

Scientific Database in a Dilemma about Becoming Public Instrument

—The Rise and Fall of the BIND from a Contemporary History's Perspective

Ang Li, Lie Sun

Institute for the History of Natural Sciences, CAS, Beijing
Email: li_ang@ihns.ac.cn, sunlie@ihns.ac.cn

Received: Apr. 27th, 2016; accepted: May 6th, 2016; published: May 13th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The influence of scientific databases to research activities is becoming more and more significant. The rise and decline of a database could be transient, but vividly reflect the interactions between society and the booming development of science and technology. The non-profit Biomolecular Interaction Network Database (BIND) was once very prosperous but faded just in a few years. Its open access concept finally ended with an unsuccessful commercialization. In this process, BIND tried to establish a new paradigm in the scholarly world. Meanwhile, it was also a result of the affection of emerging technologies like Open Science data. Establishing and running a massive scientific database is not a pure technique activity within the scholarly community. The attitudes of governments and enterprises determined the fate of BIND to a large degree. Considering scientific database as a “public instrument” leads to a dilemma. The apparently different attitudes towards scientific database among various scientists, governments, and enterprises, in fact, reveal the social cost of the transition from scientific knowledge to public knowledge in the contemporary era.

Keywords

Scientific Database, BIND, New Paradigm, Public Knowledge

科学数据库的“公器”困境

—当代史视角下的BIND兴衰

李昂, 孙烈

中国科学院自然科学史研究所, 北京
Email: li_ang@ihns.ac.cn, sunlie@ihns.ac.cn

收稿日期: 2016年4月27日; 录用日期: 2016年5月6日; 发布日期: 2016年5月13日

摘要

科学数据库对学术研究的影响日益显著, 其兴衰虽稍纵即逝, 却是当代迅猛发展的科学技术与社会相互作用的生动写照。非营利性的生物分子相互作用数据库——BIND曾兴盛一时, 却在数年后迅速衰落, 免费获取的渠道最终走向并不成功的商业化之路。在此过程中, BIND既是建立学术新范式的一次尝试, 也是受新兴技术如开放式科学数据库影响的结果。建立与运行大型科技类数据库并非只是学术共同体内部的一项纯粹的技术活动, 政府与企业对科学数据库的态度很大程度上决定了BIND的命运。科学数据库的“公器”困境, 表面上是不同科学家与政府、企业对科学数据库的不同态度, 实质上反映了当代科学知识转为公共知识的社会成本。

关键词

科学数据库, BIND, 新范式, 公共知识

1. 引言

当代史在某种程度上是社会受科学和技术进步冲击的结果[1]。使科学知识成为“公器”即公众文化的一部分, 这是一种自18世纪末以来的信念, 它变革了以往对科学知识的看法并推动了社会的进展[2]。二百余年之后, 电子信息技术的出现发展极大地影响了知识的发现、保存与传播的方式。特别是自上世纪80年代以来, 全世界诞生了数以万计的科学数据库, 其“公器”的属性如何得以体现?

如果说科学是一项公共事业, 并且具有显然的公益特性, 那么科学数据库亦然。随着信息技术与各学科的结合, 数十年来全世界各种类型的科学数据库几乎遍及各个知识门类, 层出不穷。自1980年代起, 研究人员已发现这些数据库能够“直接导致新的发现”[3], “对于科学研究越来越重要”[4], 它们既是已有科学研究结果或中间过程的信息载体, 也是进一步研究的基础和新知识的孵化器。设法使科学数据库成为“公共知识”, 这是不少此类数据库创立者的理想。然而, 由于建立与运营的形式不同, 科学数据库并不天然就是开放式数据(Open Science data, OSD)。因其拥有者和管理模式的不同, 其开放程度和发展状况差别很大。据不完全统计, 在全球学术类的数据库(含相关搜索引擎)中, 采用完全开放获取(Free)的约占43%, 例如著名的arXiv, Bioinformatic Harvester等, 而其余多是会员注册收费的商业运作形式[5], Biological Abstracts, IEEE Xplore 和 Science Citation Index 等均在其列。

