

科技成果社会影响力评估的国际经验及启示

范英杰^{1,2,3}, 徐芳^{1,3*}

1. 中国科学院大学公共政策与管理学院, 北京 100090

2. 国家自然科学基金委国际合作局, 北京 100080

3. 中国科学院科技战略咨询研究院, 北京 100190

摘要 对英国的卓越研究框架(REF)、荷兰的标准评估协议(SEP)和美国的STAR METRICS这3个典型的科技成果社会影响力评估方案进行比较, 识别了其评估对象、评估程序和评估目标方面的差异, 重点围绕这3个方案的评估内容指标和评估方法进行了深入分析。结合国际经验, 认为中国应当将科技成果社会影响力评估置于公立科研机构绩效评价的核心地位; 需要在贯彻国家战略导向和体现评估对象、评估目标和区域差异的基础上开发符合中国国情的科技成果社会影响力评估内容和指标体系; 要构建大数据评估并使用数据驱动方法进行科技成果的社会影响力评估。

关键词 科技成果; 社会影响力; 评估内容; 评估方法

随着公共财政资金支持基础研究的力度不断加大, 如何认识和评估科技成果的社会影响力成为一个至关重要的问题, 引起了学术界和政策制定者的广泛关注。美国、澳大利亚和欧洲各国对国家、机构等项目层面上科技成果的社会影响开展了评估。如英国的卓越研究框架(research excellence framework, REF)^[1], 荷兰的标准评估协议(standard evaluation protocol, SEP)^[2], 澳大利亚的研究质量框架(research quality framework,

RQF)^[3]等, 均针对大学、科研机构的科学研究成果的社会影响开展综合性评估。此外, 利用数据挖掘技术, 通过算法找到大型数据库中的关联和模式, 并以有意义的方式加以呈现, 从而达到理解已有数据的目的, 是相对较为新颖的社会影响评估模式。目前, 比较典型的基于数据挖掘的评估框架有美国的STAR METRICS (Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competi-

收稿日期: 2019-01-12; 修回日期: 2019-03-15

基金项目: 中国科学院战略研究与决策支持系统建设专项(Y9J0191601); 国家自然科学基金面上项目(71771205); 中国科学院青年创新促进会项目(Y9J0361601)

作者简介: 范英杰, 研究员, 研究方向为科技政策与管理, 电子信箱: fanyj@nsfc.gov.cn; 徐芳(通信作者), 副研究员, 研究方向为科技政策与管理、科技评价, 电子信箱: xufang@casisd.cn

引用格式: 范英杰, 徐芳. 科技成果社会影响力评估的国际经验及启示[J]. 科技导报, 2019, 37(14): 18-25; doi: 10.3981/j.issn.1000-7857.2019.14.003

tiveness and Science),其第二阶段的目标便是使用革命性的数字技术获取科研投入在科学、社会、经济以及劳动力产出等方面的影响,代表了科技成果社会影响评估的一个新的方向。

中国在建设创新型国家的过程中,科技成果的社会影响力评估已经被放在毋庸置疑的重要位置。2016年“科技创新三会”上,习近平总书记强调:“要改革科技评价制度,建立以科技创新质量、贡献、绩效为导向的分类评价体系,正确评价科技创新成果的科学价值、技术价值、经济价值、社会价值、文化价值。”这表明了对研究成果多维度影响(包括社会影响)的认可与重视,这也意味着科技评价的未来方向之一是科技成果的社会影响评价。在这种形势下,如何认识和评估科技成果的社会影响力值得所有科研人员和管理人员深思。

尽管中国学术界和实践界围绕科技成果评价开展了大量的探索^[4],但对科技成果社会影响力评估的关注还远远不够。科技成果社会影响力评估的缺失,进一步导致了只关注期刊影响因子,不求解决重大科技和社会问题的价值导向,严重影响了国家创新驱动发展战略的实施^[5]。鉴于此,本研究分析英国的卓越研究框架(REF)、荷兰的标准评估协议(SEP)和美国的STAR METRICS等典型的科技成果社会影响力评估方案,聚焦其关于科技成果社会影响力评估的内容与方法,旨在为中国未来开展相关理论和实践探索提供启示与建议。

