

大数据时代 数据开放

国际协定
缩略版

《协定》是“科学国际 2015”会议的成果。“科学国际 2015”是全球四大顶级科学组织年度系列会议的首次会议，这些组织是：国际科学理事会——ICSU、国际科学院合作组织——IAP、世界科学院——TWAS、国际社会科学理事会——ISSC。通过这些系列会议，四大组织代表全球科学界探讨科学领域国际政策。

《协定》确定数据革命的机遇和挑战是当前全球科学政策的主要议题、并提出了应该采纳以应对这些挑战的基本原则。各国政府和政府间机构把数据开放作为维持科研严谨性、为发达和发展中国家最大化数据革命带来的公共利益的根本前提。《协定》代表科学界就此发出了独特的声音。

参加“科学国际 2015”的合作伙伴将从国家和国际层面，推进各自成员和其它科学代表机构对这些原则进行讨论、采纳并给予支持。

《协定（扩展版）》提供了更深层次的依据，以及用于支持开发数据开放系统和方法的具体实践。

1. 大数据时代

最近几十年的数字化革命是一起世界性历史事件，其影响之深堪比印刷术的发明、且影响更为广泛。人们获取、存储、处理和即时传送庞大和复杂数据的能力前所未有地急剧上升，对科学产生了深远的影响¹，改变速度令人生畏。2003年，科学家历时10年、花费10亿美元，才完成人类基因组图谱。如今，这一过程只需短短几天和少量经费（1000美元）²。“大数据”（高通量的数据流出于计算系统）和“广泛数据”（大量数据集通过语义关联创造更深层意义）是这场革命的引擎，为自然、社会和人文科学带来新的机遇。

2. 机遇

数据富集时代的科学机遇在于：发现那些曾经遥不可及的新模式；连接和关联系统的不同方面，从而更好理解其行为；描述复杂性；通过对复杂系统状态进行迭代描述和模拟来预测其动态行为。

与此相关的研究领域众多：天气和气候预测；了解大脑机制；全球经济行为；评价农业生产率；人口预测；阐明历史；以及众多当代全球挑战，如环境变化、传染病、大规模移民等，都需要融汇多学科的观点和数据。

3. 挑战

抓住这些机遇对科学的工作和组织方式提出了严重挑战。数据开放是共通的使能思维。

数据开放势在必行

公共资助型研究的基本作用在于增加对人类判断、创新、社会福利及个人幸福至关重要的知识和理解。通过允许数据和思想在各方网络互动中公开且迅速广泛地流动，数字化革命的技术和过程提供了一个增强科学生产力和创造力的强大媒介。如果要实现这场科学的社会化革命，那么我们必须采纳一种默认的基本立场：当一个涉及数据收集的研究项目完成时，源自公共财政支持所产生的数据应可公开访问及再利用。

保持“自我校验”

科学结论的证据（数据）公开是科学进步的基石。它使得论证逻辑得以检验，观测或实验的可重复性得以验证，并由此支持或否定这些观点。当文章发表其科学结论时，证据性数据、支持数据再分析的相关元数据、以及用于计算机数据处理的代码应同时公开用于检验，从而保证自我校验这一重要过程。近期对多个学科已出版文章结论的检测，证实了存在较高的不可重复性，这凸显了大数据时代强化数据开放过程的迫切需要。然而仅仅开放还不够，数据须智能化开放，也即它们应当：可发现、可访问、可理解、可评价、可（再）利用。

调整科学推理

许多我们试图通过大（关联）数据和广泛（关联）数据捕捉的众多复杂关系，是经典统计方法的分析能力难有凑效的。它们要求更深层次的数学方法（包括拓扑方法）来保证从大数据和广泛数据得出有效推论。数据密集型机器分析和机器学习正变得无处不在，对科学发现意义重大。模式的复杂性能被机器识别却不易被人类认知过程掌握，由此提出关于人机界面的深刻问题，以及怎样才算21世纪的一名合格的研究人员？

道德制约

开放数据原则对研究人员和研究主体具有伦理意蕴。由于它要求超越研究者（数据创建者）的个人利益，新的贡献认可方式亟待制定。数据主体的隐私需受到保护。在开放共享的体制下，数据将自由传播，未来的使用可能失去控制，而匿名化程序已被证实无法保证个人记录的安全。

开放的全球参与

大数据和开放数据有巨大的潜力使不太富裕的国家受益，特别是最不发达国家（LDC）。然而，LDC的国家研究体系通常资源贫乏。如果他们不能参与基于大数据和开放数据的研究，他们与其他国家的差距将在未来几年成倍增长。他们将无法收集、存储和分享数据，无法参与全球研究计划，无法作为全面合作伙伴在气候变化、医疗保健和资源保护方面贡献力量，且无法充分受益于这些工作。没有全球参与，全球解决方案将化为泡影。因此，新兴国家和发达国家都有明确且直接的利益——帮助充分调动LDC的科学潜力，从而为完成联合国可持续发展目标做出贡献。

¹ “科学”指可合理解释、可靠应用的知识的系统化组织。如同英语之外的其它语言，“科学”涵括了所有领域，包括人文、社科，以及STEM学科（科学、技术、工程、医学）。

