

“工业生物技术知识环境” 建设及应用

马俊才^{1*} 吴林寰¹ 刘斌² 邢雪荣² 于建荣³ 邓勇⁴ 杜晓萌¹ 徐萍³ 夏芸³ 陈方⁴

1. 中国科学院微生物研究所, 北京市 100101
2. 中国科学院生命科学与生物技术局, 北京市 100864
3. 中国科学院上海生命科学信息中心, 上海市 200031
4. 中国科学院成都文献情报中心, 四川省成都市 610041

摘要

鉴于工业生物技术的重要地位, 中国科学院设立了“先进工业生物技术创新基地”。为了使现代信息化手段更好地服务于基地建设, 我们启动了“工业生物技术战略研究和知识环境建设”。知识环境的信息化成果主要体现在管理平台、虚拟实验室、网络化办公、战略研究等方面。该知识环境的建设是一种全新的信息化工作和服务模式的探索, 将信息化平台工作建设与实际的科学研究活动紧密结合, 对于加强信息技术在科研中的应用, 提高科学研究活动的效率等方面有着积极的意义。

关键词: 工业生物技术; 信息化; 知识环境



The Development and Application of Knowledge Environment of Industrial Biotechnology

Ma Juncai^{*1}, Wu Linhuan¹, Liu Bin², Xing Xuerong², Yu Jianrong³, Deng Yong, Du Xiaomeng¹, Xu Ping³, Xia Yun³, Chen Fang⁴

1. Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
2. Bureau of Life Science and Biotechnology, Chinese Academy of Sciences, Beijing 100101, 100864
3. Shanghai Information Center for Life Sciences, Chinese Academy of Sciences, Shanghai 200031, China
4. Chengdu Library of Chinese Academy of Sciences, Chengdu 610041, China

Abstract: In view of the importance of industrial biotechnology, Chinese Academy of Sciences establishes the Advanced Industrial Biotechnology Innovation Base. In order to facilitate its work with modern informationization achievement, we launch the project of Strategic Study and Knowledge Environment Development of Industrial Biotechnology. The project has made great progress in management platform, virtual lab, networking and strategic study. As a brand-new informational means and service pattern, the knowledge environment delicately integrates the informationization platform and scientific research, which can enhance the application information technology and improve the efficiency of research work.

Keywords: Industrial biotechnology; Informationization; Knowledge environment

1. 引言

工业生物技术是继医药生物技术、农业生物技术之后,国际生物技术发展的“第三次浪潮”。它以微生物或酶为催化剂进行物质转化,与现代工程技术有机结合,大规模生产人类所需的化学品、医药、能源、材料等,是解决人类目前面临的资源、能源及环境危机的有效手段,发展前景十分广阔。作为新兴的绿色化工产业,工业生物技术的产业化领域主要包括生物能源、生物基产品、绿色工艺过程、生物炼制、代谢组学与代谢工程、生物催化与生物转化、发酵与下游过程工程、特殊环境微生物资源与能源植物资源等。世界经合组织(OECD)指出:“工业生物技术是工业可持续发展最有希望的技术”^[1]。

世界上的主要发达国家和地区都针对工业生物技术领域制定了重大规划与项目。例如,美国众议院于2007年12月18日通过了修改后的新能源法案,大力鼓励生物燃料乙醇的发展;欧盟先后实施了“欧洲能源政策”、“欧洲能源技术战略计划”、“欧盟战略能源技术计划”、BIOCUP项目等一系列政策及项目规划,均涉及生物质能源的应用和推广;巴西则是推动生物燃料产业发展的先锋,实施了提高汽油中的乙醇比例、提高柴油中的生物柴油比例、推广生物燃料“生态证书”、完善对生物多样性和遗传

资源的保护和管理等一系列政策法规^[2]。

我国也十分重视工业生物技术的发展,制定了中长期科技发展规划,明确提出要发展工业生物技术,并把生物技术作为未来高技术产业迎头赶上的重点,加强生物技术在农业、工业、人口与健康等领域应用的总体部署,并在国家中长期科学与技术规划中将其列为重点研究领域^[3]。中科院也意识到工业生物技术具有重要的战略地位,认为及时了解国际先进工业生物技术发展动态信息、开展工业生物技术战略研究、分析各国的战略规划、总结和预测工业生物技术产品和市场,对我国工业生物技术的发展是非常重要的。有鉴于此,中国科学院特别在院创新工程三期中提出,将“先进工业生物技术创新基地”列为科技发展的“1+10”基地之一。