生命科学也好, 其他学科也罢, 数据库是否应该成为公共知识的来源, 已经在学界引起了纷争[6]。为理解此现象, 本文追踪了一个生物学数据库 BIND 变迁史。BIND (the Biomolecular Interaction Network Database)是收集管理生物分子相互作用信息的一个大型数据库, 由加拿大多伦多大学 C.W. Hogue 教授领导的研究团队开发, 受加拿大与新加坡政府资助, 于2001年对外公布, 并成功与1982年开始建立的核酸序列数据库 Genbank 数据库对接, 面向研究者完全开放获取。不过2005年之后两国政府都决定不再提供经费支持, 因此该数据库从那儿之后就停止了更新。政府资助的 BIND 一度是学术“公器”, 其改弦易辙背后的原因耐人寻味。

2. BIND 的兴衰

BIND 数据库的正式开发主要基于多伦多大学的 Christopher W.V. Hogue 指导其学生 Gary D. Bader 于 2002 年完成的博士学位的研究。论文题目就是《BIND 的开发与使用：储存与分析蛋白间相互作用数据》[7]。事实上，BIND 的最初设想是由 Tony Pawson 博士在 Bader 入学之前就提出来的，他是 Hogue 的同事，也是 Bader 的博士导师之一。2000 年 Bader 和 Hogue 首次发表了关于 BIND 的文章，介绍其开发背景、数据库结构、使用方法以及它与其他数据库交互引用的情况等等[8]。次年，BIND 正式上线，最初的数据库包括了 1000 个相互作用的记录、6 条生化代谢途径和 40 个分子复合物。它支持用户上传信息，但上载的信息要经过管理员整理才能正式进入数据库[9]，由此较好地兼顾了扩展性与数据质量。

2002 年底 BIND 成为当时最大的蛋白质间作用(PPI)的数据库，容量已经扩大到 6186 对相互作用，851 个复合物和 8 条代谢途径。其 PPI 被分成 3 大类：二元分子相互作用(binary interaction)、分子复合物(molecular complex)以及生物途径(biological pathway)，它们分别从不同层面呈现了分子间的相互作用关系。收录的 1500 种生物分子之间的 200,000 种相互作用的数据，包括蛋白质之间的相互作用，以及蛋白质与 DNA、RNA、小分子、脂质以及糖类物质之间的相互作用。

自 BIND 创立之初，Hogue 等人就将其定位为公共数据库，希望维持以政府资助来维持其非营利性发展。该数据库发展最迅速的时期，也正是获得政府资助最顺利的阶段。2002 年 BIND 获得了来自加拿大政府的关注与官方的资助，此外还有公益基金的支持，总额 2900 万加元。同时，著名服务器厂商 Sun 公司还提供 500 万加元的软硬件开销。这些投资效果十分显著，至 2003 年底，BIND 中记录的相互作用已经达到 42,000 余条，并有望在 3 年内添至 8 万条记录。

BIND 的政府资金来源在此后几度出现问题，但仍坚守公共数据库的初衷。2003 年加拿大政府承诺的经费并没有全部到位，尤其是安大略省应该出的将近一千万加元被无限延期[10]。Hogue 转而向与他有密切合作的新加坡方面需求解决。2004 年他以新加坡国立大学为依托与新加坡政府的经济发展局(Economic Development Board of Singapore, EDB)合作成立了 Blueprint Initiative Asia Pte.Ltd。它虽然叫做公司，但仍然是非营利的[11]。然而新加坡承诺的总额 1800 万美元为期 5 年的经费投入，因为种种原因没有全部兑现。

经费不足最终将 BIND 逼出了公共数据库的行列。2005 年 11 月 16 日，Hogue 在 Blueprint 的网站上发表了一封公开信，宣告由于经费无以为继，已经终止了对 BIND 的更新，并对以往支持表示感谢云云。2006 年，汤森 - 路透科技公司成为新东家[12]。与以往不同，这次变更使 BIND 最终走向了商业化道路。虽然 Hogue 后来在自己的主页中称 BIND 是头一个将加拿大政府投资的知识产权向加拿大企业转移的成功例子，但从他以往坚持让 BIND 保持为公共数据库的立场看，也许他本人并不愿意看到这次转移。到目前为止，BIND 中的部分信息仍可从 Unleashed 的网站免费获取，不过已经无法下载整套数据了。

