

生物多样性公约

Distr.: General
17 June 2025
Chinese
Original: English

执行问题附属机构

第七次会议

2026 年 8 月 4 日至 12 日，内罗毕

临时议程*项目 5

遗传资源数字序列信息

委托研究以透明和负责任的方式公开提供和获取遗传资源数字序列信息的数据库等工具和模型的报告的执行摘要**

作为包括卡利基金在内的多边机制临时秘书处的公约秘书处的说明***

一. 引言

1. 生物多样性公约缔约方大会在第 [16/2](#) 号决定中，通过了公正和公平分享利用遗传资源数字序列信息所产生惠益的多边机制包括一项名为“卡利基金”的全球基金的运作模式。
2. 探讨可能的新工具和模式，如数据库，以便以透明和负责任的方式向所有缔约方开放遗传资源数字序列信息，任由获取。缔约方大会在同一决定第 5 段邀请缔约方、其他国家政府、土著人民和地方社区以及相关组织提交关于这个问题的意见，并在第 6 段中请执行秘书（a）综合提出的意见和（b）委托开展一项研究，探讨将遗传资源数字序列信息以透明和负责任的方式公开提供和任由获取的备选方案。
3. 为此，秘书处发布了第 [2024-115](#) 号通知，收集到的所有意见已作为信息文件印发。汇总的意见摘要载于 [CBD/SBI/7/5](#) 号文件。
4. 秘书处还委托编写一份报告，说明提供和获取数字序列信息的现有工具与模式，对其益处与局限性作出评估，并为满足透明和负责任地使用遗传资源数字序列信息制定可能需要的新工具和新模式提出建议。
5. 本文件提供了这项研究的执行摘要。整份研究报告作为信息文件提供，其中列出了所有参考材料。

* [CBD/SBI/7/1](#)。

** 本文件未经正式编辑。

*** 关于包括卡利基金在内的多边机制模式的案文第 28 段规定设立一个秘书处，模式附文五概述了其职能（见第 [16/2](#) 号决定，附件）。在该秘书处设立之前，公约秘书处的相关官员将担任包括卡利基金在内的多边机制的临时秘书处。

二. 方法和局限性

A. 研究方法

6. 这项研究使用了来自文献的信息、各方提交的意见、非正式访谈以及为这次研究开展的两项调查所收到的回复。

7. 在 40 次非正式访谈中，受访者的选择兼顾了各种学科、地域和文化的代表性，以期收集信息、深入了解各种观点和世界观、查明差距并找出可能的解决方案。受访者构成如下：缔约方/政府代表占 22%、土著人民和地方社区代表占 22%、学术研究人员占 19%、数据库管理者占 17%、产业界代表占 14%和地方社区代表占 6%。从地域分布来看，来自欧洲、北美洲和大洋洲的受访者比例偏高，这很可能是因为来自这些地区的数据库管理者、公约秘书处工作人员以及土著人民和地方社区代表的受访人数较多的缘故。此外，仅使用英语进行访谈也可能使受访群体出现偏差。

8. 对学术界和私营部门开展的调查，其主要目标包括：（a）了解是否有必要建立新的数字序列信息数据库；（b）查明现有数据库网络中的差距；（c）为发现的差距提出解决方案。为确保受访者在地域、学科/行业及文化背景方面有广泛的代表性，开展了大量外联活动。对学术界的调查共收到 154 份回复，其中来自欧洲（45.8%）、北美洲（19%）和大洋洲（11.8%）的受访者比例较高，这一趋势与整体受访情况一致，原因也可能相同。值得注意的是，尽管努力争取私营部门参与，并不断完善调查问卷，结果只有 9 位私营部门代表做出回复。这种参与情况不足以进行人口或统计上的分析。

B. 局限性

9. 这项研究的局限性：

(a) 进行研究的时间只有七个月，因此无法进行同行评议；

(b) 目前尚无国际公认的遗传资源数字序列信息的定义；

(c) 访谈和调查仅以英语进行，这可能限制了参与人数；

(d) 在两项调查中，为了使用更广泛适用的术语，使用了“（元）数据库”一词来表明数字序列信息数据库，这令一些私营部门的受访者感到困惑。对该术语的使用已加更正，但可能已限制了私营部门的回复数量。