“先进工业生物技术创新基地”的定位是一个知识型组织,其管理理念大力提倡知识的整合、创新和集成,认为良好的战略决策支持和信息化的管理平台对于工业生物技术研究以及相关的产业化发展起着举足轻重的作用。为此我们提出,要利用信息化管理平台,将基地的科研进展、专家队伍、管理经验等重要的无形资产转化为可管理、可共享、可利用的信息网络资源,为研发中心、技术平台、重点实验室和工程中心的建设和重大项目的实施提供支持。

为贯彻这一理念,我们开始进行“工业生物技术战略研究和知识环境建设”。知识环境的建设思路是运用各种信息化手段,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;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;建立项目进展日志,实现基地的信息化管理,并与ARP系统实现联合,将科研活动和科研管理有机结合在一起;以“中国工业生物技术信息网”为门户,加强与院地合作局的合作,促进我院生物技术成果的信息共享、成果转化和院地、院企合作。

“先进工业生物技术创新基地”是中科院创新工程“10+1”基地中,第一个开始知识环境建设的基地,在基地的信息化应用方面是一个新的突破。

2. 建设思路及技术路线

知识环境建设集成了国际与国内工业生物技术领域较为全面的、多类型的信息资源,应用相应的数据挖掘工具,揭示了工业生物技术领域的知识概念之间的内在联系,建立了工业生物技术知识分类体系,实现了对各种信息资源进行组织与管理,从而构建了规范化、数字化的工业生物

