

文章编号:1003-2053(2004)03-0225-08

科研评价和科学计量学的研究纲领 ——二者关系的历史演变与重新定义

洛埃特·雷迭斯多夫

(Science and Technology Dynamics, University of Amsterdam, The Netherlands)

摘要: 科研评价以科研表象为基础。在提高研究系统的表象质量的过程中,无法避免与这些研究系统迥异的接收系统为指标赋予意义。由于决策影响到所要探究的系统,这种结构产生的张力正在推动“二战”以来科学指标的进一步发展。本文从历史的角度讨论了科学指标的出现和所涉及的方法论问题,介绍了欧盟在跨国层次上的政策协调以及因特网作为全球交流的媒介而带来的新的发展动向。科学、技术和创新政策在不同层次上以不同目标日益发展,有望对评价话语加以区别。

关键词: 科学计量学;科研评价;历史演变

中图分类号:G301

文献标识码:A

科学计量学指标应用于科研评价始于 20 世纪的 60-70 年代,首先出现在美国,然后出现在欧洲各国。在那之前,一方面,除了同行评议制度外,科研评价并没有正规的形式,另一方面,只是在系统的宏观层次(如国家层次)上才使用各种经济指标。经济指标(如 R&D 占 GDP 的百分比)被总部设在巴黎的经济合作与发展组织(OECD)开发成为国际性的指标。例如,《弗拉斯卡蒂科学和技术活动测量手册》说明了科学技术对经济发展的重要性,这种重要性在 20 世纪 50 年代的经济统计上已经显露出来。

从任务的角度(例如,为了军事目的),认为科学知识可以按要求组织并加以控制的想法是“二战”的结果。在那以前,知识的智力组织很大程度上取决于学科构成和专家交流的内部机制。“二战”期间,通过基于知识的开发和任务导向的军事任务影响了科学技术的发展(如曼哈顿工程),因此,有必要在 1945 年制定一种适合于和平时期的新的科学技术政策。

1945 年,万尼瓦尔·布什给美国总统的报告《无止境的前沿》呼吁回到自由的科学组织。科学的质量应该通过科学精英的内部机制加以控制,例如,通过同行评议制度。其它西方国家纷纷仿效美

国的国家科学基金会^①模式,例如,荷兰在 1950 年建立了基础科学研究基金会(ZWO)。事后看来,人们可以认为这是科学政策的建制阶段,主要的政策手段是建立专门机构对投资加以控制,进而实现对科学的支持。

1 人造地球卫星的冲击

1957 年,苏联人造地球卫星的发射扭转了科学政策的局面。在任务导向的研究中,实施非自由模式的苏联似乎比西方更成功。在人造地球卫星的冲击之下,美国艾森豪威尔总统感受到了科学精英与军方结成的联盟要求加大对科学系统投资的压力。除了他对“军事-工业联合体”的压力发出警告之外(在告别演说中),他还阐明了鲜为人知的“第二警告”:目前在重视科学的研究和科学发现的同时,我们应该,我们也必须对同样存在的一种相反的危险保持警惕,那就是公共政策本身可能成为科学技术精英的俘虏。”^[2]

美国科研系统的重大重组就是这一时期的产物,特别是美国为了应对人造卫星的发射和军事威胁建立了国家宇航局(NASA)和高级研究计划署(ARPA)^[3]。

^①事实上,1947 年美国国家科学基金会的第一个法案被杜鲁门总统否决了,因此基金会的成立延迟到了 1950 年。后来更加详细的斯蒂尔曼报告(《科学和公共政策》,1947 年出版)^[1]激活了联邦政府在资助大学基础研究中的实际角色。

收稿日期:2004-03-29;修回日期:2004-04-30

作者简介:洛埃特·雷迭斯多夫(Loet Leydesdorff),男,荷兰阿姆斯特丹大学教授,著名科学计量学家,普赖斯奖获得者。

与此同时,诸如土地、劳动和资本等传统的经济因素越来越明显地不能解释西方经济的持续增长。“剩余物”^[4]只有用知识经济的兴起来解释^[5]。除了NATO^[2]的军事协调,发达工业国家在1961年还成立了经济合作与发展组织(OECD)^[3]负责组织和协调成员国之间的科学技术政策。正如上文所指出的,在1963年催生了弗拉斯卡蒂手册^[6],在对比的基础上使用参数对科学和技术进行统计上的监测。通过国家之间的比较,有可能提出潜在的优势和劣势问题。在60年代后期,国家的科学技术政策开始在OECD成员国中出现。

