

DOI: 10.13376/j.cblls/2023167

文章编号: 1004-0374(2023)12-1543-02

· 专辑: 生物医学大数据 ·



赵国屏, 分子微生物学家, 中国科学院院士, 中国医学科学院学部委员, 发展中国家科学院院士, 美国微生物科学院院士。现任中国科学院上海营养与健康研究所生物医学大数据中心首席科学家, 国科大杭州高等研究院首席教授, 复旦大学生命科学学院微生物学和免疫学系主任, 山东大学公共卫生学院健康医疗大数据研究院/国家健康医疗大数据研究院首席科学家; 兼任中国生物工程学会合成生物学专业委员会名誉主任, 上海生物工程学会名誉理事长。研究工作涉及微生物生理生化、基因组学、系统与合成生物学以及生物信息学等领域。曾参与启动中国人类基因组计划及相关生命“组学”研究, 克隆若干遗传病致病基因; 主持若干重要微生物的基因组、功能基因组、比较和进化基因组研究, 解析SARS冠状病毒分子进化机制。在细菌蛋白质乙酰化组和肠道微生物组等领域作出若干开创性工作。组建并领导中国科学院合成生物学重点实验室, 在酵母染色体重构、代谢组与代谢流量组研究、天然化合物细胞工厂制造、基因编辑技术研发等方向上, 实现重要突破。近年来, 参与组建并领导中国科学院上海生命科学研究院(现营养与健康研究所)生物医学大数据中心, 为建立国家生物医学大数据治理体系开展预研工作。

序 言

赵国屏

(中国科学院上海营养与健康研究所, 生物医学大数据中心, 上海 200031)

生物医学大数据不仅是生物医学领域开展数据密集型研究的基石, 成为与人口健康、社会发展和国家安全相关的战略资源, 而且还是利用人工智能赋能“大健康”产业发展的核心生产要素之一。如何把海量的生物医学数据转换成能发挥研究基石和产业要素作用的“生物医学大数据”, 即实现其“价值”, 是长期以来生物医学领域的相关人员、机构乃至政府部门都深感“机遇与挑战”的大问题。但是, 该问题基本上未被全面、系统、深入地加以讨论并探索出解决途径。究其原因, 可能是源于该问题的核心, 既与海量生物医学数据与生俱来的“跨尺度、多源性、高维度、细粒度、异质性复杂体系”的“低价值密度”的数据特征有关, 又与数据的产生者、收集者、使用者之间在“管——安全与私密, 用——投入/产出, 放——竞争/合作”等一系列“复杂利益关系”的社会特征有关。2016年, 当时的中国科学院上海生命科学研究院, 依托其所属马普计算

生物研究所, 秉承“面向生物医学科技前沿、面向国家生物安全需求、面向生物医药产业主战场、面向人民生命健康”的初心, 以“先投入战斗, 然后再见分晓”的只争朝夕精神, 建立“生物医学大数据中心”, 在整整7年的时间里, 努力探索解决这一问题的途径, 获得了一些体会和经验。他们特别认识到对于生物学、健康医学等数据而言, 标准化规范整合与开放共享是使数据“要素化”的重要环节; 而且只有秉承“安全管理、信息共享、标准增值、技术创新、尊重产权、高效利用”的理念, 将生物、医学、工程、技术等多学科交叉融合并在工程技术层次上紧密协作, 才能形成落实这一环节的数据治理体系, 源源不断地供给高质量的生物医学大数据。

鉴于生物医学大数据这几年实践中形成了多方面的合作伙伴, 本刊特委托他们组织“生物医学大数据”专辑, 邀请我国在该领域长期从事数

据基础性工作的研究团队从多个视角向读者介绍大型数据中心的建设经验及相关数据技术的进展，回顾健康与疾病人群队列研究历史及趋势，总结微生物组大数据研究动态，以期通过梳理生

命科学、环境科学、公共卫生等领域的大数据进展，为生物医学大数据科学相关的工程、技术与研究提供新思路，促进我国高质量的数据供给与应用。