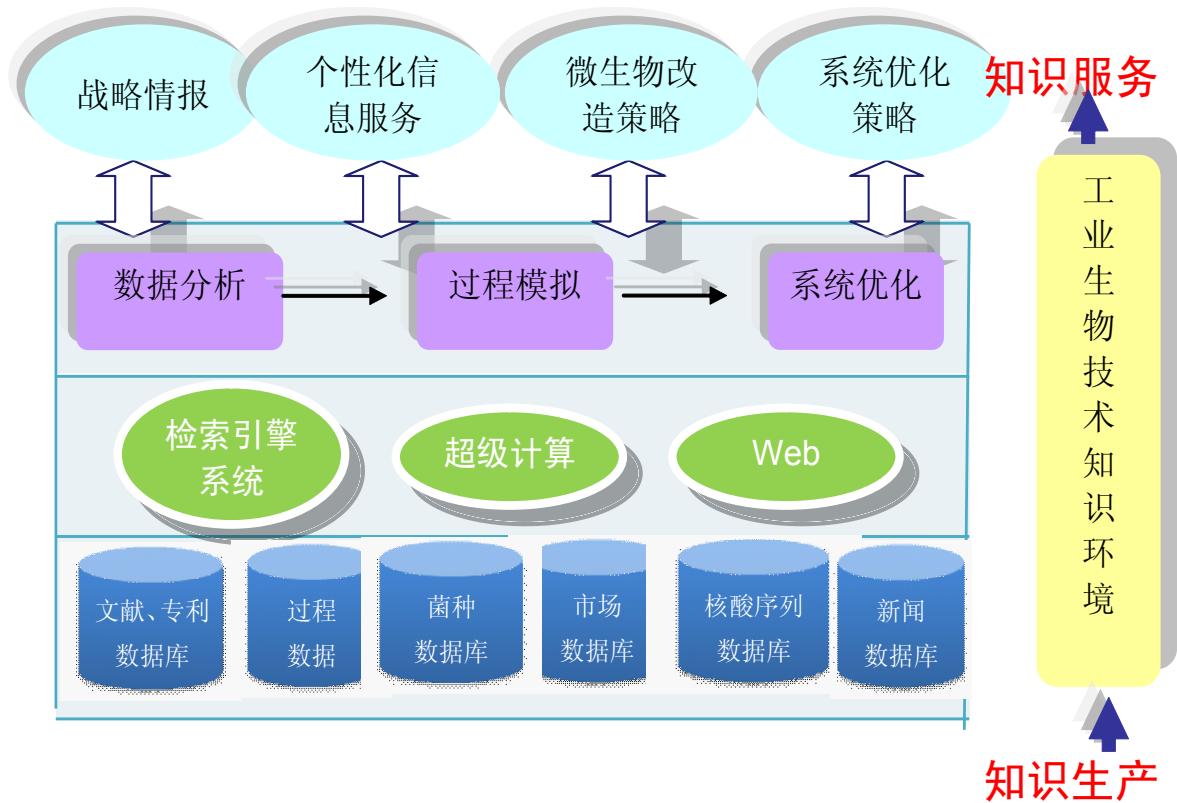


图1 知识环境建设思路

► 技术战略研究与集成知识环境，是将来进行数据挖掘工作、生成工业生物技术发展战略报告的数据基础。

知识环境的建设思路是集成包括文献、专利数据库、工业生物技术过程工程数据库、菌种资源数据库、市场数据库、核酸序列数据库和新闻数据库等在内的各种数据资源，利用信息化手段和IT技术，如检索引擎技术、超级计算平台和网络技术，实现一个从数据分析到过程模拟到系统优化的工业生物技术系统过程，该过程的结果会产出一系列的产品，也就是知识环境的信息产品，包括战略情报的分析报告、面对专家的个性化信息服务、针

对微生物的改造策略和系统优化策略。从而实现从知识积累到知识生产并进行知识服务的过程。

3. 信息化成果

知识环境建设开展两年多来，通过参与建设的中科院微生物所、上海生命科学信息中心、中科院成都文献情报中心、中科院青岛生物能源与过程研究所等单位的通力合作，在信息化建设方面已经取得了一系列成果，在院内外取得了极大的反响。知识环境的研究成果创新性地以一系列的信息产品的形式进行组织，各个信息产品的定位有所区分和侧重，覆盖到我国工业生物技

术的领导、专家和科研人员对战略研究、知识管理和信息服务的不同需求。为基地学科部署、领域方向调整、重大项目的立项、重要规划、路线图的制定提供战略决策支持，还为院相关各局、科技部、发改委和基金委相关领导、专家提供知识服务。

信息化成果主要体现在管理平台、虚拟实验室、网络化办公、战略研究等方面。

1) 数据库整合

在平台建设过程中，数据提取、转换和加载(Extraction Transformation and Loading, 下简称ETL)起着至关重要的作用。工业生物技术领域的信息来源包括生物实物资源信息、生物物种

数据、知识产权信息、生物信息学数据等多种，其数据源既有关系型数据库，也有来自文件系统的数据(如文本文件，csv文件，Excel文件等)，还有web数据源。ETL过程从多个数据源提取业务数据，清理数据，然后集成这些数据，并将它们整合进数据库中，为数据分析做好准备^[4]。

在数据库整合方面，我们深入研究了PubMed，ISI Web of Knowledge等web文献数据库的结构特征，在此基础上，针对工业生物技术领域相关的文献、专利等情报数据特点，分别设计了文献、专利数据库的数据模型；针对不同数据源（ISI Web of Science，Pubmed等）的数据格式，设计开发了相应的ETL工具，实现了异构数据源的数据向Mysql数据仓库的ETL过程。目前，已经集成了生物能源、生物材料、生物催化与生物转化、发酵与下游工程技术和生物资源等领域，包括聚乳酸、脱落酸、生物丁醇、生物柴油、沼气等主题在内的十个主题的文献数据库和聚乳酸、脱落酸、生物丁醇、生物柴油、沼气专利等在内的9个专利数据库。并进一步在此整合的数据仓库的基础上，采用Apache + PHP + Mysql的web开发工具，采用基于MVC的ThinkPHP框架，开发了基于web的工业生物技术专题数据库的检索查询系统，初步实现了基于web的信息浏览、检索、统计功能。

2) 信息平台

先进工业生物技术基地涉及到中国科学院院内的18个研究所，建设南北两个研究中心，在8个重要方向部署60-70个重要方向项目，建设5个关键技术平台，涉及资源研究、组学研究、分子操作、过程优化、应用开发5大研发团队。新的组织管理模式，对基地的管理工作提出了严峻的挑战和更高的要求。

