
NIH 为大数据计划投资 3200 万美元以确保掌握其资助项目的主要数据

2014 年 10 月 9 日，NIH 宣布将发展新策略以应对迅速扩增的复杂生物学数据集，即 BD2K 计划（Big Data to Knowledge）。BD2K 在 2014 财年首次获资将近 3200 万美元，总投资计划于 2020 年达到 6.56 亿。

随着 DNA 测序和成像等生物医学研究技术的迅猛发展，相应数据的产生正在超越研究者的处理能力。BD2K 计划旨在支持新方法、软件、工具及培训项目的发展以寻求加速数据的处理与再发掘。该计划希望在隐私得以保护的前提下，能够为挖掘巨大数据集探索新方法，最终应用于改善人类健康。如，预测乳腺癌、心脏病等疾病的高危个体，及其治疗和预防的更佳方法。

BD2K 将创建 12 个研究中心，分别主攻某一特定科学数据问题。同时还将支持培育有关数字发现索引（data discovery index）、数据科学训练以及相关人力资源开发等项目。

从分子层面到群体层面，在生物学维度内对数据进行消化吸收将是生物医学研究的未来趋势。包括我们每一个人在内的健康信息也将是大数据问题。确保能掌握其资助项目的主要研究数据是 NIH 优先关注的问题。

不可小觑的是，生物信息数据利用面临诸多挑战。包括：数据定位问题；适宜的数据获取和分析软件的研发；不同类型数据缺乏统一格式，不同研究群体间数据标准难以统一；保护隐私的同时简化数据共享的政策制定。

BD2K 计划的 4 个主要组成部分：

- 大数据计算卓越中心（Centers of Excellence for Big Data Computing）：发展创新性方法、软件、工具和其它资源。该中心预期成果与大数据科学的各个方面广泛相关，如数据整合与利用、基因组数据分析以及电子健康记录数据管理等。
- BD2K-LINCS 数据协调与整合中心（BD2K-LINCS Perturbation Data Coordination and Integration Center）：该中心将是基于网络的细胞标志综合数据库（Integrated Network-based Cellular Signatures, LINCS）的数据协调中心，支持其数据研发项目。
- BD2K 数据发现索引整合联盟（BD2K Data Discovery Index Coordination Consortium, DDICC）：该中心旨在从整体角度对生物学数据进行挖掘分析和索引归类。
- 培训与人力资源开发（Training and Workforce Development）：用于培训大数据处理分析

所需要的专业技术人员。

原文标题: NIH invests almost \$32 million to increase utility of biomedical research data

原文链接: <http://www.nih.gov/news/health/oct2014/od-09.htm>

检索日期: 2014-10-15

中科院心理所信息中心