

Reinventing Discovery: The New Era of Networked Science

重塑发现： 网络化科学的新时代

澳大利亚 迈克尔·尼尔森 (Michael Nielsen) 著

祁澍文 石雨晴 译



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

重塑发现：

网络化科学的新时代

[澳大利亚] 迈克尔·尼尔森 (Michael Nielsen) 著

祁澍文 石雨晴 译

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

REINVENTING DISCOVERY: HOW ONLINE TOOLS ARE TRANSFORMING SCIENCE

By MICHAEL NIELSEN

Copyright: ©2011 BY MICHAEL NIELSEN

This edition arranged with Louisa Pritchard Associates and The Science Factory through BIG APPLE AGENCY, INC., LABUAN, MALAYSIA.

Simplified Chinese edition copyright: 20XX W.E Time DigiTech Ltd.

All rights reserved

本书中文简体版授权予电子工业出版社独家出版发行。未经书面许可，不得以任何方式抄袭、复制或节录本书中的任何内容。

版权贸易合同登记号 图字：01-2014-7111

图书在版编目(CIP)数据

重塑发现：网络化科学的新时代 / (澳) 尼尔森(Nielsen, M.) 著；祁澍文，石雨晴译. -- 北京：电子工业出版社，2015.2

书名原文：Reinventing discovery: the new era of networked science

ISBN 978-7-121-25385-0

I. ①重… II. ①尼… ②祁… ③石… III. ①互联网络—普及读物 IV. ①TP393-49

中国版本图书馆CIP数据核字(2015)第003732号



书 名：重塑发现：网络化科学的新时代

作 者：[澳大利亚] 迈克尔·尼尔森(Michael Nielsen) 著 祁澍文 石雨晴 译

策划编辑：胡 南

责任编辑：郑志宁

印 刷：三河市双峰印刷装订有限公司

装 订：三河市双峰印刷装订有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编100036

开 本：720×1000 1/16 印张：18 字数：250千字

版 次：2015年2月第1版

印 次：2015年2月第1次印刷

定 价：55.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

作者简介：

迈克尔·尼尔森 (Michael Nielsen)，澳大利亚量子计算科学家。研究领域包括量子计算、开放科学、人工神经网络。曾供职于Los Alamos National Laboratory，2004年被授予澳大利亚“Youngest Academic”称号，并获昆士兰大学联邦基金。

目 录

01 重塑发现 001

第1部分 放大集体智慧

02 在线工具令我们更加聪明 019

03 重新分配专家注意力 029

04 在线协作模式 059

05 集体智慧的局限与潜力 093

第2部分 网络化的科学

06 世间全部知识 121

07 科学的民主化 169

08 开放科学事业所面临的挑战 225

09 开放科学的当务之急 245

附录：博学者项目所解决的那道数学难题 273

01

重塑发现

蒂姆·高尔斯 (Tim Gowers) 不是人们司空见惯的那种博主。他是剑桥大学的一名数学家，还是菲尔兹奖得主，该奖是数学领域的最高荣誉，通常被称为数学界的诺贝尔奖。所以他的博客都是在分享数学观点和与数学有关的深入见解。

2009年1月，高尔斯决定有步骤地进行一场与众不同的社会实验，工具就是自己的博客。他挑选了一道意义重大的数学难题，称这道题目是他自己非常“想要解开”的。不过，他既没有自己独自攻关，也没有邀请几名亲近的同事来合力攻克此题，他决定要将解题之事彻底公开化，所谓公开就是在自己的博客上贴出相关的想法和部分的进展。他还发表了一份公开邀请函，邀请众人来帮助解题。

任何人都可以关注此事的进展，如果有主意，也可以在博客的评论区阐述自己的想法。高尔斯希望众人同心协力能比一个人冥思苦想的力量更强大，他希望能够有不同专长和不同视角的人参与其中，以彼此启发，最终用集体的力量将这道数学难题化难为简。他将这个实验命名为“博学者项目”。