10. 在对私营部门的调查中，6 名受访者来自拉丁美洲和加勒比地区，3 名来自其他地区。本文仅对这些受访者的调查结果进行定性总结。

11. 当前全球数据库网络的规模非常复杂，这个网络由数千个在不同治理体系和供资机制下运行的数据库组成。此外，还有一些网络中的数据库设于非缔约方国家。

C. 这项研究中遗传资源数字序列信息的范围

12. 鉴于各方对《生物多样性公约》下遗传资源数字序列信息的范围存在不同看法¹，缔约方大会同意继续使用“数字序列信息”一词进行进一步讨论，尽管迄今尚未对

¹ 第 15/9 号决定，序言部分，第 6 段。

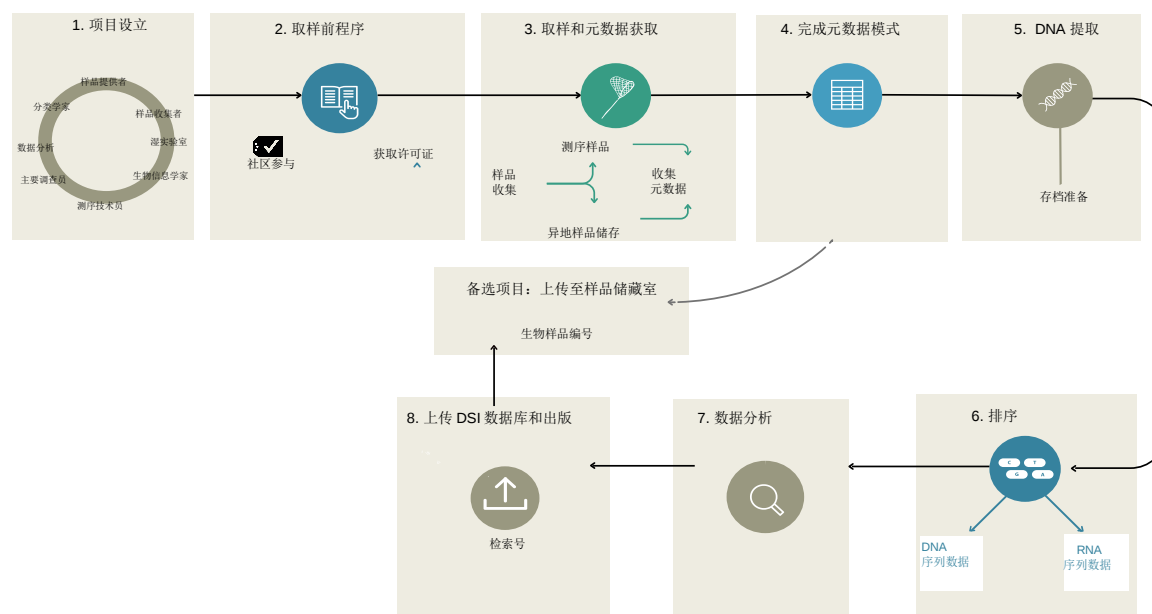
此词进行定义。在没有此词定义的情况下，这项研究中的遗传资源数字序列信息被理解为包含 DNA、RNA 和蛋白质。

三. 背景情况

A. 遗传和基因组研究过程

13. 从项目建立到提交公共数据库，遗传资源数字序列信息的一般工作流程凸显出它的复杂性，具有多步骤、跨学科、跨机构且有时具有国际流程的性质。

数字序列信息的生成和上传至公共数据库的过程²



B. 全球数字序列信息数据库网络

14. 遗传资源数字序列信息数据库的数量多寡很大程度上取决于其选用的定义。考虑到这项研究的范围，目前数据库的数量在 2,000 个到 6,000 个之间，这些数据库相互关联，也相互依存，既包括保存原始记录的主要数据库（例如，欧洲核苷酸数据库），也包括访问和利用主要数据库数据的次要数据库（例如，Ensembl、UniProt）。这些数据库在资金、历史、范围、数据传输、注册要求、数据共享、用户服务和许可证等方面各不相同。1982 年，欧洲核苷酸数据库、日本 DNA 数据库和 GenBank 组成了国际协作核酸序列数据库，该组织在各个站点之间进行数据镜像，如今已成为全球最大的公共原始数字序列信息数据库。

15. 为数据库获取统计数据极具挑战性，需要考虑诸多细节：（a）年度运营成本、年度提交量和下载量以及全部记录数量，通常包含生物多样性以外的其他数据类型，例如人类数据；（b）年度运营成本不包括数据库基础设施的初始成本；（c）数据