为此，我们设计了一套信息化的管理平台，这是知识环境的重要性和创新性的体现。知识环境建设了一系列与基地的项目和管理相关的数据库，记录了基地所管理的平台和项目在经费投入、人员组成、建设成果等各方面详细的数据。这些数据库使基地的管理工作井然有序。例如，结合工业生物技术基地建设，通过信息化手段推动基地综合管理，建立基地建设进展数据库，以全程跟踪和记录工业生物技术基地的建设步伐和发展轨迹；结合工业生物技术基地的队伍建设和人力资源管理，建立基地科技专家数据库；结合工业生物技术基地的重大项目和重要方向项目的实施，建立项目管理数据库，服务于基地整体层面的信息采集和决策管理。依托这些数据库，我们可以进行综合统计，进一步挖掘隐性知识，形成基地管理的知识积累。

3) 虚拟实验室

先进工业生物技术基地涉及的实验室不是独立的个体，必须加强交流，密切合作。为了给科

学家们提供一个更好的信息交流和科研合作平台，我们建立了基于网格（GRID）和WEB SERVICE环境的工业生物技术网上虚拟实验室。这是e-Science概念在工业生物技术领域的重要应用。通过虚拟实验室，科学家既可以便捷、有效地与其他地区或其他领域的研究人员及时地讨论研究进展、科研成果等情况，同时也可以利用网络发布业内信息、提交战略需求，进而形成一个跨地区、跨领域、满足工业生物技术基地这一分布式和网格特殊组织形式要求的信息化协同工作环境。将网格化的协同工作环境与具体的科学研究相结合，从而形成覆盖整个基地甚至我国工业生物技术领域的网络交流平台，是知识环境一个创新性的尝试。

4) 网络化办公

我们充分地发挥了网络覆盖面广、使用方便的优势，在日常工作中实现了较为全面的网络化办公。除了上面提到的信息化管理，网络化在我们日常工作中的应用还体现在诸多方面：首先，我们充分调动一线科研人员的积极性，参与知识环境的建设，及时提供信息、知识，使他们不仅是知识环境的用户，而且是重要的建设者；其次，我们利用现代化IT技术，对“中国工业生物技术信息网”等门户网站进行维护，实时发布信息。

5) 战略研究服务

由于工业生物技术所涉及的学科范畴较为复杂，包括基础

►生物学、生物技术、化工技术、工程技术等领域，体现出较强的学科交叉的特点，导致决策制定者和科学家花费了大量的时间和精力却难以得到准确的情报、文献、专利等资源，对相关领域的产业发展和动态缺乏全面和系统的把握。因此迫切需要建设一个非常有效的、可利用的资源和信息服务平台，为政策制定和科研活动提供支持。

在“工业生物技术战略研究与知识环境建设”过程中，我们充分考虑到该领域的特殊性，借助网络信息技术手段搜集国内外工业生物技术领域较为全面的、有代表意义的、多类型的动态信息资源，采用情报研究领域的知识管理技术，开发相应的数据挖掘工具，揭示工业生物技术领域的知识概念之间的内在联系，建立工业生物技术知识分类体系，实现对各种信息资源进行组织与管理，提供系统的战略决策情报、学科情报与竞争情报，从而构建规范化、数字化的工业生物技术战略研究与集成知识环境。

我们通过及时搜集整理和分析工业生物技术领域动态消息，深入研究国家经济社会发展相关战略需求和本领域国际科技发展前沿与趋势，取得了一定的成果。例如，通过对国内外工业生物技术领域著名公司和研究机构的发展状况的调研，结合专家的建议，我们推出了《生物柴油》、《生物丁醇》、《聚乳酸》、《纤维素乙醇》等一系列

工业生物技术发展态势的专题报告；我们还为“先进工业生物技术基地”在“纤维素乙醇的高温发酵和生物炼制”等重大项目的立项和“生物质资源科技领域”发展路线图制定提供了可靠的战略与信息服务。