博学者项目开始了，但起步缓慢。在高尔斯开放博客，供众人探讨此数学难题的头7个小时中，评论区无人光顾。不过，7小时后，第一则评论出现了，一位来自英属哥伦比亚大学的数学家约瑟夫·绍伊莫希（Jozsef Solymosi）就此题目提出了一个变型，这个变型没有高尔斯的那个复杂，但绍伊莫希认为这个变量可能更有助于解题。15分钟后，亚利桑那州的一名中学教师贾森·戴尔（Jason Dyer）也参与了进来，给出了自己的想法。仅仅3分钟后，同为菲尔兹奖得主，来自加州大学洛杉矶分校（UCLA）的数学家陶哲轩（Terence Tao）也发表了一则评论。此后，评论数量如火山喷发一般快速增长：接下来的37天内，27人参与，共发表800则数学评论，总字数超过17万。纵览这些评论，你能看到不同观点以同样惊人的速度被提出、被提炼、被抛弃。你也能看到顶级数学家犯错、走错路，看到他们亲身上阵，连最平凡的细枝末节都要追查到底，为了寻得一个解法百折不挠。有时是一开始就错了方向，有时是中间误入歧途，但经历了所有失败后，你会看到那些深刻的见解越来越接近最终的答案。高尔斯这样形容，博学者项目这种方式之于“常规研究就好像众人推车之于

开车。”项目仅仅进行了37天，高尔斯便宣布，他确信，这些博学者们不仅仅解决了他最初提出的那个难题，还解决了一个更难的问题，而他的那道题目不过是这个问题的一个特例。谈到此次经历，他说“这可以是我数学生涯中最激动人心的6周之一吧。”此后还有数月的整理工作要做，虽然这些后续工作的量更多，但最核心的数学难题已被攻克。（如果你对高尔斯提出的数学难题有兴趣，想了解更多细节，请参见附录。如果你只想继续阅读本书，该内容并不会影响阅读，可放心跳过。）

博学者们坐不住了。自高尔斯首创先河后，十来个博学者项目和类似博学者项目出现了，所提出的有些问题甚至超过了高尔斯那个题目的难度，更加野心勃勃。超过100名数学家和其他领域的科学家参与其中；大规模协作的风潮席卷了整个数学界。正如首个博学者项目一样，这些项目中有一些确实取得了巨大的成功，加深了我们的数学认识。而其他项目的成果就比较有限了，并未能达到预设的目标（有时候也是因为他们太过于好高骛远了）。无论如何，大规模协作已成为攻克数学难题的一种强而有力的新方式。

为什么大规模在线协作在解决数学难题上能行得通？部分原因在于，即使最顶尖的数学家也可以从不同知识背景的人那里，学到大量能够弥补自身不足的东西，并受到启发，以自己之前绝不会考虑到的角度来思考问题。在线工具创造了一个共享空间，提供了协作发生的平台；它还提供了短期集体工作存储功能，让许多人能够参

与进来，迅速改进各家观点。这些工具让我们能够成倍地扩大创意交流，相反，如果缺了它们，这些互动就不会是发展之必然，而是需要非常偶然的运气才能有幸获得的了。大规模在线协作加快了问题的解决速度，并扩大了人脑所能解决的问题的范围。

博学者项目不过是冰山一角，它背后还有更宏伟的图景，那幅图景为人们展现了在线工具正在如何改变科学家进行科学新发现的方式。这些工具都属于认知工具，能够有效增强我们的集体智慧，提高我们的聪明才智，让我们优秀到足以解决最艰深的科学难题。要了解这一切为什么那么重要，就需要回顾17世纪近代科学的起步阶段，回顾伽利略观察到木星卫星，牛顿得出万有引力定律等众多重要发现诞生的时期。

伽利略、牛顿等科学巨人留给后世最为重要的遗产并不是那些一次性的科学突破，而是科学发现方法本身，即了解自然运作规律的方法。在17世纪初，哪怕是最微不足道的科学进步也离不开杰出的天才。不过，通过不断发展科学发现方法，早期的科学家保证了到17世纪末，诸如此类的科学进步会是稀松平常的事，任何合格的科学调查都可能有所发现。过去非天才不可的事变得普通，科学也随之迎来了爆炸式的发展。

科学发现方式的这些改进比任何单一的科学发现本身要更为重要。这些改进拓展了人类思维能力的边界，让人们得以迈进自然新领域中进行探索。在当今社会，在线工具为我们进一步发展科学发