1 西方国家科技成果社会影响力评估的典型方案

1.1 英国REF评估

卓越研究框架(REF)是2008年英国高等教育拨款委员会(Higher Education Funding Council for England, HEFCE)提出的高等学校科研水平评估框架,评估HEFCE拨款的投入产出结果与影响,鉴定科研资助机构投资的成效,以证明投资的合理性并向纳税人交账,同时也为改善管理或资金分配提供合理依据。

REF评估的前身是英国高等学校科研水平评估。20世纪80年代后期,英国在大学拨款中引入择优分配政策(policy of selective funding),为了改进科研经费拨款的合理性,采用了基于公式的拨款模式(formula

based funding model)——通过英国高等教育拨款委员会(HEFCE)组织的高等学校科研水平评估实施完成。自1986年开始,HEFCE每4~5年组织一次评估,迄今已经于1986、1989、1992、1996、2001和2008年共进行了6次评估活动。前2次被称为RSE评估(Research Selectivity Exercise),后4次被称为RAE评估(Research Assessment Exercise)。

与之前的评估方法相比,REF评估更加强调突出科学研究成果在学术价值以外的经济社会贡献,在原有的“科学产出质量”和“科研环境”指标上新增了“社会影响力”评估指标。2008年首次REF评估实践在英国国内和国际上都引起较大反响,“根据非学术影响决定研究经费分配是一个相对较为新颖的观念,REF是第一个在整个研究系统内进行应用的例子,所取得的结果值得深入研究。”^[6]2014年REF评估结束以后,HEFCE对其影响力评估结果已有一系列深入考察和研究,尤其对评估过程中暴露出的问题已有深入剖析以及相应的改进建议。在此基础上提出了REF 2020评估方案,并对社会影响评估的相关框架做了微调。经过多年的实践,REF在研究的社会影响评估方面所做的探索为科学的社会影响评估奠定了坚实的基础。

值得注意的是,英国学界对REF研究成果的社会影响力评估有很大争议,认为可能会阻碍“蓝天”研究的发展,让应用研究比基础研究更占优势,也让那些更容易展示影响力的研究占得先机,而来自各行各业的同行评估专家团的参与也可能会影响到学术自治。虽然REF的成功开展部分压制了这种声音,但英国学界目前对此项评议仍有较多争议。英国大学联合会(University and College Union)甚至曾发起过“站出来力挺科研”的请愿活动,反对REF的影响力评估,认为科学与其社会或经济应用之间存在清晰界限,应“抵制粗糙的应用概念”^[7],这一点还有待日后观察。

1.2 荷兰SEP评估

荷兰是具有较为完善的科学研究体系中系统性评估的欧洲大陆国家,2018年路透社发布的“欧洲最具创新力大学”百强榜单中,荷兰有9所大学入选。荷兰高效的科研管理制度和良好的科研环境是促成大学发展的重要因素,其中科研评估制度独具特色,对于不断反思和改进科研工作起着重要的引导和推动作用。

荷兰自20世纪90年代就开始对公共资助研究进行系统性评估,形成了标准评估协议(SEP)。SEP评估

包括外部评估和自评估两个方面,外部评估要求在自评估之后进行,接受外部同行专家的评审。评估标准和流程每6年修订一轮次,迄今完成了2003—2008年、2009—2015年两轮评估,逐步重视对公共资助研究“社会影响”的评估。在第一轮SEP评估时,“社会相关性”已被列为评估标准之一,但对研究社会影响及贡献的评估却一直未有合适的方法体系,并因此一直被科学界所诟病。为此,荷兰科学界进行了不懈的探索,并在2006年发起了“情境依赖的评估研究”(evaluating research in context, ERiC)项目。ERiC的目标是为评估建立标准、指标和方法体系,为完善第二轮SEP提供依据^[8]。其中,ERiC所设定的研究的“社会影响”评估框架影响了之后的SEP,在第二轮2009—2015的SEP中,“社会影响与活力”与“质量、产出、可行性”共同成为评估的4个主要标准。2014年,第三轮覆盖2015—2021年度的SEP由荷兰大学协会(The Association of Universities in the Netherlands, VSNU)、荷兰国家科学研究组织(The Netherlands Organization for Scientific Research, NWO)和荷兰皇家艺术和科学院(Royal Netherlands Academy of Arts and Sciences, KNAW)联合制定并授权发布,进一步将主要的评估标准凝练为“研究质量、社会相关性以及活力(策略)”,把科学研究的社会影响作为与科学研究质量同等重要的评估指标。第三轮SEP正在研究如何度量和报告“社会影响”,面临建立合理评估框架的挑战。