3. “公器”的由来：建立学术新范式的尝试

BIND 等生物学数据库之所以出现，主要是生命科学的学术范式发展到一定阶段的历史产物。可从以下三个方面来说明。

3.1. 顺应了学术共同体的新需求

近数十年来，分子生物学实验室的基本研究范式逐渐成型，即在数据库中已有的信息进行分析的基础上，设计实验，进而完成对问题的判断。20 世纪 70 年代，DNA 测序技术以及转基因技术出现之后，分子生物学呈现出爆发式的增长。到了 80 年代，PCR (polymerase chain reaction)技术的发明又给它注入了一针新的催化剂，从此几乎所有生命科学领域的前沿课题都集中在分子层面了，每天由全世界的生物

实验室产生出数以千万碱基对计的核酸序列, 从中可能翻译出的蛋白质也有成百上千。如此海量数据的产生, 导致了生物信息学这样的分支学科产生, 它侧重利用计算机对生物数据进行分析。1991年在英国的《自然》杂志上刊登了一篇 W.Gilbert 撰写的短文, 指出在生物学研究领域正在兴起新的范式, 因为所有的“基因”都将可以从数据库里读取, 所以今后生物学研究的起点将是理论猜测开始, 然后才转向实验去验证假设。

随着越来越多的生物基因组序列被测定, 人们开始更多的关注基因功能的研究。生物分子要实现其功能, 首先要与其他分子相互作用, 因此对于感兴趣的分子, 证明它会与其他哪些分子有相互作用, 是研究其功能的首要步骤。鉴于生物分子的多样性, 以及实验结果的复杂性, 一个研究者如果想在某个生物分子的功能上进行一些开拓性的研究, 往往需要阅读上百篇文献才能获取有关的基本信息。这种情况下, 如果有一个数据库汇总各种生物分子相结合的信息, 将会使研究的准备阶段效率大为提高。随着越来越多的蛋白质组学研究成果的发表, 人们对于方便的存储和管理这些信息的需求也就越强烈, 也正是出于这种需求, BIND 数据库应运而生。

3.2. 受到“开放式科学数据”理念的影响

随着科学数据库的数据量与数据复杂性的增加, 科学数据库的管理视为“最重要的新兴需求之一”[13]。“开放式的科学数据(Open Science Data, OSD)”能够提供一种高效、廉价获取科研数据的方式。OSD 的思想萌芽出现于 1950 年代。随着互联网的发展, 数据作为公共资源的呼声不断, 1995 年 OSD 的构想得到国际科学理事会(International Council for Science, ICSU)的支持。此后, 开放数据的获取得到越来越多的学者与公众的认可。不仅一些影响大的机构和学科领域, 如 NASA、地学、化学等开展了相关行动, 生物学领域著名的“人类基因组计划”也是积极响应者。BIND 的管理理念显然承袭了 OSD 的模式, 在 Bader 的博士论文可找到相关表述。2005 年 XML 大会明确可扩展标记语言(XML)对 OSD 的支持。BIND 很快就找爱荷华州立大学为 BIND 做了相应的技术改造, 使得数据符合 XML 规范, 并提供了 XQuery 查询机制, 提高了强有力的数据存储和分析能力, 数据索引、排序、查找、相关一致性等[14], 而且还符合美国(ASN)与万维网联盟(W3C)制定的相关国际标准, 例如 ASN.1 与 XML DTD [9]。数据库升级后, BIND 的管理者与团队仍然坚持原有的理念——为全世界的研究者提供“免费的”信息与数据分析工具[15]。

3.3. 以非营利性作为应对同质化竞争的一种策略

在本世纪到来之前, 分子生物学的大部分工作集中在研究生物分子的结构, 对核酸而言, 主要指其序列; 对蛋白质来说, 还包括折叠而成的高级结构。世界各国建立了大大小小的许多数据库管理这些信息。这些数据库的内容虽有差异, 但彼此也有重叠。为了形成特色, Hogue 团队赋予了 BIND 若干重要的特征, 以确保其在同行之间的优势。从 BIND 的内容和数据形成方式上看, 它有如下特点:

1) 内容全面。BIND 囊括了生物分子间, 以及生物分子与非生物分子间的各种相互作用。收集分子间相互作用信息的数据库, BIND 并不是第一个, 比如加州大学洛杉矶分校在 1999 年就构建了一个关于蛋白质相互作用的数据库(Database of Interacting Proteins, 简称 DIP), 根据最新的统计其中已经收集了 57,683 条蛋白质相互作用的记录[16]。更早些时候建立的有专门收集大肠杆菌代谢途径的 EcoCyc 数据库, 它是由斯坦福研究所(Stanford Research Institute, 简称 SRI)维护的, 现在已经扩展到了 506 种生物。到了 2000 年前后, 同类数据库大量涌现, 有些专门收集某些分子的相互作用, 有些专门收集某种模式生物的相关数据, 它们大多只是某个大学或研究所下属的项目, 因此数据类型各有侧重, 并且规模都比较小。在 2003 年《核酸研究》杂志对所有生物学数据库所做的统计中, 列举了 9 个关于分子间相互作用的比较有代表性的数据库, BIND 是其中唯一全面收集各种分子相互作用信息的, 并且这个记录至今仍未被打破。

2) 数据质量高。BIND 的数据大部分由人工生成, 因此质量高, 可靠性强。在设计之初, BIND 就着重考虑了数据质量的问题。数据主要有两个来源: 经过同行评议的文献和全球重要的生物信息机构的发表物。这些来源的数据库须经过人工处理后才入库。因为一个 BIND 记录的生成需要由两个管理员配对完成, 其中一个先阅读文献、从中提取信息、录入数据库, 另外一个则要对录入的信息进行审读, 校验。一般来说完成这样一个过程需要两天, 但考虑到时差的关系, 如果由分别处于北美和亚洲的两个管理员配对合作, 就能在 24 小时以内完成。虽然在设计上也允许用户上传数据, 但校验的过程实际上还是需要管理员亲自阅读相关文献, 这样才能保证数据生成标准的一致性。管理员们平均每个月要审读 2000 篇以上的最新研究文献[17], 从中提取分子间相互作用的信息, 并按照 BIND 的格式要求进行录入和注释。相比之下, GenBank 的数据生成方式基本是由使用者上传, 经过数据库管理者简单的校验后就可上网共享。

3) 与其他数据库的良好关联性。一个数据库与其他数据库的关联性高低, 将极大的影响这个数据库的价值。由于曾经在 NCBI 工作过, Hogue 从一开始就使用了 NCBI 提供的软件编写工具来开发 BIND, 并且 BIND 一上线就与 Genbank 等 NCBI 所属的数据库实现了交互检索。不久之后, BIND 又与包括 Nature, Science 在内的 77 种刊登分子生物学研究文章的杂志建立了合作关系, 这些杂志会把相关文章在发表之前交给 BIND 管理员们阅读, 让他们从中提取出分子间相互作用的信息, 输入数据库, 并赋予数据库编号(BINDid), 这样当文章发表时就会附上 BINDid 以便读者参考。这种深层次的合作是 BIND 的创新, 并且至今同类数据库大多都无法做到的, 因而也为 BIND 增添了不少附加值。同时对于某些信息, 如果其他数据库中已经做过相应的记录, BIND 也会给出链接, 以便用户追溯, 而这种连接通常都是单方向的。由此也带来一个好处, BIND 与其他同类数据库相比, 效率可能是最高的。比如 DIP 从 1999 年到 2009 年的 10 年间只产生了 5 万多条记录, 而 BIND 从 2001 年到 2005 年就产生了将近 20 万条记录(远远多于最初设想的 8 万条)。

4) 注重生物学家的研究需求。BIND 设计者要求制作数据时, 全面考虑生物学家们如何才能方便快捷的从中获取所需的信息, 并且始终保持着数据生成标准的一致性, 以及与其他数据库之间良好的兼容性。BIND 的界面非常友好, 在 Drug Discovery 杂志做的一次调查中, 受访的科学家承认, BIND 使用起来很容易; 在信息日益增加的情况下, 想从中挖掘出自己研究中需要的数据越来越困难, 而 BIND 对他们的研究非常有帮助[18]。相比之下, 半途出家的生物信息学家开发的数据库往往缺乏对使用者的关怀, 在《核酸研究》做的一项调查中发现, 有一半的生化学家对基因组数据库中的工具根本就不知道是干什么用的[19]。