例如,统计表明:“二战”之后,物理学按照各国内的利益以及国际合作的利益成功地加以组织(如:CERN^[4])。其它科学共同体(如:分子生物学)要求给新的发展提供更多的预算空间和全面增加学科预算。1965年—1975年间,在国家层次上更为理想地处理这些问题的手段是对学科预算增长率的细分。总而言之,焦点仍然是(财政)输入指标,而这套系统在较低层次上仍依赖于同行评议以提供更周密的决策^[7]。

2 输出指标

人们对科学计量学的关注起源于科研评价之外的另一种兴趣。在20世纪50年代和60年代期间,科学共同体本身逐渐意识到科学情报和文献的数量在战后不受控制地膨胀。除信息检索的用途之外,由Eugene Garfield科学信息研究所创造的科学引文索引(Science Citation Index)很快被认为是使标准客观化的工具^[8-9]。输出指标(即出版物数量和引文数量)的逐渐引入,可以在社会层次上,也可以在科学共同体内部层面上得以合理化。这是因为在社会层次上,它使决策者和科研管理者可以利用经济效率的论点;在内部层次上,跨学科范围的质量控制很难做到合理化,除非同行评议之外还有其他客观的标准可供使用。

1976年Francis Narin的开创性研究《评价的文献计量学》^[10]在美国国家科学基金会的赞助下(并非容易地)出版。Henry Small^[11]提出一种基于科学论文共引(co-citation)对科学进行数字图像映射的

方法。Small的方法试图凝练学科结构的特性,而Narin却重视自上而下的等级结构^[10,12]。这种方式对美国国家科学基金会和美国国家卫生研究院这类面临跨学科框架下进行预算分配的资助机构很有吸引力。

从期刊聚类入手,有可能对科学作出实质性的和有等级秩序的分类吗?在衡量的标准方面,比如期望引用率与观察引用率,权重对出版物有用吗?^[13-14]由此,科学计量学研究纲领在宏观层次上日益形成。美国在70年代的研究关注科学文献的组织,80年代,欧洲已在机构的基础上应用这些指标进行科研评价。

在Martin与Irvine^[15]对射电天文学各种(昂贵的)设备的相关研究绩效作了研究之后,基于机构单元(institutional units)的评价观念在决策者中占据优势。荷兰莱顿大学在80年代率先提出了一个采用产出指标的周密模型。该模型将出版物和认可程度(作为反馈参数的引文)作为产出指标引入院系的财政计划中^[16]。这种理念在英国资助大学研究的科研评价活动(research assessment exercises)的模式中得到推广。

在国家层次上,其它欧洲国家的研究预算模式的合理化没有效仿英国^[17],但是在90年代期间,各种压力使得评价活动不得不使用出版率和引用率。例如,1990年德国统一之后,对东德的研究成果进行广泛的评价立即被列入科学政策的议程^[18]。但是,科学计量学的指标能负载科研评价的政治负担吗?

3 方法论的复杂化

出版物和引文分析已成为科研评价的标准工具。然而,一些方法论问题仍然是悬而未决的。随之产生的不确定性有时在决策过程和科研管理中是否应用这些工具作为标准而产生的犹疑中反映出来了。科研评价的立足之地有多么不稳定呢?

首先,人们可以提出科学知识生产和控制的分析单元问题^[19]。科学的智力组织与它们的机构组织不一致。而且,可以认为,这些组织层次之间的关

^② NATO:North Atlantic Treaty Organization 北大西洋公约组织(译者注)

^③ OECD的前身是OEEC,欧洲经济合作组织,该组织在战后的马歇尔计划中负责美国和加拿大的援助分配。

^④ CERN:European laboratory for particle physics,欧洲粒子物理实验室(译者注)

系在学科中是不同的^[20]。从这个观点看,按照机构参数进行“同类相比”的假设是有问题的。新的科学发展(如人工智能)在非常不同的机构环境下出现,想要作公平的比较,人们或许首先应该定义一个“认知”的分析单元。然而,科学的智力组织不可能轻易地被观察或者被测量^[21]。