与英国科学界对影响力评估的反感不同,荷兰科学界目前似乎已经非常接受将社会影响作为其研究评估标准之一的现实。如前所述,2008年以各大学、KNAW和NWO下属研究所为代表的荷兰科学界曾主动提出SEP(2009—2015)评估应加强社会影响力(societal relevance)评估,定位以及标准设置。同时,为了更好地为社会影响评估建立标准、指标和方法体系,荷兰科学界进行了不懈的探索,并在2006年主动发起了“特定情境下的研究评估”(evaluating research in context, ERiC)项目。在与荷兰皇家艺术和科学院(KNAW)院长的访谈中,他认为荷兰大部分科学家已经接受了用社会影响和相关性衡量其研究产出的做法,目前荷兰在此方面的目标是如何更好地改进在此方面的评估质量,以更好地鼓励各学科的均衡发展。经过SEP评估的长期完善,很多学者认为荷兰已经拥有了较为成熟的评估模式,积累的经验值得在其他国家推广应用^[9]。笔

者认为在社会影响评估框架和体系建立起来,在学术界和社会之间达成充分共识之前,荷兰SEP所采取的“双向对话”以及“与拨款的弱关联”的模式,可能更容易被科学界所接受。

1.3 美国STAR METRICS评估

STAR METRICS计划开始于2009年,由白宫科技政策办公室指导国立卫生研究院(National Institutes of Health, NIH)和美国科学基金会(National Science Foundation, NSF)联合发起,目标是通过收集科研资助部门、科研管理部门现有的数据,搭建科研资助机构和大学、科技政策研究者共享的科研管理以及政策研究数据平台。在此基础上构建成熟的分析工具集合,来反映联邦科学投资在经济、社会以及科学上的成效。具体分析目标分为两个阶段:第一阶段重点关注科研投资如何创造就业岗位,第二阶段重点关注科研投资更广泛的影响。分析研究成果不与资金分配直接关联,而被用来接受社会问责、辅助资助战略决策(图1)。

STAR METRICS项目框架是由大学或科研资助机构自愿提出参与意愿,参与科研机构每季度接收一份来自STAR METRICS团队发回的本单位分析报告,主要以可视化图标的方式展示受资助科研工作情况。分析报告能够清晰、直观地反映联邦政府的科学投资是支持的谁、被支持的机构分布在什么位置,以及各科研机构在科研投资方面的其他相关信息。现阶段计划已覆盖美国大部分联邦政府投资的科研机构。

STAR METRICS的分析结果主要用以向大学董事会报告科学投资对当地就业影响的数据依据,以获得相应支持。这些信息还可改善信息问责和机构管理与绩效。在适当情况下,STAR METRICS团队也会对来自不同科研机构的汇总数据进行综合分析,并将结果向公众公开,以满足评价政府科研投资是否能够有效促进就业的要求。同时,科研机构也可利用这些信息去评估和改善本单位的绩效。

该框架出台后在国际上受到广泛关注并产生了很大影响。日本和欧洲已经加入了该项目。2011年4月,日本借鉴STAR METRICS提出了“发展科学、技术与创新政策方法学的倡议”(Initiative for Development of Science of Science, Technology and Innovation Policy);欧洲委员会受STAR METRICS启发也将类似方法应用于其科研评价工作之中进行了探索。