可见, 做高品质且非营利性的“公共数据库”是 BIND 团队的信念。首席科学家 Hogue 一直致力于把 BIND 打造成可以为全世界生命科学领域的研究, 尤其对药物研发, 疾病防控等领域的研究提供重要参考资源的公共平台。事实也确如此, 在 2004 年 Nature Cell Biology 的编辑栏目, 有专文介绍 BIND 并予以了很高的评价[20]。当 BIND 的知识产权出让给 Unleashed 公司之后, 该公司网站的注册人数迅速攀升了 10 倍, 这也可以侧面说明 BIND 在业界的受欢迎程度。2004 至 2005 年间, BIND 升级了数据库, 建立了与多个世界顶尖杂志的合作关系, 并与诺华公司的热带病研究所(Novartis Institute for Tropical Diseases 简称 NITD)成立了一个登革热的合作研究项目。即便在 2005 年加拿大完全停止了对 BIND 的资助, 该项目仍持续了一段时间。

4. 困境的由来: 开放模式折戟经费难关

为了保证数据质量, BIND 的数据主要靠人工生成, 而维系这一高成本运营, 则需要持续的高投入。经费需求最终演变为开放模式与政府资助的对接困难, 由此直接导致了 BIND 的衰落。

4.1. 政府政策的不连续性殃及财政资助

BIND 投入几乎都源自加拿大政府。在 BIND 总共接受的 1780 万加元(当时约折合 1200 万美元)经费中, 三分之二强来自加政府管理的非营利性组织 Genome Canada, 剩下的接近三分之一来自安大略省研究发展基金(Ontario R&D Challenge Fund, 简称 ORDCF), 另外还有来自加拿大自然科学与工程研究理事会(National Science and Engineering Research Council of Canada, NSERC), 加拿大健康研究所(Canadian Institutes of Health Research, CIHR)等单位的小额资助。

加拿大是一个两党制的国家, 其保守党对基础科学的保守态度十分鲜明, 2008 年, 执政的保守党政府解雇了四年前由前总理保罗·马丁(Paul Martin)任命的国家科学顾问阿瑟·卡蒂(Arthur Carty), 并关闭了他在政府的办公室, 可以说是个代表性的事件。在自由党执政期间, Hogue 相当轻松的获得了所需的经费。当 2005 年政府面临换届之际, 不但新经费的申请失败了, 而且已经承诺过的也没有到位。很明显这其实是国家政策的变化[21]。Hogue 把这些归结为他“无法控制的一系列不幸的环境变化”[11]。

其次, BIND 开展国际合作遭遇失败。维护一个大型的、服务全球科研人员的公共数据库, 也许不是应该由一个国家承担的责任。BIND 在发展过程中曾经设计了一个很美好的国际合作框架, 涉及美、欧、亚洲地区的三个国家。事实上从 2004 年下半年到 2005 年上半年这段时间, Blueprint 北美和亚洲的两个分部同时运转时, 曾经建立了非常高效的合作关系。不过继加拿大政府之后, 新加坡政府也取消了原本答应的长期投资。新加坡对于独力维护一个知识产权不属于她的数据库顾虑重重。

4.2. 开展国际合作遭遇失败

维护一个大型的、服务全球科研人员的公共数据库, 也许不是应该由一个国家承担的责任。BIND 在发展过程中曾经设计了一个很美好的国际合作框架, 涉及美、欧、亚洲地区的三个国家。事实上从 2004 年下半年到 2005 年上半年这段时间, Blueprint 北美和亚洲的两个分部同时运转时, 曾经建立了非常高效的合作关系。不过继加拿大政府之后, 新加坡政府也取消了原本答应的长期投资。新加坡对于独力维护一个知识产权不属于她的数据库顾虑重重。

4.3. 公共数据库难获企业资助的青睐

BIND 为了保持其作为公共数据库的地位, 也没有更多地寻求来自企业的支持。这虽然保证了数据库的独立管理与运营, 但资金始终捉襟见肘。企业对非营利性组织的赞助, 因为缺乏经济利益的驱动, 所以很大程度上只是偶然事件, 很少持久下去。事实上, 遇到以上困难的科学数据库不止 BIND 一家。早在 1999 年, 就有调查发现, 三分之二的生物数据库都面临经费来源不稳定的问题[22]。容易获得资助的多是创新型的项目, 对于正在进行当中的、需要长期维持的基础项目, 并不怎么感兴趣[23]。商业化的数据库也许可以在细节上做得更好, 但服务范围总归有限。而“倾斜性收费”带来的收益又未必能满足维持数据库发展的所有需要。