定义分析单元的另一种方法是把操作建立在科学期刊文献所反映的科学发展的基础之上。科学文献被组织在相对离散的期刊簇中,例如,生物化学期刊中的一篇论文不会经常引用凝聚态物理的论文,反之亦然。既然学科之间的出版率和引用率是不同的,这些文本的分析单元和评价的机构单元之间的关系就产生了有待于进一步研究的问题。

文献的相对分解性是Narin试图将集总的期刊—期刊引文数据库进行聚类的核心问题。无论如何这种聚类算法提供了一点参考。而在给定的任一时间,这种结构不能说明科学在过去的一段时间里的动态发展。人们希望学科专业是并行发展而不是分层发展的。

此外,Narin还建议选定一种期刊以便使历时性的比较成为可能。然而,发达工业国家比那些处于保守体系的研究机构更趋向于在新兴领域(相应的新期刊)发表文章。所谓的“英国科学的衰落”^[22]——20世纪80年代期间政治辩论的一个激烈主题,事后看来,在某些方面可以被看成是人为的方法论的产物。既然失之东隅收之桑榆,英国在一个动态的数据库比在一个固定的设置中更稳定^[23-25]。

4 多变量和动态分析

期刊文献能被认为是一个通过文章的引文集总以及正文和标题的词共现而编织在一起的巨大网络。因子分析方法通过编组,使我们能及时地假设这个网络随时的潜在结构。然而,因子分析往往太强调计算以至于在单一的运行中无法说明整个数据库,更不用说全面地和动态地说明这些问题了。

图分析方法基于展示关系的回归(自下而上)程序,因此在大型数据集中更容易得到应用。然而,按照定义,相关分析的结果只能显示关系和层级。在网络中,编组的位阵分析——即按照关系而扩展——在类型上是不同的^[26]。但是,如同指出的那样,位阵分析仍然是局限在比较有限的数据集中。

巴黎的国家高等工业学院(*école Nationale Supérieure des Mines*)的一群研究者提出要重视词和词间关系(“共词”),以替代引文分析和共引文分析^[27]。这种方法的优点是词和词共现不仅出现在科学期刊中,也出现在政策报告和专利应用文件中。在跨领域的“转译”中,词间关系的强度可以作为一种表达概念存在价值的指标吗?这些作者认为,如果能够将词和词共现作为“转译”指标,那么,按绩效评价科研将成为可能。^[28-29]

对共词模式的分析,在技术上类似于美国科学信息所(*ISI*)Small小组开发的共引分析^[11]。与此同时,*ISI*小组制作出了基于图分析算法聚类技术的《科学地图集》^[30]。然而,由于最初的研究定位在结构和战略位置的层级关系上,因此,这种数字图像映射方法就变得有缺陷了。结构只有分解成系统的潜在维度才能被分析。这些维度使用因子分析技术才可以被揭示^[31-33]。

当科学作为交流网络按照词间关系加以建造时,可以预想,集总的构造(*aggregated constructs*)能在较长时间里根据规则加以区别,而这些规则对专业化的、相对自治的科学交流结构的智力组织的进一步发展是起作用的^[34-35]。各种各样的论点继续不断更新并且自反地重写它们对于有关历史的理解。从现时理解的事后观点看,“自组织”(范式的)论点既发生在活跃的科学家中间,也发生在政策机构的层次上(或许有时间滞后)。后者只有通过财政和机构手段指导并且影响科学的发展。然而,研究的前沿不同于机构组织,它是伴随着科学交流的动态机制而发展的。

例如,人们现今所认为的“生物技术”或者“人工智能”与决策者和科学家在八十年代最初提出这些领域优先计划时所认为的相关范畴相去甚远^[36]。随着历史的发展,重建和重写历史的双重任务产生了重建中的“双重解释”的需要^[37-38]。可观察的变化和选择机制(网络层次上的潜在的特征矢量)会随着时间而改变^[39]。

尽管变化在数据中是可见的,但选择机制仍然是潜在的,因此只能被假设。一方面,需要构建一些用于数字图像映射的维度。另一方面,构建的变量是“软的”,即可以讨论和重构。现在看来,普赖斯^[40]梦想把科学计量学的数字图像映射变成一门相对硬的社会科学可能是根本错误的^[41-42]。当数据和观点潜在地改变时,分析者的立场不再被视为:

是中立的。

在 90 年代期间,科学计量学的科研评价蒙受分离破碎之苦,这一点社会科学家都深有体会。科研评价日益成为评价机构的问题。1994 年,当《科学计量学》杂志发行专刊来讨论“小科学计量学,大科学计量学和超科学计量学? (Little Scientometrics, Big Scientometrics and Beyond?)”时,危机变得明显化了^[43]。这场反思不仅变得比以前更不确定,而且所反思的主题也随着人们对技术科学和基于科学的创新的重视而开始改变。

5 欧盟和研究评价领域

尽管科研评价作为国家机构层次上的议程已经形成,但是欧共体 1986 年的单一法案 (Single Act) 和 1991 年的马斯特里赫特条约 (Maastricht Treaty) 表明欧洲内部正在逐步向跨国技术和创新政策转变。欧盟的政策反复提到科学和技术,因为科学和技术被认为是成员国共同遗产的堡垒。然而,“辅助性”原则规定欧盟委员会不应该涉入主权国家自己能解决的问题。因此,欧盟的联合研究计划得不到发展,除非集中实施基于科学的创新行动的迂回战术^[44]。

通过奖赏那些利用国家优势和机构间的互补性而形成的研究群体,欧盟的研究、技术与发展 (RTD) 网络促进了跨国和跨部门的合作。这样,一种“大学-工业-政府”的三螺旋关系得到欧洲决策层次上的强化。预期中的协商和转译系统趋向于按照国家的机构体制增加动态的迭加^[45]。

除了事前的机构合理设置外,相互作用也可以被事后的目标所优化;在不断重复的时间段内,网络系统可以逐渐提供它们自己的动力。根据这个观点,机构层可以被认为是一种网络的保持机制,通过它的功能趋向于更进一步的发展。通过固定“次等级系统”(next-order system) 层,欧盟已提供了对这些功能的反馈。迭加系统可以是一种网络模式的概念——Gibbons^[46] 等人称为“模式 2”——或者是一种“国际化”的概念,这是进化经济学家在先前的“国家创新系统”的比较研究中提出的^[47-48]。

除了国家层次上普遍使用的机构测量之外,新模型可被看作是新的进化,因为它们为测量提供了启发。例如,国家政府在发展跨学科计划中已经取得一定的成功^[49-50]。然而,欧洲的政界更注重一种

转变,即从传统的科学出版物转向成果,转向“可交付使用的物品”(deliverables),并使之商品化,它们较少地受到国家研究理事会和科学共同体的评价方案的阻碍。这些“可交付使用的物品”的科研评价模式中已经成为下一轮政策制定的要点。如此看来,科学文献滞后于研究议程^[51]。

一种(跨国)科研评价方案可以根据研究者给官方提供的“可交付使用的物品”的可靠性,将用户服务能力对它们进行分类排序。在“模式 2”中,不仅社会的科研工作,而且项目和计划的管理组织的科研工作都可以通过对用户的服务而日益发挥作用^[52]。应该注意的是,这种转变不一定意味着科学的商业化,因为这种机制仍然主要是机构的而非市场的^[53]。

Leydesdorff 和 Etzkowitz^[54] 已经论证,要将“革新”作为与知识发展相结合的“创新系统”运行的分析单元。然而,描绘一个以知识为基础的创新系统本身就回避了问题。进化经济学家已经强调系统的国家性特征^[47,48,55],而强调技术发展的全球性^[56-57]。将部门层次作为最相关的参照系统^[58]。个人^[58]则力主将诸如生物技术这样的新的技术作为知识整合的框架。各种创新子系统交织在一起,会组织,并且相互推进(或者相互制约)。因此,科研评价的参照系统本身日益成为了一个问题。

6 因特网与电脑计量学(cybermetrics)和网页计量学(webmetrics)的发展

20 世纪 90 年代期间出现的因特网再一次改变了形势。全球化发生在机构层次之上,它加强了科学、技术以及具有不同的潜在协调机制作用的经济之间的直接联系。那么,机构组织(例如国家层次上)就可以被认为是在全球环境中通过资源和赢得信誉而求得自我发展(或停滞)的小环境。

虽然参与交流的组织在交流网络层次上只提供了表象,但是这类表象提供了原始材料,但是这类表象在网络中可以像“行动作用”(actants)一样加以传播。表象之间的相互作用开始在网络上共振(resonance)。同时,行动作用背后的作用者便日益黑箱化,即表象的系统也许会变得依赖于它们在网络层面上的表象,而参与交流的系统则自反性地知道它