STAR METRICS框架也面临着巨大的挑战,最主要

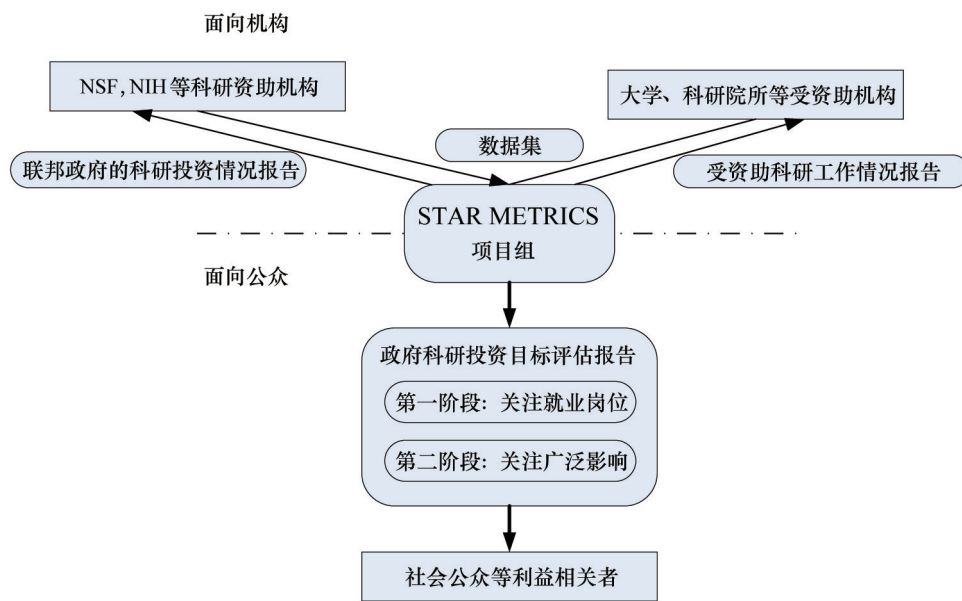


图1 STAR METRICS评估的基本程序

Fig. 1 Basic process of STAR METRICS evaluation

的问题是如何选择,并只选择那些可能能够反应研究与社会影响之间直接因果关系的数据。另外在实际应用中,围绕STAR METRICS的问题远比其能回答的多,如,联邦政府机构的数据库如何与公司和大学的数据汇集到一起、数据的取舍;“科学影响”的定义与准确区分、短期内的分析是否可靠等,但无论如何这代表了科学研究影响力测度的一个方向。在完成第一阶段的部分工作以后,2015年5月4日,STAR METRICS团队代表Sally Rockey在项目网站发布声明,宣布将该项目的所有资源迁移至正在建设中的联邦政府报告,这意味

着STAR METRICS项目自2016年1月1日起停止了数据收集工作,也意味着STAR METRICS并未有机会按计划开展第二阶段的科学投入社会影响数据的收集与分析工作。虽然政府更替与此有很大关系,但项目本身所面临的问题和挑战也表明该评估框架和方法在现实应用中遭遇困境。

本研究选取英国REF、荷兰SEP、美国STAR METRICS评估项目,分别从评估内容、程序、方式和目标方面,对上述评估项目进行比较分析,分析结果如表1所示。

表1 REF、SEP、STAR METRICS评估机制比较

Table 1 Comparison of REF, SEP, STAR METRICS evaluation mechanisms

评估对象		评估程序	评估目标
REF	高等学校科研水平	每4~5年为一周期,由HEFCE开展第三方评价	鉴定科研资助机构投资的成效,以证明投资的合理性并向纳税人交账,同时也为改善管理或资金分配提供合理依据
SEP	公共资助研究(可以是高校、科研机构、项目组等)	6年为一周期,先自评后外部评估	SEP评估大部分基于被评估单位所提供的自评报告,所提建议也只供被评估单位参考。SEP中各个被评估单位的打分之间并无有效的核准与审计
STAR METRICS	评估联邦政府投资的科研机构	建立多方共通支撑的开放式数据收集方法,成果分析和展示视角的数据挖掘和可视化	旨在通过收集科研资助部门、科研管理部门现有的数据,反映联邦科学投资在经济、社会以及科学上的成效——第一阶段重点关注科研投资如何创造就业岗位;第二阶段重点关注科研投资更广泛的影响