这一知识环境对无论是对相关领域的决策者、科研人员，还是对企业界人士，都具有重要的意义。借助工业生物技术领域的战略情报研究，相关管理者和决策者可以了解各国制订战略规划的背景、内容，各种计划的实施方略、组织方式，以及计划的实施情况，总结各国的战略规划的成功经验，使之在制定工业生物技术的发展计划过程中有所借鉴；借助战略情报研究，科研人员可以了解、追踪国内外在工业生物技术领域的最新发展动向，准确把握国际、国内战略规划与研究计划、政策，合理调整研究方向，使之适应国家战略规划 and 产业化的实际要求；借助战略情报研究，企业界人士能够分析和预测工业生物技术领域的基础研究方向和产品开发潜力，明确市场定位。战略研究服务兼具专业咨询、决策依据与信息导向的作用，通过个性化的信息定制服务，将信息与资源及时和准确地传递给最需要的人。因此，加强工业生物技术领域的战略情报、企业与产品的竞争情报以及学科情报的研究，可以为国家的战略决策、科学家的研究活动、企业的生产提供科学的情报咨询与决

策依据，具有重要的理论意义和现实意义。

4. 意义及前景

工业生物技术战略研究与知识环境建设是一种全新的信息化工作和服务模式的探索，将信息化平台工作建设与实际的科学研究活动紧密结合，对于加强信息技术在科研中的应用，提高科学研究活动的效率等方面有着积极的意义。

开展工业生物技术战略研究和知识环境建设的目的在于为了促进“先进工业生物技术创新基地”的科技创新活动，为决策部门、科研管理者以及科学家提供快捷、及时的工业生物技术信息和决策报告。到目前为止，知识环境已经为中国科学院先进工业生物技术基地所涉及到的数十个研究所、近百位科研人员和科研管理工作提供服务，它为管理与咨询决策提供参考依据，为科研与开发人员提供全面及时的研究方向评述和市场需求分析以及深层次、高层面的战略研究与知识决策环境。

知识环境的建设已经初具规模，也摸索出了一套适合的发展模式。要真正地使知识环境有效地运转起来，就必须与科研领域的专家和科研人员相结合。因此，我们从建设之初，就一直是以从科研活动的实际需求出发为原则进行的，并取得了比较好的效果。在建设工业生物技术知识

仓库时，对科学家的需求进行了实际的调研，根据科研活动的情况，进行数据库的设计，并且通过虚拟实验室和网格环境，可以将数据积累与科研活动有机地联系起来并且在科学研究中应用。在进行战略情报研究的过程中，也是通过网络环境的有效组织，将专家的智慧通过网络环境聚集并且通过信息化的手段转变为可以直接利用的知识。同时，我们

也为专家提供了可以订制的个性化信息服务，科研人员可以通过网络选择他们所关注的新闻、文献、专利、会议等各种信息进行订制，知识环境会将所订制的信息直接发送给科研人员，避免了信息搜集的复杂性和冗余信息的干扰。

我们的目标是要将知识环境的概念和信息管理模式在我国工业生物技术甚至整个生物技术

领域进行推广。我们还将广泛开展与国内外工业生物技术领域企业的合作，将知识环境的服务对象从科研人员拓宽到工业生物技术领域相关企业，为企业提供相关信息咨询和决策支持服务，增强科研人员与企业的有效交流，促进科技成果向实际生产力的转化。为我国的工业生物技术的科研和生产发展提供技术和智力支持。



参考文献:



[1] Organization for Economic Cooperation and Development. Biotcchnology for Clean Industrial Products and Processes: Towards Industrial Sustainability. Paris: OECD report, 1998.

[2] 江洪波, 等. 工业生物技术领域重大规划与项目. 2008 工业生物技术发展报告. 北京: 科学出版社, 2008: 2-29.

[3] 曹竹安. 抓住大好时机 发展工业生物技术. 生物技术产业, 2007, 2.

[4] 苏庆. 工业生物技术数据仓库ETL系统的设计与实现. 中国科学院微生物研究所硕士论文, 2008.

收稿日期: 2009年4月16日

作者信息



马俊才

中国科学院微生物研究所信息网络中心, 主任、正高级工程师, 研究方向为生物信息学、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、LINUX机群系统、IT技术在生物学领域的综合利用等。