘信息化的方舟,激荡出一朵朵耀眼的科技浪花。

1 微生物资源信息化现状与挑战

微生物学是现代分子生物学发展的三大支柱之一,其在整个生物学发展中起到了不可估量的作用。而微生物数据资源是微生物资源共享和开发的关键环节,数据资源的丰富性、准确性和共享水平决定着整个微生物学领域研究和应用的综合能力。与实物资源相比,微生物信息资源是最有可能实现共享的一种资源。通过信息技术,建立统一的数据标准,为微生物资源研究的各个环节提供包括数据管理及共享、数据分析、计算模型等在内的支撑,促进信息资源的共享

* 修改稿收到日期:2013年4月17日



中国科学院

从而带动对微生物资源的开发和利用,对微生物资源研究和生物技术发展具有重要的意义。

世界各国科研人员对信息资源的共享需求,促进了生物学数据库的大发展。目前,世界上最著名的核酸一级数据库有3个,即由美国国立生物技术信息中心(NCBI)建立和维护的GenBank数据库、由欧洲生物信息学研究所(EBI)维护的EMBL核酸序列数据库和日本的DDBJ数据库^[1]。而伴随着人类基因组计划(HGP)的实施,产生了为此计划保存和处理基因组图谱数据的基因组数据库(GDB)。蛋白质数据库主要包括:PSD(PIR国际蛋白质序列数据库)、SWISS-PROT(蛋白质序列注释数据库)、PROSITE(蛋白质位点和序列模式数据库)、PDB(蛋白质空间结构数据库)、SLOP(蛋白质结构分类数据库)和COG(蛋白质直系同源簇数据库)等。相关功能数据库主要包含:KEGG(京都基因和基因组百科全书)、DIP(相互作用的蛋白质数据库)、ASDB(可变剪接数据库)、TRRD(转录调控区数据库)、TRANSFAC(转录因子及其结合位点数据库)^[2]。自2010年起,核酸序列数据开始爆发式的增长,每年数据产出达到数百PB级,耗资数百亿美元以上,推动了众多重要的科学发现。截至2011年,全球至少有超过1300个在线生物信息学数据库。截至2011年3月,已经完成测序的细菌基因组达到876个,还有648个基因组完成了测序草图^[3]。

微生物数据资源的开发和共享极大地推动了科学的进步,但其在发展过程中也不可避免地面临诸多问题和挑战。由于资源和环境分布的局限性,微生物学研究也不可避免地受到区域限制,全球性的合作势在必行。然而,在全球,还没有一个比较统一的项目或者领导性的机构,来协调各方资源,促进交流合作和资源的共享。21世纪初,经济合作与发展组织(OECD)曾经推动过全球微生物资源中心网络(GBRCN),欧盟也推动过欧洲微生物资源及信息共享项目(CABRI),但是这些计划由于缺乏共享机制和技术力量支持等原因,都

未能建立一个稳定运行的成熟的国际性数据平台;此外,各国对微生物资源开发和信息化应用的进程不一,也是阻碍世界微生物学整体发展的一个重要因素。

具体到我国的微生物数据资源领域,我们与发达国家的差距主要体现在两个方面:一是在我国的微生物学领域中,缺乏统一的、近乎覆盖全国的信息平台和知识环境,用来整合调动全国各主要微生物学科研机构,服务我国国内的微生物学研究与应用;二是世界知名数据库均为欧美发达国家所开发,这就使我国的微生物资源信息化在国际学术界处于从属地位。因此,开发具有国际影响力的微生物学数据库和信息平台十分必要。在此基础上,还应进一步主导建立全球性权威微生物学数据库,提升我国在国际微生物学界的地位。

2 中科院微生物所的科研信息化实践

认识到科研信息化的重要性及其在微生物学的巨大潜在服务能力,中科院微生物所结合自身的研究方向,开展了一系列科研信息化的有益尝试。承担了世界微生物数据中心(WDCM)的建设工作,在WDCM的框架下,开发了微生物菌种信息挖掘与服务平台,并在中科院生命科学与生物技术局、信息化工作领导小组办公室的支持下,与上海生命科学院、天津工业生物技术所、青岛生物能源与过程所、成都文献情报中心、对地观测与数字地球科学中心等单位合作,实施了包括工业生物技术知识环境建设、应用微生物研究网络信息中心建设等在内的一系列科研信息化项目,力求将微生物学研究与应用学科应用紧密结合起来,以学术研究服务产业应用。本文将选取若干案例加以介绍,为业界同仁以及对科研信息化应用感兴趣的读者提供借鉴。