2 科技成果社会影响力评估指标与方法的比较分析

英国 REF、荷兰 SEP 与美国的 STAR METRICS 这 3 个框架在社会影响评估方面做出了有益的探索。以下将围绕“什么是科学的社会影响”和“如何评价科学的社会影响”两个问题,从科学社会影响的定义与指标设定,以及科学社会影响评价方法两方面进一步探讨。

1) 英国 REF 相关的社会影响评估。

英国的 REF 对影响力(impact)定义为:“主要指研究成果在学术以外,为经济、社会、文化、公共政策或服务、健康、生活环境或质量带来的影响、变化或益处”。REF 评估框架和提交指南(Assessment Framework and Guidance on Submission)中,列出了可以视为影响力的包括但不限于:任何地理位置(无论是本地、某个地区、国家或国际上)的观众、受益人、群体、选区、机构或个人在活动、态度、认知、行动能力、机遇、表现、政策、实践、过程或理解方面产生的效果、改变或效益,也包括减少或防止伤害、风险、成本或其他负面影响。

评估指标方面,REF 按照学科特征,将评估对象划分为 4 个主要类别,包括 A 组集中在生命与医学类;B 组主要是理工工程类;C 组集中在社科管理类;D 组则主要是人文艺术类学科。每类采用不同的评估指标,最大程度上保证评价标准在不同学科之间的适用性。其中影响力指标方面,生命与医学(A 组)、理工工程(B 组)和社科管理(C 组)的影响力指标包括健康和社会福利;社会、文化和创造力;经济和商业;公共政策和服务;生产;从业者和服务;环境;国际发展等。人文艺术类(D 组),影响力范围更为广泛,与前三类的差异也更明显,还涉及公民社会、教育事业、公共话语权等方面。

针对每一个影响力指标的测量指标,REF 还提出了可参考的指标池,供被评单位和评估专家在进行材料准备和评审时参考使用。如在健康和社会福利方面的建议参考的测量指标是临床结果、公共行为或健康服务得到改善的指标(如挽救生命,感染率降低);生活得到改善的指标;消费者对健康风险和利益意识增强的证据;对动物福利法规或指导方针变更的书面证据;临床实践中的相关证据;临床操作手册变更的书面证据;治疗或护理的成本降低等。

评价方法方面,REF 要求被评估人以个案叙述的方式列出研究的相关社会影响,并采用“个案研究

(case study)”方法进行评估。评估材料的设定上,REF 影响力评估主要基于高校或科研机构提供的两类材料影响力案例研究(REF3b)和影响力模板(REF3a),案例研究提供了研究产出影响力的案例,模板则是围绕便于将研究转化为影响力这一目标设置的更为广泛的战略。REF 主要是通过评价专家组对被评估单位递交的各种材料进行打分,并在不同学科组之间进行核准完成评估。具体是,评估将科研机构按学科分类,分别组成评价专家组,从研究产出质量(权重为 65%)、社会影响(权重为 20%)与研究环境(权重为 15%)3 个方面对被评估单位所递交的材料进行打分和排名,每份案例和模板都有不止一位评审人进行评估和打分。案例打分后,有一个不同分数的核准过程,不同大学科组之间也有一定的核准。除了核准以外,还有一个步骤就是审计,评估组成员可以要求对个案和模板进行审查,以核实是否符合打分的各项标准,也可以对案例中提到的证据源进行核查。审计并非在所有案例中系统开展,只是基于分评估组的要求。此次评估中,74%的高等教育机构(higher education institutions, HEI)有至少 5%的案例接受过审计^[9]。

Jonathan Grant 等借助文本挖掘分析(主题模型、关键词搜索和信息提取)工具,对 REF 的影响力评价方法效果进行研究,结果显示,英国高等教育机构研究成果的社会影响力广泛、深远,给人留下深刻印象,不同类型的高等教育机构往往在不同类型的影响力方面有专攻。影响力个案研究中使用标准化的信息清单和影响力定义,为未来分析提供了丰富的资料来源,表明个案研究是评估研究成果社会影响力的合适方式^[10]。