现方式提供了崭新的机遇，这个机遇直到近代科学起步的阶段才为人们所察觉。我相信，科学方法，即取得科学发现的途径在下一个20年里的巨变要比之前300年的累积量都要多。

博学者项目不过是这场巨变的一个方面而已，展现的是科学家通过协作创造新知的方式的一个转变。巨变的第二个方面在于科学家发现知识背后意义的能力的巨大提升。举个例子，大家想想看，你经常在新闻报道中看到研究说“这样和那样的基因导致了一种这样和那样的疾病。”而让这些研究成为可能的功臣是人类遗传图谱，它是在过去20年中聚合而成的。其中最著名的便是科学家2003年完成的人类基因组图谱。

但2007年完成的人类基因组单体型图（HapMap，全称haplotype map）虽然没那么有名，却可能有着更重大的意义，该图谱记录了不同的人的遗传密码在什么位置有怎样的差别。这些基因变异在很大程度上决定了我们对同种疾病的不同易感性，而人类基因组单体型图能告诉我们那些变异可能出现在哪——它并不仅仅是某一个人的遗传图谱，它是全人类的遗传图谱。

人类遗传图谱是全世界许许多多生物学家共同的工作结晶。他们每次在自己的实验室获得一大批新的遗传数据，就会将这些数据上传至基因银行（GenBank）等集中式在线服务系统中，基因银行是一款令人为之惊叹的遗传信息在线数据库，它的运营商是美国国家生物技术信息中心（the US National Center for Biotechnology

Information)。基因银行会将所有的遗传信息整合到一个单一、公开、可自由存取的在线数据库中，该数据库是数千生物学家工作成果的汇总。这些信息的规模之大基本已不可能纯人工分析了。幸运的是，世界上任何一个人都可以自由下载这份遗传图谱，并使用计算机算法对其进行分析，这样一来，也许会在人类基因组领域取得前所未有的发现。只要你愿意，你也可以立刻登录基因银行网站，开始浏览遗传信息。（基因银行和其他资源的链接，请见347页的“资料来源说明”。）实际上，这就是这些研究能够将基因和疾病联系起来的原因：从事这些研究的科学家先要找到大量患有某一疾病的人，还要找到大量没有患该疾病的人作为对照组，然后再利用人类遗传图谱找到疾病发病率和两组研究对象遗传差异之间的联系。

科学界各领域都在广泛使用着类似的科学发现模式。许多领域的科学家都正在利用在线协作方式创建庞大的数据库，以绘出宇宙的结构、全球气候的结构、全球大洋的结构、人类语言的结构，甚至所有物种的结构。将数百位或数千位科学家的工作整合到一起，我们就能够集体绘制出整个世界。有了这些聚合出来的图谱，任何人都可以利用计算机算法挖掘出事物之间不曾被发现的联系。稍后，本书将给出各种各样的实例，从追踪流感爆发的新方式的发现到两个环绕彼此运动的超巨型黑洞的发现。我们正在一块一块地，用全球所有的知识砖块修建一栋庞大而复杂的知识大厦。这栋大厦太过庞大，任何个人都不可能独自完成它。不过，新兴的计算机化工具

能帮助我们找到这些知识背后的意义。

如果说，博学者项目展示的是科学家协同创造新知的方式的转变，基因银行和这些遗传研究展示的是科学家发现知识中潜藏的意义的方式的转变，那么第三个大变化就是科学和社会之间关系的转变。这种转变的一个例证是星系动物园（Galaxy Zoo）网站，该网站招募了超过20万的网络志愿者，来帮天文学家对星系照片进行分类。志愿者会一边看着星系照片，一边回答各种问题，如“这是一个旋涡星系还是一个椭圆星系？”，“如果这是一个旋涡星系，那么旋臂转动的方向是顺时针还是逆时针？”这些照片都是程控望远镜自动拍摄的，此前没有任何人看见过。你可以把星系动物园当做是对宇宙居民的一次普查，这也是有史以来规模最大的一次普查，截至目前已经对超过1.5亿个星系完成了分类。

