

实现欧洲开放科学云

【译者按】 欧洲是世界上最大的科学数据产出地，但由于基础设施不足、研究力量碎片化，导致大量科学数据未被充分利用。当前，欧洲正在积极建立开放科学云（EOSC），旨在整合现有的数字化基础设施、科研基础设施，为欧洲研究人员和全球科研合作者提供共享开放的科学云服务。2016年10月，欧盟委员会发布《实现欧洲开放科学云》报告，分析了实现EOSC面临的形势和发展EOSC的关键因素，并针对政策、治理和实施提出了对策建议。赛迪智库信息化中心对该报告进行了编译，希望供我国相关部门参考。

【关键词】 欧洲 开放科学云

欧洲开放科学云（EOSC）的理念是建立一个集聚各类研究、创新与教育的数据和工具，可供研究人员、创新人员、企业以及公民发布、搜索和反复使用的全球共享云环境，确保欧洲研究与创新形成知识积累，为破解全球难题、推动欧洲经济繁荣提供支持。EOSC 建设内涵是：一是旨在消除使用研究数据和工具过程中出现的技术、法律和人员障碍，为跨学科、跨社区和跨地区访问服务、系统与数据流提供支持，助力欧洲数字化单一市场发展。二是 EOSC 涵盖了人类知识、科研资源、标准、最佳操作规范和底层基础设施等内容建设，在支持开放数据的搜索、访问、互操作和反复使用的同时，确保敏感和保密数据的安全。三是 EOSC 的发展还需与全球建立专业的数据监管机制，确保 EOSC 成为欧洲科学与创新的前沿。

一、发展欧洲开放科学云的重要意义

传统的科学研究评估、筛选、资助和奖励体系都建立在数据规模较少的年代，能对理论和新工具的研究提供有效支持和激励。但是，面对当前以数据发布、工具共享、知识开发以及数据管理为核心的科学研究方式，该体系的支持和激励效果就微乎其微了。因此，有必要建立一种新的体系框架或依据，为开展科研活动提供有效支持。

在科研领域,受数据采集和使用方法不当等原因影响,约 50% 的研究数据与实验无法复制,而大部分(可能超过 80%)的数据未能有效、长久地存储起来。据预测,2014 年至 2020 年间,欧洲将在数据生成研究上投入 1200 亿欧元,而每年的资本消耗将十分惊人,这还未包括低效数据分析所带来的损失、创新停滞,以及创新知识不适用于社会造成的经济影响。如果能实现全球互联网数据与服务可以平等的反复利用,以及建立良好的数据监管等机制,将显著减少上述损失,有助于推动科学创新。

欧洲拥有相对健康的研究基础设施,包括特定领域的研究基础设施、跨领域的信息通信基础设施以及跨学科的合作和服务,许多成员国还制定了为科学研究开放数据访问和使用的基础设施建设和使用计划。这些都为推动 EOSC 建设奠定了重要基础。

在此背景下,面向所有数字的研究和合作,要想实现基础设施及其数据和服务的无缝访问和反复使用,特别是公共与私有领域相关利益者都能参与,就必须进行一次巨大的、根本性变革。这要求通过科学研究领域和信息通信技术领域之间深入、平等地合作,培养形成新的思维方式,使所有知识创新都能推动 EOSC 发展,并从 EOSC 获得受益。正如每个时代所经历的社会结构和基础设施变革一样,欧盟正在向开放科学时代迈进,EOSC 是实现并支撑这一变革的重要设施。

二、发展欧洲开放科学云面临的形势和挑战

（一）面临形势

对数据专家的依赖程度提升。这一点在最低估数据专家的学术界尤为显著。新的数据和软件发布形式正在形成，缺少核心的专业数据知识可能会使欧洲失去在科学领域的领导地位。如果欧洲要向开放型、数据型科学转型，那么就需要在研究评估中对数据和软件发布形式予以信任，并将其纳入到推广决策之中。