2) 荷兰 SEP 相关的社会影响评估。

荷兰 SEP 中关于研究的社会影响的评估中一直使用社会相关性(societal relevance)概念。SEP(2016—2021)的界定是,社会相关性指的是受政府资助的科学研究,在“针对特定的经济、社会或文化目标团体,政策方面的咨询报告,公共辩论”等方面贡献的质量、规模和相关性。SEP 将社会相关性评估指标具体分为 3 类(表 2)。

SEP 评估中,被评估单位需以案例叙述的方式列出其研究社会影响,内容也需包括案例、相关证据和在影响力方面所采取的战略。SEP 既对过去 6 年所取得的成果有回顾式评估,也有对未来在此方面战略的前瞻式评估。因此被评估单位在自评报告中须详细列出

表2 SEP的评估指标
Table 2 SEP evaluation criteria

类别	指标
可展示的产品	报告;针对非学术读者的专业期刊中的文章;针对社会目标群体的其他产出;科普活动(针对大众的讲座和展览等)
可展示的产品应用	专利/许可;社会各方对研究设施的使用;与社会各方合作的项目;合同研究等
可展示的获承认的标志	公众奖项、增值基金;社会各方支付的职位数量;民间团体机构的会员等

未来5~10年在社会影响方面拟采取的战略及举措。SEP评估大部分基于被评估单位所提供的自评报告,所提建议也只供被评估单位参考。SEP中各个被评估单位的打分之间并无有效的核准与审计。

相比之下,英国REF是按照统一标准来衡量所有被评估单位的全国性的综合评估框架,其框架和指标的设定更为详尽、复杂和专业。而荷兰SEP的评估因为基于被评估单位的自评报告,所开展的评估和结论更具针对性,但不同被评估单位之间可比性要弱化很多。这说明根据评估目标不同,影响力评估的内容与程序可以专门设定,不必拘泥于定势。

3) 美国STAR METRICS相关的社会影响评估。

美国STAR METRICS对社会影响的评估关注科学研究支持对创造工作岗位和经济发展的影响。由于《2009年美国复苏与再投资法案》中直接投资基础科学研究76亿美元,并承诺科学研究资金将创造工作机会,推动美国经济复苏。为了向公众展示科学研究资金如何对美国经济社会产生影响,美国政府提出了STAR METRICS项目。评估设立了“科学成果促进就业岗位创造”和“关注科学研究的广泛影响”两阶段目标,重点关注经济、社会、健康以及知识产出等方面的社会影响,基于数据平台构建“科技对经济效益影响”(如就业岗位创造)、“科学研究成效”(如通过引用和专利来测

度新科学的产生和接受),以及“科研投资的社会效益”(如公共健康方面)的测度方法。

评估方法方面,现有STAR METRICS评估采用科学计量方法。STAR METRICS评估计划响应2005年联邦政府提出的发展“科学政策的科学学”(science of science policy, SoSP)倡议,重点开展定量化、可视化和逻辑化的科技数据挖掘和展示方法,构建基于证据的联盟政府科技决策模式,为政府科技政策提供定量化支持。因而,STAR METRICS在评估方法上,一方面重视数据基础设施,着力于构建评估数据采集和处理平台,关注现有数据质量的有效性和数据提取的简易性;另一方面,STAR METRICS将开发和扩展现有的研究工具,综合利用国内或国际的已有各维度数据,把公共数据与机构数据库相捆绑,建立可扩展的测度体系以降低参与成本,满足利益相关人的期望,同时将资助机构和受资助方的负担降至最低。

综上,REF研究的社会影响评价主要用于政府未来的资助拨款决策;SEP主要为了帮助被评估单位提高科学研究质量,发现不足;STAR METRICS则帮助被评估单位更好地进行数据梳理,以向政府交账。分别从影响力界定、评估方法和指标三方面,对上述评估项目进行比较分析,分析结果如表3所示。

表3 REF、SEP、STAR METRICS的影响力界定、评估方法和评估指标比较

Table 3 Comparison of the definition of influence, the evaluation mechanism, and evaluation criteria of REF, SEP, and STAR METRICS