² 尽管这个流程呈线性展示，但根据具体项目情况，某些步骤可能需要重复进行；例如，由于样本质量不佳而需重新采样，或因测序运行效果不理想而需重新测序。

库工作人员还可能承担数据库以外的其他工作，因此工时数仅为估算值；（d）随着人工智能的发展，年度下载量正迅速成为一项意义不大的统计数据。考虑到这些因素，日本 DNA 数据库的年度运营成本为 700 万美元，存储了 63 亿条数字序列信息，门户网站浏览量为 180 万次，拥有约 30 名工作人员。规模较小的生命条码数据系统的年度运营成本为 96 万美元，但仅存储了 2,050 万条数字序列信息，门户网站浏览量为 190 万次，需要约 29 名工作人员。许多数据库未公开其 API 和 FTP 的访问统计数据。然而，欧洲核苷酸数据库报告称，到 2025 年，其 API 请求量达到每月 1.92 亿次，由此推算，每年的请求量约为 23 亿次。

16. 数据库网络还包括托管访问数据库和私有数据库。私有数据库可分为两类：一类是存储内部生成的、不对外公开的遗传资源数字序列信息（例如，制药公司的数据）；另一类是需要付费才能访问和使用其数据的数据库（例如，BaseCamp Research）。托管访问数据库是指用户必须事先获得授权才能自由访问和使用其数据的数据库。

17. 还有两种类型的托管访问数据库：要求用户在访问和使用数据库之前注册的数据库（例如，GISAID）；需要通过申请才能访问数据库或数据所有者的特定数据集的数据库（例如，Aotearoa Genomics Data Repository）。托管访问数据库模型在人类数据领域很流行，但在生物多样性数据方面不太受欢迎，主要是存有敏感数据，例如与土著人民和当地社区相关的数据。

C. 数据治理

18. 数据治理是一套操作原则、实践/框架和工具，确保数据在其整个生命周期得到准确和合法管理。有效且高效的数据治理对于成功管理公共数字序列信息至关重要。道德或效率的意义会因文化/地理/资金状况、数据类型和/或具有不同目标的项目而异。第 16/2 号决定第 10 段提到了 FAIR（数字资产的可找寻、可访问、可交互和可再用）原则、本土数据治理的 CARE（集体利益、控制权、责任和伦理）原则、TRUST（透明度、责任、以用户为中心、可持续性、技术）原则和教科文组织的《开放科学建议书》。下表列出了每组原则。

表
第 16/2 号决定提到的数据治理原则概述

原则	FAIR	CARE	TRUST	教科文组织第三部分
年份	2016 年	2019 年	2020 年	2021 年
重点	（元）数据	人与目标	数字数据库	开放科学
原则	1. 可寻找 2. 可访问 3. 可交互 4. 数字资产可再用	1. 集体利益 2. 控制权 3. 责任 4. 伦理	1. 透明度 2. 责任 3. 以用户为中心 4. 可持续性 5. 技术	1. 透明度、审查、批评和可重复性 2. 责任、尊重和问责 3. 协作、参与和包容 4. 灵活性 5. 可持续性

D. 数字序列信息背景下的人工智能

19. 随着生成遗传资源数字序列信息的成本不断下降，数据库内的数据量继续以前所未有的速度增长（例如，GenBank 的规模每 18 个月增加一倍）。基因组数据类型和相关信息（注释、元数据等）的多样性也在扩大，因此传统的生物信息学工具难

以跟上步伐并保持相关性。人工智能是遗传资源数字序列信息领域的一项不断发展的技术，随着数据集的不断增长和多样化，人工智能尤其有用。

20. 随着技术进步，人工智能在基因组学领域的准确性不断提高，其在基因组学领域的应用不断增长。然而，它的缺点也应该被考虑到：训练人工智能的不平等数据集的偏差可能会在其结果中被放大、由于现有的数字鸿沟而导致的有限参与可能会扩大、能力差距也可能会扩大、这种快速增长的工具的能源成本造成了越来越大的环境影响和增加了能源需求的压力和竞争。最后，与遗传资源数字序列信息相关的元数据中的现有偏见和差距限制了人工智能工具在生物多样性研究中的效用。