跨学科合作日益密切。随着科学家越来越多地使用来自其他学科的原始数据资源、受监管数据和分析工具，跨学科合作已变得至关重要。但目前，任何人要想获得其他人的数据都存在困难，即便是在自己的学科领域内也不例外，从其他学科找到相关数据更是难上加难。在缺少合适的元数据标准、数据工具和搜索引擎的情况下，研究人员也是“巧妇难为无米之炊”。比如，医疗研究人员想要使用来自社交媒体和“自我量化”（个体通过设备自己记录）的数据，但无法在术语、数据格式和元数据标准完全不同的情况下判断出哪些是有价值的数据集。这类问题在跨学科数据上被进一步放大。因此，文本转换和数据挖掘技术对于 EOSC 建设、研究分析和支持跨学科使用而言必不可少，但这方面常常会受到法律制约。在欧洲，这类问题已成为委员会讨论的主题之

一，并且在 2016 年可能会将文本转换和数据挖掘涉及事宜列为法律除外事项。

从科学研究向创新转化的进程明显加快。虽然还远远未达到最佳状态，但知识发现的速度已超出了社会转换与创新能力。欧洲社会最具创新力的中小型企业相对较弱，始终制约着创新发展。尽管支持跨学科研究和创新项目，以及建立公共或私立联盟的政策文件已经多到来不及解读，但在实际操作过程中，欧洲融资与审批环境过于苛刻，严重阻碍了重要的跨学科研究和创新合作项目的发展。

基础设施生态系统复杂程度不断提升。随着新型计算和存储技术以及高级数据算法的迅猛发展，使得仅仅靠扩大数据基础设施，无法有效解决数据规模不断扩大所带来的挑战，因此，欧洲需要好好重新思考数据基础设施生态系统建设问题。越来越多的学科数据增长速度超出了数据转换速度，使得综合数据集规模不断扩张，导致数据无法有效地从一个地点向另一个地点转移，并且在许多情况下，数据隐私与保密性的有关法律条款严禁数据随意移动到数据采集环境之外。因此，包含并行和分布式计算的轻量化工作流程（比如处理虚拟机）越来越倾向于在数据所在地进行访问，并只从安全数据库中带走计算结果，这就是一个轻量化计算容器访问数据（反之亦然）服务设施的早期形态。这种方法

将释放巨大的分布式计算潜力，但需要大幅提升硬件设施的高性能计算能力，集中化的数据中心是解决高性能计算问题的关键，但还需要补充更多的支撑基础设施。

机器理解变得必不可少。许多数据集的规模庞大且复杂，只能由强大的计算机进行处理，并发现其中可以被人类用户获取的可用知识。机器在数据生成、处理和分析方面已变得必不可少。为了帮助机器理解人类语言，以及人类通过机器处理获取知识，数据格式、术语/标识和来源都必须得到规范设置，但支持这两项流程的工具存在根本性的差异：模式识别工具主要用于机器，而认知解读的工具主要用于人类。因此，无论发布的是什么样的研究成果（论文或数据），是否能够通过机器执行都是需要考量的重要因素。

诚信研究的重要性显著提升。目前的科学研究严重缺乏可复制性，再加上科学欺诈，对科学界的声誉造成了巨大的破坏。造成这一现象的原因之一是如何在有效呈现数据、相关方法和工具的同时，又允许他人复制结果方面缺少深厚、严谨的理论基础。一点细微区别，都可能造成文献中提到的关于不能在任意地点轻易复制结果的结论成立。随着数据监管以及核心数据处理和分析专业知识的加强，可复制性和欺诈信号的提前检测能力将会增强，而研究资产的反复使用也会得到优化。

（二）主要挑战

EOSC 建设面临的最主要的挑战是跨学科数据分析的复杂性，而非数据规模本身，这种挑战来自社会而非技术。主要包括以下几点：

一是数据专家严重缺口是全球面临的共同难题。数据专家既不是指从事计算机研究的科学家，也不是指纯正的数据或计算机科学或软件领域的工程师，而是指技术型数据专家——在所从事的内容分析领域足够专业，能够从实验和提案编写阶段到数据发现周期结束，从始至终为研究团队提供日常专业咨询。这类专家会确保优秀的数据管理计划成为优秀研究规范的必要组成部分（包括数据反复使用、监管和合理的预算制度），帮助制定合理采集新数据的规范或要求（数据格式、元数据规模、数据标准、采集来源、发布形式、链接和分析等），并为数据分析提供充分的技术支持。