2.1 世界微生物数据中心(WDCM)的建设

随着计算机技术的飞跃发展和互联网的渗透普及,国内外形成了一系列微生物分类数据库。用户可以据菌种的部分特征搜索目标微生物,并

能通过网络获得相关微生物的生化指标等在线资料,这些都为微生物研究者提供了巨大的信息资源,信息的快速收集极大地扩展了微生物研究人员的研究范围以及研究能力。同时,高度发达的网际互联为微生物学的信息交流、资源共享和国际合作带来了前所未有的机遇。

世界微生物数据中心(WDCM)隶属世界菌种保藏联合会(WFCC)和联合国教科文组织(UNESCO)的微生物资源中心网络MIRCEN (UNESCO Microbial Resources Centers Network),是微生物数据资源领域规模最大、成员最多、最具国际影响力的科学组织之一,也是科研信息化的践行者和推动者。目前WDCM收集了全世界68个国家584个保藏中心的微生物资源信息。中心在国际生物多样性公约的框架下促进微生物资源的共享利用;同时借鉴国际先进的微生物资源管理经验,竭力提升全球微生物资源保藏工作的信息化管理水平和数据质量,以期使发展中国家微生物资源的研究和开发利用更好地融入国际研究计划,建立良好的沟通平台,形成与世界对话的能力。

建设WDCM是微生物所促进全球科研信息化的成功案例。就我国而言,它丰富了中国微生物信息资源,为我国微生物技术创新和新品种创造提供物质基础,增强了我国微生物产品的研发能力,并建立了良好的沟通平台,使中国微生物资源的研究工作更好地融入国际研究计划,提升中国微生物学领域的国际影响力;就全球范围来看,世界微生物数据中心提供了一个有效整合全球微生物资源数据的机制和平台。这一平台帮助实现了对全球微生物及其遗传资源的盘点,在国际生物多样性公约旨在保护资源输出国利益的前提下,促进微生物资源的共享利用。

2.2 微生物菌种信息挖掘与服务平台建设

微生物资源是人类赖以生存和发展的重要物质基础和生物技术创新的重要源泉,是目前生命科学领域研究和产生数据最丰富、对数据分析最深入、方法研究最成熟的学科。为了很好地应对微生物研究过程中对数据的需求,我们构建了微生物菌种信息挖掘与服务平台。该平台包括全球微生物资源目录(Global Catalogue of Microorganisms, GCM)和生物资源引用分析系统(Analyzer of Bio-resource Citations, ABC)。

GCM是为全球保藏中心和微生物学家服务的整合型全球微生物数据资源平台。平台将开拓与菌种资源相关的基因组、蛋白质、结构、功能、应用等多方面异构大数据。形成自主开发、只针对微生物菌种资源的综合结构化数据库。同时嵌套多种数据分析工具,为菌种资源提供信息查询、数据分析、功能预测、开发利用等多方面的支持与服务。目前,已有来自全球19个国家36个保藏中心的超过23万余株微生物资源数据加入了该平台。预计未来,该计划将覆盖全球超过100个保藏中心的数据资源。这对提升我国在全球微生物资源数据共享及利用的引领地位具有重要意义。

ABC是利用搜索引擎、文本挖掘等信息技术,对生物资源类文献进行信息采集,数据整理和引用关系挖掘的Web平台。在此过程中建立起关于生物领域资源类文献的元数据信息库及全文库。并在此基础上,挖掘出生物领域资源类文献对生物资源的引用信息。ABC系统集成了微生物在科学文献、专利、基因组、核酸序列等方面的数据挖掘信息,提供给微生物保藏机构、科研人员、保藏微生物使用者全面的微生物资源引用信息。

全球微生物菌种保藏目录和生物资源引用分析系统的建设和成功实施,对于建立



中国科学院

统一的全球微生物菌种目录体系,对主要的保藏中心的目录的标准化整理,提供统一的检索出口;深度挖掘了微生物资源引用情况和生物信息学数据分析,同时规范化了全球各注册微生物菌种保藏中的记录提交字段;更加方便了微生物菌种资源的交流与利用;提升了我国在国际微生物保藏领域的引领作用。对世界微生物资源保藏组织、微生物资源领域专家学者以及微生物资源的研究使用者提供一个知识全面、追踪更新、方便易用的查询分析平台。

2.3 工业生物技术知识环境建设

我国十分重视工业生物技术的发展,在《国家中长期科学与技术规划》中将其列为重点研究领域。中科院也从国家战略规划的高度考量,特别在其“10+1”基地建设规划中,设置了“先进工业生物技术基地”。