	影响力界定	方法	评估指标
REF	政府资助的研究成果,在学术以外,为经济、社会、文化、公共政策或服务、健康、生活环境或质量带来的影响、变化或益处	个案研究方法	生命与医学(A组)、理工工程(B组)和社科管理(C组);健康和社会福利;社会、文化和创造力;经济和商业;公共政策和服务,生产,从业者和环境,国际发展等 人文艺术类(D组),影响力范围更为广泛,与前三类的差异也更明显,还涉及公民社会、教育事业、公共话语权等方面
SEP	受政府资助的科学研究,在“针对特定的经济、社会或文化目标团体,政策方面的咨询报告,公共辩论”等方面贡献的质量、规模和相关性	案例研究方法	可展示的产品 可展示的产品应用 可展示的获承认的标志
STAR METRICS	评估联邦政府投资的科学研究资金如何对美国经济社会产生影响	科学计量方法	科技对经济效益影响 科学研究成效 科研投资的社会效益

3 中国开展科技成果社会影响力评估的建议

从全球发展趋势看,开展科技成果社会影响力评估势在必行。多元化的科研活动产生形式多样化的科研成果,也会产生不尽相同的社会影响力,这对科技成果社会影响力的评估方法和标准体系提出了挑战。从上述国际经验中,可以得到如下启示与建议。

首先,将科技成果社会影响力评估置于公立研究机构绩效评价的核心。当前,无论是从公共研发资金绩效管理角度,还是从回应社会公众关切角度,都离不开对科技成果社会影响力的合理评估。中国应当将科技成果的社会影响力评估作为破除量化评价导向、深化科技评价制度改革的突破口。具体可借鉴 REF 和 SEP 的做法,将社会影响力作为高校和科研机构基础研究绩效评价的重要指标,实施周期性评估,并据此投入公共研发经费和其他公共科研资源。

其次,开发符合中国国情的科技成果社会影响力评估内容与指标体系。为了真正推动中国开展科技成果的社会影响力评估,必须探索符合中国国情的评估内容与指标体系。在评估内容和指标体系设计方面,首先应当贯彻国家重大战略导向,将科技活动服务国家重大战略作为评判其社会影响力的首要标准。在此基础上,要体现三方面的差异。一是评价对象的差异。一方面,正如 REF 的做法,不同学科科技成果社会影响力的体现有所不同,应当设置不同的指标。另一方面,不同层次的高校和科研机构定位不同,在社会影响力评估时也需要区别对待。二是区域差异。不同区域间经济社会发展不平衡是中国的重要国情,也是社会影响力评估内容和指标设计的基本出发点。三是评估结果用途的差异。科技成果社会影响力评估可以用在科技资源投入、宣传展示、绩效考核等多个方面。根据评估结果用途的不同,评估指标和内容也应当有所侧重。美国 STAR METRICS 的评估指标具有极大的灵活性,可以为不同的评估目的所服务,这为中国科技成果社会影响力的评估提供了有益的启示。

三是搭建大数据平台,使用数据驱动的社会影响力评估方法。科技成果的社会影响力评估涉及内容多,难以使用传统的评估方法。大数据等新一代信息技术的发展为科技成果社会影响力评估提供了新的思路,这也正是美国 STAR METRICS 的发展方向与重要

启示。一方面,中国需要构建一套类似于 STAR METRICS 的基于证据的科技决策系统,建立跨部门的科研活动全过程数据平台,为科技成果社会影响力的评估提供数据基础。另一方面,需要积极采纳新的技术手段,使用数据驱动的社会影响力评估方法。一是借助数据挖掘、数据整合、自然语言处理、机器学习等技术,结合使用正式数据和图像、用户生成内容等网络数据开展评价。二是使用各类工具手段和数据分析模型,使得评估结果可视化、精细化和智能化。

参考文献(References)