面对理想与生存的抉择, BIND 不得不放弃作非营利性的“公共数据库”的理想, 商业化最终还是成为它的宿命。

5. 结论

科学技术与社会的互动是当代非常突出的现象, 同样也当代史非常重要的研究对象。BIND 兴衰史虽然短促, 却凸显了科学知识作为“公器”在当代所面临的困境。

BIND 之兴表明, 数据不同于以往的“散乱的”知识, 科学数据库具有不可替代的重要性。信息化时代与基因组时代的到来, 意味着数据库的运用能导致新知识的发现[24], 成为知识与智慧的来源[25], 而

数据密集型科学发现(Data-Intensive Scientific Discovery)甚至有跃升为当代学术新范式的趋势。BIND 的高质量与非营利性受到学术共同体的青睐, 是因为它在建立学术新范式过程中的成功尝试。

BIND 之衰, 源自科学数据库在当今社会的非“公器”属性。换言之, 科学知识之“公器”并不天然地导致科学数据库具有相同的属性。正如同“知识无价, 而书本有价”的俗理。科学数据库面临许多重大的技术难题, 比如大数据设施、交互式数据探索、可视化的数据管理与使用等, 其应该具有何种特性, 仍难有定论。一般认为它应该不那么昂贵, 且使用方便[26]。可是这些需求之间并不容易协调, 比如数据库拥有良好交互特性、数据质量高、安全可靠的数据库, 往往需要大量资金与人力的投入, 而产出的不确定性又降低了公共资金的兴趣。Hogue 在宣告 BIND 项目终止的公开信的末尾无奈的提了这样的问题: “数据库对今后生命科学领域的研究是必需的。问题是它们应该免费供给公众使用还是应该收取注册费呢, 像 BIND 这样的数据库是否应该保持非营利性并完全开放呢?” BIND 的最初设计是希望生物分子相互作用的研究数据能成为学术“公器”, 但苦于缺乏持续的公共资金来维系。

经费在很大程度上压制了学术共同体的理想信念, 也决定了 BIND 的兴衰。然而, 问题并未就此终结。科学数据库该成为公共的知识来源, 还是应走商业化之路, 这之间始终存在一种难以偏废的张力。1998 年 BTSF 所做的一项研究表明, 数据库作为一个经济体类似于学术期刊, 它们对整个科学共同体来说必不可少[22]。若从功能、管理和传播等方面来看, 科学数据库很可能将演化为学术期刊的一种类型[27]。期刊的历史经验或能为科学数据转为公共知识提供某种思路。科学数据库之困, 表面上科学家与政府、企业对公共数据库所持的不同立场, 实质上反映了当代科学知识转为公共知识仍面临着巨大的社会成本。

参考文献 (References)