被反映^[59]。

在这些条件下,科研评价只能通过参考被评价的表象自反性地定位自己,因为它分析不清楚(有时不易被理解)所要精确反映和代表的事物^[60]。文本嵌入语境中,而在具体的交流编码中后者为前者赋予了意义。利用不同数据库或就相同数据得到的不同观点而形成的不同话语可以服务于不同的听众。可以通过接触面(交界面)实现表象的功能,从而对表象加以评价。

例如,像一种新药那样的创新,其创新的意义对于置身于市场的公司和对于蒙受疾病痛苦的患者,或者对于开发药物的科学家和药剂师是不一样的。后者可能使用《化学文摘》(Chemical Abstracts)作为他们的参照系统或者将“医学在线”(Medline)作为检索和参考的工具。科学共同体所使用的药物原名对大多数用户是陌生的,后者仅仅通过其商品名称才知道同样的物质。可以认为,对于不同的对象,如分子生物学家和从事临床实验的医学科学家,评价方案是不同的^[61]。

使用这些不同话语的机构在他们的环境(语境)中只具备与其相关话语进行交流的有限能力。这种话语的构建使每个机构能够着重于自己的交流质量。凌驾于“模式2”之上的研究承担着囿于形式从表象到表象,而不触及其实质的风险。虽然为这些复杂的相互作用建模是一件合乎情理的事,但是抽象的模拟结果却与科研评价的实际见识差异很大。当后者能够凭直觉去理解并且转译成政策时,前者首先要求得到理论上的解释。

7 自反的科学计量学

对于这些既定的复杂的动力机制,面对科研评价以及其它基于知识的交流活动的评价,我们还有多大的选择余地?“一切固体都可以融于空气中”^[62],但这种融化仍然留下了交流的痕迹。一个交流系统跨越交界面与另外的交流系统交流,如果共有的信息能被充分地重复,系统就“锁定”(locking)并且暂时稳定在交界面上互相形成的一个共演化(co-evolution)状态。人们怎样才能描绘出界面交流的特征及其质量呢?

交流系统(如自反性的话语)不可能被看作具有清晰描绘的既定系统。然而,交流的系统(子系统)可以被重建,因为它们是历史上构造的,并且在

它们的轨迹已被稳定的范围内,这种分析单元才能服从测量,最终服从管理^[63]。人们能够观察事件,但是事件仅仅与话语联系时才有意义。然而,人们可以根据交流特征的理论信息对交流系统进行假设。通过首先提出“当系统处于交流状态时,什么被交流?”的问题,分析者可以重建所要研究的交流系统的代码。只有在提出这实质性的问题之后,人们才能有意义地提出“这些编码是如何被交流的?”问题。然后,也才能在参考系统的既定特征下提出测量指标。

由于进行交流的并可进行重新分配的是一种具体的(编码的)实体,因此,交流服从于测量。分配中引起的变化产生概率熵(用信息比特来表示)。这种信息的数学定义仍然是与内容无关的,参考系统的界定给这种不确定性赋予了意义。如上所述,参考系统的界定本身就是一项既困难又具分析性的任务,因为人们需要两方面信息,一是所研究的交流与其(在每一个时间点上)环境相关的系统之间的关系,二是所要研究的系统所赋予意义的内在发展。

从这个观点看,历史研究与科学计量学分析是相关的。在话语重建中,分析性的说明尽可能最好地再现了所要研究的话语。然而,表象除了可以起表示系统的那些作用外,表象还有其他目的。表象的质量——按照分析特征和定量精度——对成本(costs)是至关重要的,它涉及到对这些表象的管理,使之对它们所表示的系统以及它们的进一步发展为社会所接受。

在一个看似爆炸性的交流和表象的宇宙里,我们被置于何处?根据科研评价的观点,只有问清楚在一个既定的话语里是什么在交流之后,科学计量学指标的发展才能自反地改进交流的质量。可以认为,定性的话语在功能上是明确的。所要研究的话语对领域内的任务和/或科研机构的任务起多大的作用呢?