E. 土著人民和地方社区的考虑

21. 2022年，缔约方大会第15/9号决定认可了CARE原则，并认识到使用遗传资源数字序列信息所带来的货币和非货币利益除其他外应惠及土著人民和当地社区。2024年，缔约方大会第16/2号决定认可了《联合国土著人民权利宣言》³，并规定卡利基金的至少一半资金应支持土著人民和地方社区包括这些社区内的妇女和青年的自我确定需求。此外，缔约方大会第16/4号决定通过了到2030年关于《生物多样性公约》关于土著人民和地方社区的第8(j)条和其他条款的工作方案，其中要素3涉及分享利用遗传资源和遗传资源数字序列信息以及与遗传资源相关的传统知识所产生的惠益。

四. 研究结果摘要

A. 查出差距

22. 查出关于以透明和负责任的方式公开提供和获取遗传资源数字序列信息以及建立公平和公正地分享其使用获得的惠益的系统的四大差距。这些差距包括：

(a) **访问/使用差距：**由于缺乏关于谁在访问和使用数据的透明度，人们对现有的公开数字序列信息及管理信息的数据库缺乏信任；

(b) **合规差距：**公共数据库中的数字序列信息缺乏相关的合规信息（许可证信息、合规元数据），这造成重大的问责差距。一些主要数据库要求提交者勾选确认所提交数据不受限制，但这不具法律约束力，也未在所有公共数据库中得到普遍实施；

(c) **沟通/教育差距：**决策者与数据使用者之间沟通渠道有限，这被视为一种访问差距。人们对使用者是否了解《生物多样性公约》、获取与惠益分享义务以及多边机制表示关切。反之，也有人担心决策者不了解公共数字序列信息被访问和取用的细节。其他关切问题包括公共数据库的地理分布不均，这导致在制定和拟定有关遗传多样性数字序列信息的存储、提交、访问及再利用政策方面存在地理上的不平等；

(d) **关于负责任地管理遗传资源数字序列信息以及与土著人民和地方社区相关的传统知识方面的差距。**人们普遍认为，公共数据库与FAIR原则高度契合，大型数据库也在努力符合TRUST原则。然而，人们指出，在考虑符合CARE原则、《联合国土著人民权利宣言》以及《公约》关于第8(j)条的实施所制定材料的要求时，

³ 《联合国土著人民权利宣言》，2026年4月13日参阅，A/RES/61/295号决议。

存在时间和资源不足的问题；这可能对负责任地管理遗传资源数字序列信息以及与土著人民和地方社区相关的传统知识产生负面影响。

B. 拟议的工具、模型和数据库

23. 根据提交的意见、访谈和调查结果，针对查明的差距提出了18项解决方案（工具/模型和数据库），并将其归纳为三大类（括号内列出各类的拟议工具）：

(a) 构建新数据库（完整数据库、有价值数据数据库及许可证数据库）；

(b) 修改现有公共数据库网络（强制注册、网络分析工具、用于追踪学术界、私有部门活动及相关传统知识的人工智能、教育工具、元数据工具、法律工具（许可证和法律协议）、传播工具、治理工具）；

(c) 对与土著人民和地方社区相关的数字序列信息及传统知识的考虑（教育资源与培训、元数据改进、将与土著人民和地方社区相关的数据纳入公共数据库的战略、治理工具）。

24. 对各项提案优缺点的评估，请参阅信息文件的表7和表9。每项提案的评估标准包括法律与技术难度、财务负担以及对现有数据库网络支持的依赖程度。此外，相关提案还根据其与CARE、FAIR和TRUST原则的契合程度进行了评估。

25. 关于由公约秘书处托管一个涵盖遗传资源数字序列信息的完整数据库的提案、关于建立类似但仅限于更加受限数据的数据库的提案以及关于建立监测访问与利用许可证的数据库的提案，均被认为在技术和财务上极大困难，且在法律上取决于缔约方的决定。在现有数据库中建立强制性注册系统、对数据集实施许可证制度以及要求使用者在提交或访问数据前签署法律合同，这些措施均被认为在技术、法律和财务方面存在从困难到极难的挑战。然而，一些提案被认为在技术、法律和财务上可行，例如，开发网站分析工具宏观监测数据库网站的访问情况、利用人工智能追踪使用公共数据库中遗传和基因组数据的科学出版物或利用相同方法或使用相关传统知识的专利申请。此外，关于《生物多样性公约》三大目标及第8(j)条、数据治理、多边机制及其卡利基金的教育工具以及关于规范数据引用、许可证颁发和数据治理的指导方针，也被认为是可行的。最后，在数据库中设置元数据字段以及通过获取和惠益分享信息交换所传播教育工具和指导方针也被评估为技术、法律和财务难度较低的提案。为落实其中部分教育和指导工具，有人建议成立具有包容性的工作组，另有人建议将土著人民和地方社区纳入这个工作组。