- [1] Barraclough, G. and Kellett, N. (1964) *An Introduction to Contemporary History*. Watts, London, 42-46.
- [2] Golinski, J. (1999) *Science as Public Culture: Chemistry and Enlightenment in Britain, 1760-1820*. Cambridge University Press, New York, 1-10.
- [3] Gertz, M. (2010) Scientific and Statistical Database Management. *22nd International Conference, SSDBM 2010*, Heidelberg, Germany, 30 June-2 July 2010. Springer Science & Business Media, Berlin, 4-5.
<http://dx.doi.org/10.1007/978-3-642-13818-8>
- [4] Chia, L. and Idema, W.L. (2007) *Books in Numbers: 75th Anniversary of the Harvard-Yenching Library*. Harvard-Yenching Library, Harvard University, Cambridge, MA, Distributed by the Chinese University Press, Hong Kong, 420.
- [5] Wikipedia (2016) List of Academic Databases and Search Engines.
https://en.wikipedia.org/w/index.php?title=List_of_academic_databases_and_search_engines&oldid=706265886
- [6] Harvey, M. and McMeekin, A. (2009) *Public or Private Economies of Knowledge?: Turbulence in the Biological Sciences*. Edward Elgar Publishing, Cheltenham, UK, Northampton, MA, 1-36.
- [7] Bader, G.D. (2003) *Design and Use of the Biomolecular Interaction Network Database (BIND) for Storing and Analyzing Protein-Protein Interaction Data*. Doctoral Dissertation, University of Toronto, Toronto.
- [8] Bader, G.D. and Hogue, C.W.V. (2000) BIND—A Data Specification for Storing and Describing Biomolecular Interactions, Molecular Complexes and Pathways. *Bioinformatics*, **16**, 465-477.
<http://dx.doi.org/10.1093/bioinformatics/16.5.465>
- [9] Bader, G.D., Betel, D. and Hogue, C.W. (2003) BIND: The Biomolecular Interaction Network Database. *Nucleic Acids Research*, **31**, 248-250. <http://dx.doi.org/10.1093/nar/gkg056>
- [10] Kondro, W. (2005) Canadian Global Database May Move to Singapore after Loss of Funding. *Science*, **308**, 1529-1529. <http://dx.doi.org/10.1126/science.308.5728.1529a>
- [11] Hogue, C. (2005) Canadian Database in Singapore. *Science*, **309**, 875-877.
<http://dx.doi.org/10.1126/science.309.5736.875b>
- [12] Wang, K. (2007) The Other Side of Staying out of a BIND. *Nature Biotechnology*, **25**, 971-972.
<http://dx.doi.org/10.1038/nbt0907-971b>
- [13] Ailamaki, A. and Bowers, S. (Eds.) (2012) Scientific and Statistical Database Management. *Proceedings of 24th International Conference, SSDBM 2012*, Chania, 25-27 June 2012, Vol. 7338, Springer, 1-2.
<http://dx.doi.org/10.1007/978-3-642-31235-9>

- [14] Lin, Y. (2003) XML Schema of Biomolecular Interaction Network Database and an XQuery Use Case. M.S. Thesis, Iowa State University, Ames.
- [15] Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., Buzadzija, K., *et al.* (2005) The Biomolecular Interaction Network Database and Related Tools 2005 Update. *Nucleic Acids Research*, **33**, D418-D424. <http://dx.doi.org/10.1093/nar/gki051>
- [16] UCLA (2016) Database of Interacting Proteins. <http://dip.doe-mbi.ucla.edu/dip/Stat.cgi>
- [17] SRI International Is an Independent, Nonprofit Corporation. LIST OF BIOCYC PATHWAY/GENOME DATABASES. <http://biocyc.org/biocyc-pgdb-list.shtml>
- [18] Johnston, N. (2004) The BINDs That Tie. *Drug Discovery*, **9**, 59-60.
- [19] Baxevanis, A.D. (2003) The Molecular Biology Database Collection: 2003 Update. *Nucleic Acids Research*, **31**, 1-12. <http://dx.doi.org/10.1093/nar/gkg120>
- [20] (2004) Editorial. *Nature Cell Biology*, **6**, 681. <http://dx.doi.org/10.1038/ncb0804-681>
- [21] Walsh, M.P. (2005) Co-Funding versus Scientific Excellence. *The Biochemist*, **10**, 45-46.
- [22] Ellis, L.B. and Kalumbi, D. (1999) Financing a Future for Public Biological Data. *Bioinformatics*, **15**, 717-722. <http://dx.doi.org/10.1093/bioinformatics/15.9.717>
- [23] Merali, Z. and Giles, J. (2005) Databases in Peril. *Nature*, **435**, 1010-1011. <http://dx.doi.org/10.1038/4351010a>
- [24] Piatetski, G. and Frawley, W. (1991) Knowledge Discovery in Databases. MIT Press, Cambridge, MA.
- [25] Ackoff, R.L. (1989) From Data to Wisdom. *Journal of Applied Systems Analysis*, **16**, 3-9.
- [26] French, J.C., Jones, A.K. and Pfaltz, J.L. (Eds.) (1990) Scientific Database Management: Report of the Invitational NSF Workshop on Scientific Database Management, Charlottesville, Va., March 1990: Final Report (Vol. 90, No. 21). University of Virginia, Charlottesville.
- [27] Bourne, P. (2005) Will a Biological Database Be Different from a Biological Journal? *PLoS Computational Biology*, **1**, e34. <http://dx.doi.org/10.1371/journal.pcbi.0010034>