对我而言,似乎不同的目标能被区别。首先,当制定公共的科学技术政策或者在进行机构层次上的R&D决策时,有必要在系统的层次上提供优质信息。这种信息只能被认为是局部构造起来的关于特定研究问题的一部分指标。然而,众所周知,鉴赏和评价的窗口糟糕地切分了所要研究的系统^[64]。人们在这个方向上要谨慎行事。

其次,科学计量学已使我们知道不管表象多么地不完善,科学是服从于测量的。科学史、科学社会

学和科学哲学可被看作是对科学进行研究的一种定性表象。定性方法和定量方法之间的关系能被再次阐明：定性的表象和洞察力给测量提供假设和启发，而测量方法能通过考虑相互作用的条件而加以更新和精炼。

定性方法从不同的角度有助于形成模型，而定量分析的结果再一次被赋予来自各种角度的意义。然后模型可被认为是使我们的科学理论进一步发展的一台机器。然而，比起现代主义者们的观念来，自反的模型也没能实现这个功能。与所表示的系统相比，所有的表象必定还是不完整的。片段表象的总和不一定比它们的差异所提供的信息更多。考虑到各种贡献的潜在差异，定量的科学技术研究议程，或者说科学计量学不可避免地要进行方法论的转向：测量不能再回避测量什么和为什么要测量的问题。

8 结论

鉴于科学计量学的研究纲领产生于大学、政府和工业界的建制的话语语境中，三者关系的系统发展在 20 世纪下半叶已经改变了科研评价的参照系统。在知识经济时代，科学已经实现了改变研究定义的功能，全球化已经改变了国家参照系统的相关性。很显然，欧洲在跨国层次上致力于工业的国际化和创新的全球化取得了主导地位。当然，科学从一开始就被定位于国际化，而在全球化发展中科学的研究的过程在科研评价和科学计量学的测量中得到反映。

换句话说，所要研究的系统已变得更复杂。一个复杂的动力系统可以从分析的角度被分解为若干个子动力系统。例如，人们可以提出这样的问题：作为新的子动力系统，也就是具有国际科学合作与跨国合著特征，而不是国内合作和合著特征的新的子动力系统是否已经形成于 90 年代？^[65-66] 可以假设和测量一个不同的（如全球的）子动力系统吗？这个科学输出的新维度能在国家系统中为机构管理而制定的计划中加以计算吗？这些问题是在科学、经济、政治背景下形成的界面上的难题。演化的系统和子系统在不同的维度中交流，而评价已成为这些交流编码的一部分。

参考文献：

- [1] NSF. Science and technology then and now. [DB/OL]. <http://www.nsf.gov/od/lpa/news/media/2000/grdrst.htm>. 2003-09-08.
- [2] York H F. Race to Oblivion: A Participants View Arms Race [M]. New York: Simon and Schuster,
- [3] Edwards P. The Closed World, Computers and the Politics of Discourses in Cold War America [M]. Cambridge, MA: MIT Press, 1996.
- [4] Abramowitz M. Resource and output trends in the States since 1870 [J]. American Economic Review, 1956, 46: 5-23.
- [5] Rosenberg N. Perspectives on Technology [M]. Cambridge: Cambridge University Press, 1976.
- [6] OECD. The Measurement of Scientific and Technological Activities: "Frascati Manual" [M]. Paris: OECD, 1992.
- [7] Mulkay M J. The mediating role of the scientific [J]. Social Studies of Science, 1976(6): 445-468.
- [8] Price D de Solla. Little Science, Big Science [M]. New York: Columbia University Press, 1963.
- [9] Elkana Y, J Lederberg, R K Merton, A Thackray, Zuckerman. Toward a Metric of Science: The Adachi Science Indicators [M]. New York: Wiley, 1978.
- [10] Narin F. Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity [M]. Washington, DC: National Science Foundation, 1976.
- [11] Small H. Co-citation in the scientific literature: a measure of the relationship between two documents [J]. Journal of the American Society for Information Science, 1973, 24: 265-269.
- [12] Carpenter M P, F Narin. Clustering of scientific journals [J]. Journal of the American Society for Information Science, 1973, 24(6): 425-436.
- [13] Braun T, W Glänzel, A Schubert. Scientometric Indicators: A 32-Country Comparative Evaluation of Publishing Performance and Citation Impact [M]. Singapore: World Scientific Publications, 1985.
- [14] Schubert A, T Braun. Standards for citation based measurements [J]. Scientometrics, 1993, 26: 21-35.
- [15] Martin B, J Irvine. Assessing basic research: some spatial indicators of scientific progress in radio astronomy [J]. Research Policy, 1983, 12: 61-90.
- [16] Moed H F, W J M Burger, J G Frankfort, A I Raan. The use of bibliometric data for the measurement of university research performance [J]. Research Policy, 1985, 14: 131-49.
- [17] Hicks D, J S Katz. Science policy for a highly competitive science system [J]. Science and Public