为了促进基地项目之间的资源共享和合作研究,构建紧密融入用户科研过程的个性化知识环境,微生物所启动了“工业生物技术知识环境”建设工作。项目从调查研究、战略分析、知识库和知识环境建设及基地信息化管理等方面入手,结合工业生物技术基地中的项目管理、人才队伍建设、院地、院企合作等工作,整合和集成工业生物技术研究必须的生物实物资源信息、专利信息、文献情报、产品、企业、政策等信息,建立工业生物技术基地综合知识仓库、智能检索引擎和信息化协同工作环境;紧密跟踪国际工业生物技术发展趋势,针对若干重大应用需求,定期开展最新发展态势和核心关键技术信息的定制信息服务,形成中科院工业生物技术发展战略报告;结合基地的人力资源管理,建立骨干科研人员个人门户,管理组织的显性知识,并且积累、挖掘组织的隐性知识,以“中国工业生物技术信息网”为门户,促进生物技术成果的信息共享、成果转化和院地、院企合作;建立我国生物能源科学技术评价体系,为国家或相关部门特别是中科院先进工业生物技术创新基地提供生物能源科学技术发展和产业化的对策与建

议。“工业生物技术的知识环境”项目是中科院首次尝试为基地建设配置专门的信息化支撑项目,既是基地管理模式创新,也是利用信息化手段服务科研活动的有益探索。

通过知识环境建设,在工业生物技术基地各研究单元之间初步形成了一张知识网。在知识网的基础上,还结合应用微生物研究网络建设以及所际联合的网格化研究中心建设,形成面向基地的资源、研究和知识三网合一的网络管理构架。网络管理构架模式的提出和实践,增强了基地对科研布局和科研组织的调控能力,有效弥补了原有的院所垂直管理关系的局限性,增加了网络之间节点的联系,促进了资源的逐步整合共享、优化配置。从而有利于凝聚全院力量,在联合争取和承担重大项目、开展多学科合作方面迅速反应,系统应对。打破原有机制体制和管理组织形式的瓶颈问题,对基地管理工作和科研活动本身都产生了积极的推动作用。将资源网络、研究网络、知识网络结合,实现三网合一多层网络架构的管理模式是基地实现高效科研管理的有益尝试。

在“10+1”基地中率先开始知识环境建设的实践,并通过中科院科研信息化示范项目的后续支持,基地实现了信息化工作与科研、管理方式创新的有机结合,为基地管理机制的创新提供了一个范例。

2.4 应用微生物研究网络信息中心建设

微生物资源是地球上最大的、尚未充分利用的生物资源,是生物技术创新和生物产业发展的源头和基础。我国微生物资源丰富,已有相当规模的菌种储备,微生物应用研究也有一定基础,但在微生物资源的规范管理、开放共享和功能评价等方面尚存在诸多问题,因而制约了微生物资源的开发利用。为推动我国微生物资源的收集、整合、研究和开发,中科院于2010年8月正式启动了中科院应用微生物研究网络,旨在更好地发挥微生物资源和技术在支撑生物产业发展和满足国家重大需求方面的作用,推动国民经济和社

会的可持续发展。

应用微生物研究网络总体研究框架由总中心、分中心和网络实验室组成。总中心设有网络信息中心,专门进行信息化建设相关工作。信息中心通过为应用微生物网络建立科研信息化的工作环境,集成了超级计算、知识仓库、文献情报等信息化基础设施,以先进的信息化手段提升微生物资源数据保藏和微生物资源开发能力,促进跨地区、跨领域的协同工作和资源共享;以世界先进的微生物菌种资源管理和数据标准为基础,建立全国性应用微生物研究的数据网络体系,建立全面的数据管理系统和数据汇交平台;并且与我国从事微生物学研究的一线科学家紧密结合,深入挖掘数据应用需求,为我国应用微生物领域研究提供完善的数据管理系统和先进的数据挖掘工具,从而切实推动微生物资源的开发利用。

应用微生物研究网络是中科院在微生物资源研究领域的重要布局,也是一种新的组织和管理方式的探索与尝试。

在网络的整体层面,通过建立虚拟网络实验室,加强整体层面的联系和对项目宏观进度的把握,有效地促进了资源的整合和系统应对;在研究中心层面,以资源优势进行布局,充分发挥研究中心在资源的系统收集与整理、共享标准与机制、信息化建设、交流与合作等方面的关键作用。通过科研信息化的手段,对现有各领域的信息资源进行了充分的整合和数据挖掘,提供面向工业、农业、环境和海洋等应用方向的数据应用和服务的典型。