参加星系动物园的这些志愿天文学家正在进行着惊人的发现。例如，他们最近发现了一个全新的星系类型，“绿色豌豆星系”——之所以这样命名是因为这种星系看上去确实很像绿色的小豌豆，此类星系中恒星形成的速度几乎比宇宙中其他任何地方都要快。他们还发现了一例类星体镜像，这也被认为是该类型的首例，类星体镜像直径达数万光年的巨型气体云，被邻近的类星体发出的光加热而发出了明亮的光。在短短3年时间内，根据星际动物园志愿者的工作成果著成的科学论文已经有22篇了，还有更多的论文正在进行或准备中。

星系动物园只是众多在线公众科学项目之一，这些项目正在招募志愿者帮助科学家解决科研难题，而大多数的志愿者都没有接受过科学方面的专业训练。这样的例子在各个科学领域都能见到，有的志愿者是在用电脑游戏预测蛋白质分子的形状，有的是在帮助解密恐龙是如何进化的。这都是些严肃的科学项目，在这些项目中，大批几乎完全没有受过专业科学训练的志愿者可以攻克小批专家无法解决的科学难题。任何一个专家组都不可能做星系动物园在做之事——就算这些专业人士愿意专职从事此事，但也没有将数十万（或更多的）星系进行分类的时间。你也许认为他们应该用计算机来对星系照片进行分类，但实际上，这些志愿者的准确度甚至超过了最好的计算机程序。所以，参加星系动物园等项目的志愿者正在扩大人类可以解决的科学难题的范围，此举也同时改变了两个问题的答案，一是谁可以成为科学家，二是成为一名科学家意味着什么。职业科学家和业余科学家之间的界限能有多模糊呢？我们是否会在某一天看到诺贝尔奖由业余人士主导的大型协作项目夺得呢？

科学与社会之间的关系发生了巨大的转变，公众科学只是其中的一部分而已。星系动物园这类的项目是一些制度的例子，这些制度正在运用崭新的方式搭建科学与社会之间的桥梁。我们将会看到，在线工具让许多其他新式桥接制度的出现成为了可能，包括开放存

取出版¹和科学博客，前者让公众能够直接获取科学成果，后者正在帮助建立一个更开放、更透明的科学界。在为科学和社会搭建桥梁上，我们还能找到何种新方式呢？而这些新兴的桥接制度将会带来怎样的长远影响呢？

这个故事讲到现在仍是十分乐观的，讲述了可能性，讲述了正在改变世界的新式工具。实际上，这个故事还伴随着一个问题，一些重大障碍正阻碍着科学家，让他们无法充分发挥在线工具的优势。现在我们借用刚才讨论过的一个例子，即能将基因和疾病联系起来科学研究，来让大家更好地了解这些障碍究竟是什么。虽然前面我刻意忽略了这个问题，但它却是整个故事中至关重要的一环，而真正令人倍感困惑的是：为什么率先在基因银行上分享遗传数据的是生物学家呢？你一旦想到了这一点，就会发现这个选择非常奇怪：你若是一名专业的生物学家，那么对你有利的选择就应该是让这些数据成为秘密，而且越久越好。你为什么要在有机会利用这些数据发表论文或取得专利前，就在网络上分享出这些数据呢？在科学界，能给你带来工作和升职的只有论文和专利，当然，专利只适用于某一些

1 开放存取出版（open access publishing）：2001年《布达佩斯开放存取计划》（Budapest Open Access Initiative, BOAI）给其的定义为：对某文献的“开放存取”意味着它在Internet公共领域里可以被免费存取，并允许任何用户阅读、下载、复制、传递、打印、搜索和超链接，也允许用户将其遍历并为之建立索引，用作软件的输入数据或其他合法用途。用户在使用该文献时不受财力、法律和技术限制，而只须在存取文献时保持文献的完整性，对其复制和发行是唯一的限制。（译注）