- 1996, 23, 39-44.
- [18] Weingart P. Die Wissenschaft in osteuropäischen Ländern im internationalen Vergleich-eine quantitative Analyse auf der Grundlage wissenschaftsmetrischer Indikatoren [M]. Bielefeld: Kleine Verlag, 1991.
- [19] Collins H M. The possibilities of science policy [J]. *Social Studies of Science*, 1985, 15: 554-558.
- [20] Whitley R R. The intellectual and social organization of the sciences. [J]. Oxford: Oxford University Press, 1984.
- [21] Leydesdorff L. The Challenge of Scientometrics: the Development, Measurement, and Self-organization of Scientific Communications [M]. Leiden: DSWO Press, Leiden University, 1995.
- [22] Irvine J, B Martin, T Peacock, R Turner. Charting the decline of British science [J]. *Nature* 1985, 316: 587-90.
- [23] Leydesdorff L. 1988. Problems with the measurement of national scientific performance [J]. *Science and Public Policy*, 1988, 15: 149-52.
- [24] Braun T, W Glänzel, A Schubert. Assessing assessments of British science: some facts and figures to accept or decline [J]. *Scientometrics*, 1991, 15: 165-70.
- [25] Martin B R. The Bibliometric assessment of UK scientific performance: a reply to Braun, Glänzel and Schubert [J]. *Scientometrics*, 1991, 20: 333-357.
- [26] Burt R S. Toward a Structural Theory of Action [M]. New York: Academic Press, 1982.
- [27] Callon M, J P Courtial, W A Turner, S Bauin. From translation to problematic networks: an introduction to co-word analysis [J]. *Social Science Information*, 1983, 22: 191-235.
- [28] Callon M, J Law, A Rip. Mapping the Dynamics of Science and Technology [M]. London: Macmillan, 1986.
- [29] Latour B. Science in Action [M]. Milton Keynes: Open University Press, 1987.
- [30] Small H, E Sweeney, E Greenlee. Clustering the science citation Index using co-citations II: mapping science [J]. *Scientometrics*, 1985, 8: 321-340.
- [31] Leydesdorff L. Various methods for the mapping of science [J]. *Scientometrics*, 1987, 11: 291-320.
- [32] Leydesdorff L. A Validation Study of 'LEXIMAPPE' [J]. *Scientometrics*, 1987, 25: 295-312.
- [33] Lazarsfeld P F, N W Henry. Latent Structure Analysis [M]. New York: Houghton Mifflin, 1968.
- [34] Luhmann N. Soziale Systeme. Grundri? Einer Allgemeinen Theorie [M]. Frankfurt a. M: Suhrkamp, 1984.
- [35] Luhmann N. Die Wissenschaft Der Gesellschaft [M]. Frankfurt a. M: Suhrkamp, 1990.
- [36] Nederhof A J. Changes in publication patterns of biotechnologists: an evaluation of the impact of government stimulation programs in six industrial nations [J]. *Scientometrics*, 1988, 14: 475-485.
- [37] Giddens A. New Rules of Sociological Method [M]. London: Hutchinson, 1976.
- [38] Giddens A. The Constitution of Society [M]. Cambridge: Polity Press, 1984.
- [39] Leydesdorff L. Why words and co-words cannot map the development of the sciences [J]. *Journal of the American Society for Information Science*, 1997, 48(5): 418-27.
- [40] Price D de Solla. Editorial statement [J]. *Scientometrics*, 1978(1): 7-8.
- [41] Wouters P, L Leydesdorff. Has prices dream come true: is scientometrics a hard science? [J]. *Scientometrics*, 1994, 31: 193-222.
- [42] Price D de Solla. Citation measures of hard science, soft science, technology and non-science [A]. C E Nelson, D K Pollock. Communication among Scientists and Engineers [C]. Lexington, MA: D C Heath & Co, 1970, 3-22.
- [43] Glänzel W, U Schöpflin. Little Scientometrics, big Scientometrics...and beyond [J]. *Scientometrics*, 1994, 30(2-3): 375-384.
- [44] Narin F, E Noma. Is Technology Becoming Science? [J]. *Scientometrics*, 1985, 7: 369-381.
- [45] Etzkowitz H, L Leydesdorff. The dynamics of innovation: from national systems and "Mode 2" to a triple helix of university-industry-government relations [J]. *Research Policy*, 2000, 29(2): 109-123.
- [46] Gibbons M, C Limoges, H Nowotny, S Schwartzman, P Scott, M Trow. The new production of knowledge: the dynamics of science and research in contemporary societies [M]. London: Sage, 1994.
- [47] Lundvall B Å. National Systems of Innovation [M]. London: Pinter, 1992.
- [48] Nelson Richard R. National Innovation Systems: A comparative study [M]. Oxford and New York: Oxford University Press, 1993.
- [49] Van den Daele, W W Krohn, P Weingart. Geplante Forschung Frankfurt [M]. A M: Suhrkamp, 1979.
- [50] Studer K E, D E Chubin. The Cancer Mission: Social Contexts of Biomedical Research [M]. Beverly Hills: Sage, 1982.