二是陈旧的科学创新奖励与资助体系，始终要求提供纸质文件材料而非电子材料，并阻止有效数据的发布和多次使用。

三是在科学领域和信息化领域之间开展沟通交流的专业人员的缺乏，使得信息化基础设施提供者和科学专家间无法有效对接，形成鸿沟，导致各领域各自研究发展碎片化的数据开放解决方案，彼此孤立且出现越来越多的重复建设。

四是短期且相互分散的核心研究和信息化基础设施建设资助计划，并不足以支撑并实现全球科学数据的监管和有效使用这一宏大目标。且受制于数据规模和隐私等因素，大型分布式数据集越来越难以移动，仅凭集中式高性能计算不足以支持高度联合而分散的元分析与学习。

五是虽然创建第一代 EOSC 的要素大部分都已具备，却分散在 28 个成员国和各个社区团体中，而目前缺乏专门的强制措施和工具来协调推动各成员国间的 EOSC 合作共建行为。

（三）要突破的关键因素

EOSC 作为开放科学的有效组成部分，要实现有效发展，关键因素包括：需要实施新的学术交流模式，重点是确保机器的可操作性。推行现代化的奖励与酬劳机制，以支持数据共享和多次利用。培养核心数据专家，为其提供更好的职业前景。设立创新、定制的资助方案，支持支撑性基础设施与核心资源的可持续发展。真正激发多学科合作，针对审核、资金与基础设施采取具体的措施。从科学观察向创新转化需要采取专门的支持政策。EOSC 应成为一个数据基础设施公地，即发展基础设施的生态系统。在可能的情况下，EOSC 应实现数据处理的自动化，机器可操作性是关键。推行轻量化并在国际上行之有效的治理机制。开发并制定面向 EOSC 的关键性能指标。

三、发展欧洲开放科学云的对策建议

(一) 政策支持

与成员国密切配合，采取及时、积极的行动。为了让欧洲拥有一个先进、繁荣的研究与创新环境，各成员国必须通过当前工具开展全球合作，共同采取及时、有力的积极行动，建立 EOSC 这一全球可以访问的初级合作环境，使研究人员、创新者、企业和公民都能够在一个定义清晰、安全和可信的环境中，采用可持续、有价值的模式发布、搜索和彼此反复使用的数据和工具，开展研究、创新与教育。

停止关于“感知需求”的讨论。尽管 EOSC 建设需要涉及所有的欧洲科研基础设施战略论坛、网络基础设施、私营行业和来自各成员国及其他地区的利益相关方，正如欧洲科研基础设施战略论坛一样，要就建立格局和达成共识进行频繁接洽和准备，但准备阶段不应太长。对于 EOSC 建设而言，当前关于“感知需求”的谈判过程已经过于漫长，普遍的看法也都认为迫切需要尽快启动建设。

尽可能建立在现有能力和专业知识的基础之上。建立 EOSC 的提议书和其他议题中都明确表示，绝大多数人都认为 EOSC 的要素都已存在并且具有很高的质量，但问题在于它们隐藏在碎片

当中。因此，主要挑战是如何破解文化差异方面造成的制约。目前来看，EOSC 准备阶段所提出的大部分行动都能在 2017 年末前取得巨大进展，甚至完成。

将 EOSC 界定为欧盟针对基于开放协议的互联网数据与服务公平所做出的贡献。当前互联网的广泛应用避免了极少数个人或公共参与者的垄断统治，其严格标准与协议最大自由的实施了“沙漏模式”。欧洲应在实施基于最低程度严格标准的 EOSC 时采用相似的解决方案，通过开放的方式推动共同实施，从而避免在决定谁的解决方案最好这一过程上耗时伤财。同时，所有利益相关者都能参与，包括研究基础设施提供者、成员国、研究机构和企业等。所有公共与私人提供者都可以在最低参与标准和最低参与规则发布之日起，开始实施这一具有互联网数据与服务公平原型的模式。