抓住机遇

有效的数据开放要求个人、学科、国家和国际层面上的全体行动。尽管科学是一项国际性事业，却完成于各国各自的责任、组织和管理体系，各体系都需应对数据开放的机遇。研究实施机构或资助方应资助设立相关流程，为致力于数据智能开放的研究者减轻负担，支持数据开放过程。

越来越多的研究团体发现了数据共享的好处，其中包括语言学、生物信息学、化学结晶学；通过一些促进开放数据访问和使用的国际合作，他们为其学科取得了重大进展。

此外，国际机构——如国际科学理事会科技数据委员会（CODATA）及世界数据系统（WDS）、研究数据联盟（RDA）——也有责任推进和支持开发相关系统和流程，实现国际数据访问、互操作性和可持续性。

开放科学和公共知识

科学界需要与社会更紧密地对话、更广泛地寻求社会参与，以便通过对问题的交互式建构、对研究课题的协同设计、实施和应用来解决当前的众多问题。正是意识到这点，“开放科学”得以发展。当然，开放总是有合理的限制的，比如，我们需通过恰当的法律机制保护安全、隐私和所有权问题。再者，目前存在相反的知识私有化趋势，这有悖科学探究精神及自由使用思想的基本需求。如果科学事业要避免被此类压力带入误区，那么全球科学界需申明其恪守数据开放、信息开放、知识开放的原则。

4. 数据开放原则

考虑到数据革命对科学实践带来的重大挑战，“科学国际2015”参加者认为，应推进以下数据公开原则：

责任

科学工作者

i. 得到公共资助的科学工作者有义务通过创造和传播新知识来为公共谋福利，而受资助项目产生的数据是知识的固有成分。在数据被生产后，科学工作者应尽快开放此类数据，并使其可再利用。

ii. 为已发表科学结论提供证据的数据，应以智能化的开放格式与科学结论同时公开。如此，数据和结论之间的逻辑可得以严格验证，数据有效性也能通过重复实验或观测得到检验。数据应尽可能存储于管理良好、受信赖且无访问障碍的存储库。

iii. 研究机构和大学

有义务创造一个数据开放的支持性环境，包括提供数据管理、保存和分析的培训，提供相关技术支持（包括图书馆服务及数据管理服务）。雇佣研究人员的研究机构、以及资助他们的资助机构，应为参与数据开放过程的人员制定有助于其职业发展的激励措施和标准。此类标准需基于全国共识、理想情况下应

基于国际共识，以促成理想的研究人员流动模式。顺应当前国际化的时代精神，发达国家的大学及科学机构应与发展中国家合作，推动数据密集型能力。

iv. 出版商

有义务保证评审过程中审稿人可获得相关数据，应要求数据和对应的文章同时实现智能化开放式访问，应要求完整地指向和引用这些数据。出版商还有义务通过公开元数据、开放获取文本、支持数据挖掘等，来支持科学记录开放用于后续分析。

v. 资助机构

应将数据开放过程的花费视为研究经费的一部分，应为基础设施和存储库的长期可持续性提供足够的资源和政策支持。对科研影响力的评估，尤其涉及到论文引用指标时，应考虑数据创造者的贡献。

vi. 专业协会、学术团体和机构

应制定数据开放的指南和政策，并通过各种机会进行宣传和推广，影响其成员的认知规范和实践。

vii. 图书馆、档案馆和存储库

有义务开发和提供数据服务及相关技术标准，确保数据可用、且能长期访问。

开放的边界

viii. 数据开放应成为公共资助的科学研究的默认模式。例外应仅限于以下情况：隐私、安全、治安及事关公众利益的商业用途。申请例外时应针对具体情况个案处理，避免简单归类进行例外豁免。

达成实践

ix. 引用和溯源

当研究人员在学术出版物中使用他人创建的数据时，数据应以参考文献的方式进行引用，注明创建者、出处以及永久数字标识。

x. 互操作性

研究数据和（允许数据被评估、再利用的）元数据，应尽可能具有最大互操作性。

xi. 非限制性再利用

如果某研究数据尚未进入公共领域，应通过一定的权利豁免或非限制性许可将其标识为可再利用，以明确数据在再利用时只需致谢其生产者。

xii. 可关联性

开放数据应按其内容和产生背景尽可能与其他数据关联，以最大限度提高其语义价值。

本文件由 ICSU-IAP-ISSC-TWAS 工作组完成，人员包括：

- Geoffrey Boulton,
爱丁堡大学, CODATA 主席, 工作组组长
- Dominique Babini,
布宜诺斯艾利斯大学 & 拉丁美洲社会科学委员会 (ISSC 代表)
- Simon Hodson,
CODATA 执行主任 (ICSU 代表)
- Jianhui Li,
中国科学院计算机网络信息中心 (IAP 代表)
- Tshilidzi Marwala,
约翰内斯堡大学 (TWAS 代表)
- Maria G. N. Musoke,
马卡雷雷大学, 乌干达 (IAP 代表)
- Paul F. Uhler,
学者, 美国国家科学院 (IAP 代表); 独立顾问, 数据政策与管理
- Sally Wyatt,
马斯特里赫特大学 & 荷兰皇家科学院 eHumanities (ISSC 代表)

《协定（扩展版）》见：

<http://www.science-international.org>

纸质版可通过以下地址获得：

International Council for Science (ICSU),
5 rue Auguste Vacquerie, 75116 Paris, France.



www.icsu.org
www.interacademies.net
www.worldsocialscience.org
www.twas.org