- [51] Lewison G, P Cunningham. Bibliometric studies for the evaluation of trans-national research [J]. *Scientometrics*, 1991, 21: 223-244.
- [52] Kobayashi S-I. Applying audition systems from the performing arts to R&D funding mechanisms: quality control in collaboration among the academic, public, and private Sectors in Japan [J]. *Research Policy*, 2000, 29(2), 181-192.
- [53] Nowotny H, P Scott, M Gibbons. Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty [M]. Cambridge: Polity Press, 2001.
- [54] Leydesdorff L, H Etzkowitz. The triple helix as a model for innovation studies [J]. *Science and Public Policy*, 1998, 25(3): 195-203.
- [55] Skolnik E B. The Elusive Transformation [M]. Princeton NJ: Princeton University Press, 1993.
- [56] Pavitt K. Sectoral patterns of technical change: towards a theory and a taxonomy [J]. *Research Policy*, 1984, 13: 343-73.
- [57] Freeman C. Continental, national and sub-national innovation systems-complementarity and economic growth [J]. *Research Policy*, 2002, 31: 191-211.
- [58] Carlsson B. New Technological Systems in the Bio Industries-an International Study [M]. Boston/Dordrecht/London: Kluwer Academic Publishers, 2002.
- [59] Wouters P. The Citation Culture [D]. Unpublished Thesis, University of Amsterdam, 1999.
- [60] Rip A. Qualitative conditions for scientometrics: the challenges [J]. *Scientometrics*, 1997, 38(1): 7-26.
- [61] Leydesdorff L. Indicators of innovation in a knowledge-based economy [J]. *Cybermetrics*, 2001, (5): 2.
- [62] Marx K. The Communist Manifesto [M]. Harmondsworth Penguin, 1967.
- [63] Leydesdorff L. A Sociological Theory of Communication: The Self-Organization of the Knowledge-Based Society [M]. FL: Universal Publishers, 2001.
- [64] Casti J. Alternate Realities [M]. New York: Wiley, 1987.
- [65] Persson O, W Glänzel, R Danell. Inflationary Bibliometric Values: The Role of Scientific Collaboration and the Need for Relative Indicators in Evaluative Studies [R]. Paper presented at the 9th International Conference on Scientometrics and Informetrics, Beijing, 2003.
- [66] Wagner C S, L Leydesdorff. Mapping Global Science using International Co-Authorships: A Comparison 1990 and 2000 [R]. Paper presented at the 9th International Conference on Scientometrics and Informetrics, Beijing, 2003.

(刘小玲译 金碧辉校)

The evaluation of research and the scientometric research program: historical evolution and redefinitions of the relationship

Loet Leydesdorff

(Science and Technology Dynamics, University of Amsterdam, Kloveniersburgwal
48, 1012 CX Amsterdam, The Netherlands)

Abstract: Research evaluation is based on a representation of the research. Improving the quality of the representations cannot prevent the indicators from being provided with meaning by a receiving system different from the research system(s) under study. Since policy decisions affect the systems under study, this configuration generates a tension that has been driving the further development of scientific indicators since World War II. The article discusses historically the emergence of science indicators and some of the methodological problems involved. More recent developments have been induced by the emergence of the European Union as a supra-national level of policy coordination and by the Internet as a global medium of communication. As science, technology, and innovation policies develop increasingly at various levels and with different objectives, the evaluative discourses can be expected to differentiate with reference to the discourses in which they are enrolled.

Key words: scientometric; evaluation of research; historical evolution