- [1] 徐芳,刘文斌,李晓轩. 英国 REF 科研影响力评价的方法及启示[J]. 科学学与科学技术管理, 2014, 35(7): 9-15.
Xu Fang, Liu Wenbin, Li Xiaoxuan. The method and enlightenment of REF scientific research impact assessment in UK[J]. Science of Science and Management of S. & T., 2014, 35(7): 9-15.
- [2] VSNU, KNAW and NWO. Standard evaluation protocol 2015—2021: Protocol for research assessments in the Netherlands[EB/OL]. (2014-03-21)[2019-03-14] https://www.knaw.nl/en/news/publications/standard-evaluation-protocol-2015-2013-2021?set_language=en.
- [3] Australian Research Council. Australian research council corporate plan 2015-16 to 2018-19[R]. Canberra: Commonwealth of Australia, 2015.
- [4] 汪雪锋,张硕,刘玉琴,等. 中国科技评价研究 40 年:历史演进及主题演化[J]. 科学学与科学技术管理, 2018, 39(12): 67-80.
Wang Xuefeng, Zhang Shuo, Liu yuqin, et al. Forty years of research on science and technology evaluation in china: Historical and theme evolution[J]. Science of Science and Management of S. & T., 2018, 39(12): 67-80.
- [5] 陈兆莹. 关于科技评价和科技评价改革的讨论[J]. 科学通报, 2018, 63(7): 611-617.
Chen Zhaoying. Reflections on science and technology evaluation in China[J]. Chinese Science Bulletin, 2018, 63(7): 611-617.
- [6] David C, Ian M C, Molly M J, et al. Making the Grade: Methodologies for assessing and evidencing research impact[EB/OL]. (2013-01-01)[2019-03-15]. https://www.exeter.ac.uk/media/universityofexeter/research/ourresearchexcellence/describeproject/pdfs/2013_06_04_7_Essays_on_Impact_FINAL.pdf.
- [7] Stand Up for Research of University and College Union[EB/OL]. (2009-10-16)[2019-03-12]. <http://www.ucu.org.uk/events>.
- [8] Meulen B V D, Daemen F, Drooge L V, et al. Evaluating Re-

- search in Context: Pilot Study at Faculty of Architecture TU Delft[EB/OL]. (2010-03-01)[2019-03-15]. <https://repository.tudelft.nl/islandora/object/uuid:f0a713f1-1564-4b79-be66-4f5299ebba2c?collection=research>.
- [9] Van d M B, Rip A. Evaluation of societal quality of public sector research in the Netherlands[J]. *Research Evaluation*, 2000, 9(1): 11-25.
- [10] Catriona M, Molly M J, Michael F. et al. Preparing impact submissions for REF 2014: An evaluation[EB/OL]. (2015-03-25)[2019-03-15]. https://www.rand.org/pubs/research_briefs/RB9833.html.
- [11] King's College London and Digital Science. The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies [J]. *Revista Da Sociedade Brasileira De Medicina Tropical* 2015, 41(1): 27-33.

Social impact assessment of science achievements based on international experiences

FAN Yingjie^{1,2,3}, XU Fang^{1,3*}

1. School of Public Policy and Management of University of Chinese Academy of Sciences, Beijing 100090, China
2. Bureau of International Cooperation, National Natural Science Foundation of China, Beijing 100080, China
3. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China

Abstract The evaluation of the social impact of science achievements funded by the government can be considered as a breakthrough point in the reform of China's science and technology system. Besides, it is also an inevitable step to carry out the strategy of China's innovation-driven development. The lessons drawn from the evaluation activities in western countries can be a useful guidance. The paper focuses on three typical systems, the UK's Research Excellence Framework (REF) for assessing the excellence of researches in higher education institutions, the Standard Evaluation Protocol (SEP), which is used to assess researches conducted by Dutch universities and NWO and academy institutes, and the USA's STAR METRICS for creating a repository of data and tools, that will be useful to assess the impact of federal R&D investments. This paper analyzes the evaluation unit, the processes, and the objectives of the above three systems, as well as their differences, focusing on the indicators that define the social impact and the methods to realize the evaluation. Based on this analysis, it is concluded that the social impact evaluation should be at the core of the performance evaluation of public scientific research institutions in China. Following the experience of the UK's REF system, the national strategies, the differences between institutions and disciplines, as well as the regional divergence should be taken into account when developing the indicators of social impacts and the corresponding measurements in China. To facilitate the system operation, it is necessary to create a searchable database of scientific researches funded by national agencies.

Keywords assessment of social impact of science achievements; evaluation content; evaluation methods ●



(责任编辑 傅雪)