（二）强化治理

以最轻量化的全球有效治理为目标。鉴于实现欧洲开放科学云的迫切性以及利益相关者和参与者的数量和类型众多，EOSC 取得成功的合适模式并不是在某一地点建立一个管理严格的新基础设施，而是需要一个更加包容、灵活、透明和分散，并且能够实现有效合作的解决方案。欧盟委员会需要为 EOSC 建立一个轻量化的可持续合作模式，使所有参与者都能做出贡献。

仅针对需要指导的方面进行指导，这涉及技术问题、最佳时间与社会变革。虽然提倡轻量化治理，但仍需要一定程度的监管。如果没有监管制度，那么如同当今的互联网一样，一些大型公共和私人参与者可能在 EOSC 中占据不公平且适得其反的份额。EOSC 将会包含各类小型和超大型参与者，但监管者和股东应对他们一视同仁，使公民、研究者和创新者能够在—个值得信任的可持续环境中使用彼此的数据和工具。欧洲应在互联网数据与服务公平的合理指导方面起带头作用。

针对 EOSC 中提供的服务制定参与规则。为了支持广泛的参与、创新和可持续发展，EOSC 需要向所有公共和私人参与者、欧洲和非欧洲参与者开放，并且应当规定利益相关方的参与条件：需要通过一套严格实施的最低标准协议，用以指导和管理相关基础设施的开发。这套参与规则可以用于标记 EOSC 中的提供者是否值得信任，并应遵守参与规则。需要明确的是，非欧洲开放科学云批准的参与者也能够在开放科学生态系统中根据自己的意愿自由选择任何服务角色，但他们不能将自己的服务角色标注为是通过 EOSC 批准/认证的。

充分整合优秀资源，联合重要相关方并扩大优秀实践。基于互联网数据与服务公平的大部分基础设施已在一些地方正常运作，但分散在各个领域、地区和资助项目中，欧洲应尽早采取强

有力的行动整合优秀资源。信息通信基础设施社区、欧洲科研基础设施战略论坛社区和其他学科组与协会也需要明确自身最佳的参与模式。这些跨欧洲科研基础设施战略论坛组成部分中，有些已在各成员国中开始运行。所有遵守标准和签署参与规则的合作伙伴和利益相关方都应有资格参与。

（三）组织实施

1、将高级专家组¹研究报告转换为指导文件，从而明确 EOSC 计划的方向并为之提供指导

研究报告可以成为各成员国和欧盟 2020 地平线（Horizon 2020）实际发展与实施的指导文件，以及进一步建立国际共识和合作的讨论基础。

2、制定、签署和实施 EOSC 的参与规则

欧盟委员会应与各成员国对应的利益相关者密切合作，将为任何希望加入 EOSC 的参与者制定参与规则放在第一位。参与规则应基于 EOSC 中的所有数据（即所有研究对象）是公平的（除了检查是否使研究对象可以被搜索、访问、互操作和反复使用之外，公平原则不强制规定具体的实施）这一假设。欧洲开放科学云高级专家组应尽快建立和指导一个专门小组，在 2016 年起草

¹ 译者注：欧盟开放科学云高级专家组（High Level Expert Group of the European open science cloud, HLEGEOSC）是欧盟云计算战略推动实施后，组建的专门研究开放科学云政策措施的战略咨询专家团队，主要职责包括：清除各类障碍、统一标准规范、理顺各方利益、高速宽带互联互通、综合管理软件开发和联合技术项目攻关等。由欧盟 2020 地平线（H2020）提供必要的资金支持。

参与规则。随着 EOSC 的发展，遵守参与规则应作为 EOSC 中分享和提供网络资源和服务的认证标志。在任何情况下，都应分阶段执行，并且应该及时制定最低程度的参与规则，推动短期目标向中期和长期的可持续发展目标迈进。