在网络实验室和具体的研究层面,中心为配合以需求为导向,深入挖掘资源应用潜力的设计初衷,建设了生物信息学、微生物元基因组、蛋白质组学分析、合成生物学模块设计等各种数据分析和应用的平台,充分

利用计算资源为科学研究提供支撑。

应用微生物研究网络的建立,整合了全院应用微生物研究力量,发挥了相关研究所不同研究领域的优势和特长,完善了从资源保藏到资源开发和利用的产业链,前瞻性地抢占了海洋微生物和环境微生物研究的制高点,对加强应用微生物领域的知识积累、技术创新和转化,整体提升该领域的创新能力,更好地发挥微生物在支撑生物产业发展和解决国家需求方面的作用起到了重要作用。而与之相适应的科研信息化环境的建设,是应用微生物研究网络得以高效、有序运行的一个不可或缺的保障。

3 结论和建议

信息化技术在科研中的应用其实并不拘泥于某一种模式。我们可以调动各种信息化手段,根据实际需要,量体裁衣,设计出最为适合的技术方案,辅佐微生物学的研究,进而促进科研成果的应用。建议从以下几方面着手:

(1)加强针对微生物资源数据的保存、利用共享的机制及平台的研究。在构建数据库和平台过程中,建立规范化、国际化微生物资源数据模型,通过促进数据资源的共享,加强微生物实物资源的交流,促进我国、乃至世界微生物学整体学科研究的发展。

(2)建立系统的知识环境,开发协同工作平台。对各种数据、计算等资源进行整合,将数据转变为知识,并将传统科学数据库、ARP系统和超算设备等信息化手段与文献情报服务相结合,深化战略情报研究,服务科研应用,推进产学研一体化。借助知识环境,为科学家提供个性化的知识服务,提高工作效率,促进科研活动模式的转变。

(3)在从事研究的过程中,还应注重创新科研管理模式。利用信息化技术,辅以有



中国科学院

力的行政管理手段,双管齐下,优化管理模式,实现数据的定期汇交和整合,确保数据库建设和项目实行的质量。

(4)积极推动国际行动计划,逐步构建我国在国际微生物学界的主导地位。应积极倡导重要领域性国际合作行动计划,通过参与、主导制定全球微生物资源数据的整合机制和信息平台标准,吸引全球各国微生物资源保藏机构加入,建立国家和区域性节点,进一步确立我国在国际微生物资源信息共享领域的主导地位,并以此为契机逐步

建立获取、利用全球微生物资源的渠道,在微生物资源的利用方面开展与其他国家的实质性合作,推动我国及全球生物产业的快速发展。

参考文献

- 1 Baxevanis A D. Methods Biochem Anal, 1998, 39: 1-15.
- 2 Borsani G, Ballbio A, Banfi S. Hum Mol Genet, 1998, 7: 1641-1648.
- 3 Genome Online Database: <http://www.genomesonline.org/cgi-bin/GOLD/index.cgi>

Promote Microbiological Research and Application by Virtue of e-Science

Ma Juncal¹ Liu Bin² Wu Linhuan¹ Du Xiaomeng¹ Sun Qinglan¹ Wang Nan¹

(1 Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

2 Bureau of Life Sciences and Biotechnology, Chinese Academy of Sciences, Beijing 100864, China)

Abstract Information technology has greatly influenced scientific activities. Being aware of the significance of e-Science and its potentials in microbiological study, Institute of Microbiology, Chinese Academy of Sciences (IMCAS) implements a series of e-Science programs, including World Data Centre for Microorganisms (WDCM), Service Platform for Information Mining of Microbial Strains, Knowledge Environment of Industrial Biotechnology, and Research Network for Applied Microbiology (RNAM). The institute strives to integrate microbiological research and its application. This paper will discuss how to promote microbiological research and application by virtue of e-Science.

Keywords e-Science, World Data Centre for Microorganisms (WDCM), service platform for information mining of microbial strains, knowledge environment of industrial biotechnology, Research Network for Applied Microbiology (RNAM)

马俊才 中科院微生物所网络信息中心主任,正高级工程师。2006年获日本三重大学生物资源系博士。兼任世界菌种保藏联合会(WFCC)世界微生物数据中心(WDCM)主任、世界菌种保藏联合会执委、国际科学数据委员会CODATA中国国家委员会委员、中科院科学数据库专家委员会副主任委员、中科院信息化专家委员会委员、中科院典型培养物委员会理事。主要从事生物网格、并行检索技术、超大规模全文检索技术、远程异构数据库的检索引擎技术、IT技术在生物学领域的综合利用等方面的研究。E-mail: ma@im.ac.cn