科学领域。通常来说，将数据公开对你的职业生涯毫无益处，如果这些数据帮助你科学上的竞争者，那么它甚至可能对你的职业生涯造成伤害。

基因银行于1982年正式上线，但起步非常缓慢，这也是因为受到了这些原因某种程度上的影响。许多生物学家都很高兴能够在基因银行中获取他人的数据，但却对贡献出自己的数据毫无兴趣。不过，随着时间推移，这种情况发生了改变。改变的部分原因是1996年在百慕大岛召开的一次历史性会议，此次会议的与会者中有许多世界顶级生物学家，其中包括了好几位由政府资助的人类基因组计划的项目带头人。克雷格·文特尔(Craig Venter)也有出席，几年后，他领导了一个非政府的人类基因组测序项目。尽管许多与会者都不愿意在发表前就率先、单方面地将自己的所有遗传数据拿出来共享，但若公开数据成为科学界的习惯做法，那么必将大大地造福整个科学界。所以，他们坐下来就这个问题讨论了很多天，最终达成了一个共同协议，现在被称之为《百慕大协议》，协议规定所有的人类遗传数据应该即时上传网络进行共享。该协议并非一纸空文。这个会议室里的生物学家有足够的影响力，他们成功说服了几家大型科学资助机构将数据的即时共享定为进行人类基因组工作所必须遵循的要求。他们不会给拒绝共享数据的科学家提供研究资金。此举改变了游戏规则，即时共享人类遗传数据成为了一个准则。《百慕大协议》终于传达到了政府的最高层：2000年3月14日，美国总统比

尔·克林顿和英国首相托尼·布莱尔发布了一份联合声明，对《百慕大协议》中的原则大加赞扬，并敦促各国科学家采用类似准则。正是因为有《百慕大协议》和随后出现的类似协议，人类基因组图谱和人类基因组单体型图才能对公众公开。

这是一个欢乐的故事，却有个不欢乐的结尾。《百慕大协议》本来就只适用于人类遗传数据。此后，为了让更多的遗传数据得以共享，人们做了许多想要进一步发扬该协议精神的努力。尽管如此，还是有许多生物的遗传数据仍是秘密。举例来说，直到2010年，还没有规定共享流感病毒数据的全球性协议。向这一协议迈进的步伐因领导团体的争执不休而陷入了停顿。最近，一位科学家告诉我，他已经“坐拥”某种生物完整的(!)“基因组”一年多了，这个例子也向大家展示了许多科学家对共享非人类生物的遗传数据一事究竟是怎样的态度。对科学家来说，没有任何共享的动机，倒有许多不共享的理由，所以他们选择将自己的数据藏起来。结果，一种新型的数据将人类等为我们所理解的生物，与流感病毒等生物一分为二，前者所有的遗传数据几乎都已在网络上公开了，而后者的重要数据都还处于非公开状态。

听着这个故事，好像里面的科学家都是贪婪消极的。毕竟，这项研究的经费通常都来自公共资金。科学家难道不应该将自己的研究成果尽快公开吗？这些想法都没错，但实际情况却很复杂。为了弄清楚究竟是怎么回事，你需要先了解，压在那些有上进心的年轻

科学家肩头的竞争压力有多惊人。一所重点大学出现一个非常好的长期职位空缺是很难得的，而极其符合要求的申请者通常就会有数百位。年轻科学家的工作竞争非常激烈，所以每周工作80个小时以上是很普遍的。他们将这些时间都尽可能多地花在能帮他们争取到工作的一件事上：积累数量惊人的科学论文。那些论文将给研究吸引来资金，将给论文作者带来找到长期工作所必须的推荐信。在获得终身职位后，他们的步伐才会放松，但想要持续获得资金支持的话，勤奋工作这一原则仍是必须坚持的。这样下来的结果就是，许多科学家原则上表示愿意将自己的数据在发表前就分享出来，但他们同时也非常担心此举会让竞争对手获得不公平的优势。那些竞争者可以利用那些数据让自己的研究结果率先发表，或者还可能出现更糟的情况，即他们会直接窃取这些数据，作为自己的研究结果发表。只有当每个人都受到类似《百慕大协议》这样的集体协议保护的时候，共享数据才会可行。

另一些例子中也是类似的情况，科学家拒绝为其他许多的在线项目贡献资料。以网络百科全书“维基百科”为例，它有一则让科学家感到温暖的愿景宣言：“想象一下，世界上的每个人都可以自由地共享所有知识。这是我们的承诺。”你也许以为维基百科的创始人是一群渴望将世界上所有的知识都拿出来共享的科学家，但你错了。实际上，维基百科的创始人是吉米·威尔士（Jimmy Wales）和拉里·桑格（Larry Sanger）。吉米·威尔士，网名神保（Jimbo），当时