制定指导原则，尽快启动该计划，初期准备阶段应采用以下指导原则：

——EOSC 将建立在以社区为基础的轻量化可持续治理体制之上；

——EOSC 将尽可能支持现有的优秀运行机制；

——EOSC 将支持科学家和创新者，并将由用户带动发展；

——虽然需要开展合作，但这并不意味着由科学家和创新者向开发者下达建设命令，而是应该让数据专家和工程师贡献各自的知识与专业技术，了解哪些可以进行灵活开发：一方面，科学家和创新者需要统一意见，向开发者清晰、一致地表达需求。另一方面，科学家和创新者需要致力于灵活开发，通过定期反馈和用户测试，避免实施不符合目标、成本昂贵的解决方案。

——应立即培训第一批核心数据专家，以便将数据型科学的需求转换成可用于纯数据科学家与工程师讨论的技术规格。这一类新型核心数据专家还能将技术发展与限制情况反馈给纯科学家，从而推动数据科学研究发展。

——EOSC 需要制定经过社区同意、由全球共同管理的可执行协议。这些协议应是开放、透明的最低标准，使所有科学家、创新者、工程师和服务提供者能够理解并遵守；这些协议应简化到满足数据和相关服务最基本的需求、提供最基本的支持并且严格限制在使 EOSC 正常运行的范围内（类似于互联网的 TCP/IP、HTTP 和 HTML）；这些协议应考虑所有研究对象，并成为研究对象广泛、有效（反复）使用的最低要求。

——公平原则将指导如何使研究对象可以被搜索、访问和互操作，最终使它们能被反复使用和引用。在公平原则的范围内，标准与协议应被再次限制在最低程度，从而减少未来发展需要调整协议的风险。

——当前数据共享操作和机制的复杂性需要对现有对象，尤其是跨领域对象进行温和的监管而非限制性监管，并且向许多社区一样进行标识映射。

——EOSC 应为一般概念和数据格式建立具体到领域的标准和协议。具体到领域的协议应来自社区最佳操作规范，并且支持现有和未来的欧洲科研基础设施战略论坛、欧洲科研基础设施联盟（ERIC）和其他学科研究社区联合体担任领导者的角色。通用协议应由正式和非正式的国际科学组织（例如负责研究持久化标识的开放研究者组织）、信息通信基础设施社区的信息通信技术

标准设立者、立法者和硬件、软件生产者及政府来制定（比如虚拟化和分析程序、通用软件和硬件标准、单点登录、身份验证、授权和保护等）。

——参与规则应指导遵守规定的开发者和提供者以及其他利益相关者的参与。提供者的授权机制可基于欧盟和部分相关国际计划中已实施的自我报告机制。

——与协议一样，参与规则必须开放、透明，并且由用户和提供者社区联合制定并接受，其严格程度足以防止意外和非法使用，比如数据滥用、不法的科学行为、不公平的定价、供应商的锁定、垄断和不公平的用户排斥。

——尽管参与规则应具有通用性，但也应在必要时专注于EOSC的特定类别。

3、为在欧洲开发核心数据专业知识的具体项目提供资金

应在欧洲实施大范围培训计划，培养和维持所需要的核心数据专家。这应该成为一项由欧洲科研基础设施战略论坛、信息通信基础设施等组织中的主要利益相关者和联盟领衔的社区培训项目。该项工作应制定宏大的目标：