26. 针对涉及土著人民和地方社区及相关传统知识的提案，修改现有数据库以纳入由土著人民负责并管理其访问权限的标准化元数据被评为非常困难。在另一方面，将部分教育工具的重点放在管理涉及土著人民和地方社区的数据则被认为相对容易，但这需得到数据库的支持与配合。制定最低元数据要求以披露本地状况标签与声明以及地理空间坐标，其难度评估结果也与之相同。最后，监测数据库落实CARE原则的进展情况被认为难度较低，可由数据库自行作出，也可由独立机构进行。

C. 私营部门调查结果摘要

27. 私营部门调查的回复数量有限（n=9），且地域分布不均，主要集中在拉丁美洲和加勒比地区，产业的多样性也低。受访者对建立新数据库普遍持负面态度，超过半数认为此举并无明显益处。主要关切包括：与现有系统重复、互操作性受到挑

战、合规负担增加、可能侵犯土著人民和地方社区的权利以及用于保护工作的资源被分流。其他风险包括：若由《公约》负责设立数据库，则治理流程可能缓慢、数据碎片化和因增加官僚程序和潜在的惠益分享义务，导致使用遗传资源数字序列信息的积极性降低。关于修改现有数据库，受访者表示了中度至高度的支持，具体包括：（a）强制注册；（b）与缔约方共享数据库访问情况的产业汇总信息；（c）在出版物和专利披露数据使用情况；（d）纳入地理空间元数据；（e）扩展获取和惠益分享信息交换所的连接。受访者普遍支持采用知识共享许可协议，尽管也有人倾向于保持灵活性而非强制实施。最后，各方就承认土著人民和地方社区的权利与利益达成强烈共识，并广泛支持使用“本地状况标签与声明”来标记遗传资源数字序列信息。许多受访者还支持针对与土著人民和地方社区相关的遗传资源数字序列信息实施托管访问系统，尽管也有人提醒，这可能与支撑多边机制的开放获取原则相冲突。

五. 结论

28. 这项研究确定了四类差距（访问/使用、合规、沟通/教育以及负责任地管理与土著人民和地方社区相关的数字序列信息）和 18 项拟议工具、模型及数据库（表 2 和表 3），这些工具、模型及数据库可归为三类：（a）构建新数据库；（b）修改现有公共数字序列信息数据库网络；（c）对与土著人民和地方社区相关的数字序列信息及相关传统知识的考虑。

29. 对拟议工具、模型和数据库的分析表明，与新建大规模基础设施相比，渐进式、互补性的解决方案更具优势。关于建立新数据库的提案（例如，完整、“有价值”和有许可证的数据库）在提高法律确定性、建立信任、制定标准以及尊重国家主权方面具有潜在益处，并与第 15/9 号决定第 9 段中的要素高度契合。然而，在成本、增加运作阻力、可行性、与现有公共数据库网络的互操作性、重复建设、数据碎片化以及可能导致数字序列信息的获取和使用受限等方面，这些方案的弊端往往超过了其益处。

30. 相比之下，更具针对性的工具，如网站分析、利用人工智能追踪出版物、相关传统知识与专利、教育资源、治理工具和传播工具，被评估为更具可行性和可扩展性。此外，这些工具也被认为与第 15/9 号决定第 9 段及第 16/2 号决定第 10 段更为广泛地契合。它们在提供更高的透明度、问责制和可获取性的同时，对法律、资金和技术资源的需求相对较低。不过，即使是这些方法也面临着挑战，例如科学界和私营部门的采纳度、隐私问题以及需要在全球数据库网络协调进行实施。

31. 重要的是，针对与土著人民和地方社区相关的数字序列信息的提案与 FAIR、CARE 和 TRUST 原则以及第 15/9 号决定第 9 段高度契合。这些提案也有力地支持了在可见度、包容性、归属权和治理参与方面的改进，特别是通过完善元数据、监测 CARE 原则的实施、推行包容性治理以及提供教育资源等方式。这些方法被视为切实可行，但需要持续的资源投入以及土著人民和地方社区的不断参与。