——到2020年，每个成员国和每个学科内至少有一个认证机构支持各学科的数据监管实施。

——到2022年，培训数十万名能够对欧洲科研基础设施战略

论坛/信息通信基础设施活动和合作，以及长期可持续发展产生影响的认证核心数据专家。

4、制定具体的 EOSC 数据互操作计划

欧盟委员会及成员国的相关机构应指派执行机构担任数据监管的指导与管理角色，并采用以下具体措施：一是将具体领域问题的标准与协议制定委托给现有的国家与国际机构，包括欧洲科研基础设施战略论坛、科学协会/学术组织。如果这些机构缺席或无法履行职责，通过路线图或其他合适的指导工具促进它们快速发展。二是积极促进和支持统一广义领域中的多个欧洲科研基础设施战略论坛型社区开展合作，集体制定所在领域首选持久化标识（PPID）计划的最低规范，并映射到其他持久化标识。三是激励欧洲科研基础设施战略论坛层面的跨领域合作，实现更通用的语义类别，包括人员、组织和地区等。EOSC 应在参与规则中加入以人为本的应用，比如 Google Scholar、Research Gate 和 Academia.edu。此外还要规定数据联盟（RDA）的角色。虽然许多 RDA 工作组确定了用于数据型科学协作的协议和其他工具，但仍需要与欧洲科研基础设施战略论坛、信息通信基础设施和国家基础设施开展更有效的协调与合作。除了最低程度的严格协议之外，确保符合协议的模板、工具、数据集和服务具有最大的设计自由。

5、为准备阶段制定创新指导资助计划

为了促进变革与创新，并在 2020 年实现欧洲开放科学云，所有提出的措施应具备变革特征。必须制定预算合理的数据监管计划，将约 5% 的研究支出花费在数据的合理管理和监管上。变革计划应明确 EOSC 核心领域发展目标。应明确计划的范围、评估、选择和奖励机制，并确保为多套方案提供资金支持。

欧洲开放科学云变革计划中的优先行动：

——解决社会和技术瓶颈，包括：连接、安全与信任、性能、标准或格式等问题以及技术或/和社会障碍，如缺少数据发布与共享的奖励和激励机制等。

——发展和维护核心数据集，在定义明确的条件下向社区提供，包括工作流程、分析程序和现有数据集的创建。

——制定并实施 EOSC 连接、科学数据存储和计算的核心资源持续供应和资助计划。

——支持开发一个或多个用于在可信欧洲开放科学云存储库中搜索公平元数据的公共数据搜索引擎。

——开发各项技术和方案，有效衡量研究对象在最初发布后的反复使用和科学影响（例如，重要和获得认可的指标）。

——制定开放数据共享资助与奖励的计划。

——启动专项工作，明确加入 EOSC 的数据和研究对象。

——将单一登录问题与面向社会和专业人员的网络应用程序的连接问题相结合，为所有人员提供联合身份证明与证书。

——通过研究词汇库和软件应用支持词汇的广泛访问、反复使用与开发，提高互操作性。

——在欧盟成员国迅速培养充分的数据监管能力。

——使利益相关者社区参与指导专项工作，以开发和提供在线、可扩展、可多次使用的培训模块。不仅为专家提供培训，而且带动所使用的标准与操作规范的共同进步。

6、对所有研究提案强制进行充分的数据监管

欧洲需要研究项目中使用的当前和未来工具，包括 Horizon 2020 等，应仅支持能够合理解决开放数据监管问题的项目、可实现目标的项目、跨学科数据项目、合理解决可持续发展问题的项目或其他推动 EOSC 共同目标发展的项目在变革方案中通过简单流程获得资金。

除以上六条以外，还应为 EOSC 早期实施阶段制定清晰的运行时间表。

译自：*Realising the European Open Science Cloud, October 2016 by EU*

咨询翘楚在这里汇聚

信息化研究中心

电子信息产业研究所

软件产业研究所

网络空间研究所

无线电管理研究所

互联网研究所

集成电路研究所

工业化研究中心

工业经济研究所

工业科技研究所

装备工业研究所

消费品工业研究所

原材料工业研究所

工业节能与环保研究所

规划研究所

产业政策研究所

军民结合研究所

中小企业研究所

政策法规研究所

世界工业研究所

安全产业研究所

编辑部：赛迪工业和信息化研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：刘颖 董凯

联系电话：010-68200552 13701304215

010-68207922 18701325686

传真：0086-10-68209616

网址：www.ccidwise.com

电子邮件：liuying@ccidthinktank.com

报：部领导

**送：部机关各司局，各地方工业和信息化主管部门，
相关部门及研究单位，相关行业协会**

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区紫竹院路 66 号赛迪大厦 15 层国际合作处

邮政编码：100048

联系人：张滢星

联系电话：（010）88559658 15901454760

传 真：（010）88558833

网 址：www.ccidgroup.com

电子邮件：zyx@ccidgroup.com

