



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



国家天文台2023年度职工培训计划

中国科学院先导专项项目数据汇交管理要求

中国科学院科学数据总中心

刘宁

2023.11.28

≡ PART 01

政策机制解读



科学数据汇交管理的提出背景



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院成立于1949年11月，是中国自然科学最高学术机构、科学技术最高咨询机构、自然科学与高技术综合研究发展中心，同时也是重要科技项目管理机构



100+机构

- 106所研究机构
- 2所大学
- 450+衍生企业



科研设施

- 31大科学装置
- 100+观测台站
- 4国家实验室



科研投资

2022年科学技术方面支出
109亿元



跨学科领域

- 物理、化学等6个学部
- 涉及11个行业



地理分布

- 设有12个地方分院
- 分布在20多个省市



学术产出

- 十一年自然指数全球首位
- 大量数据



科学数据汇交管理的提出背景

研究团队自发管理

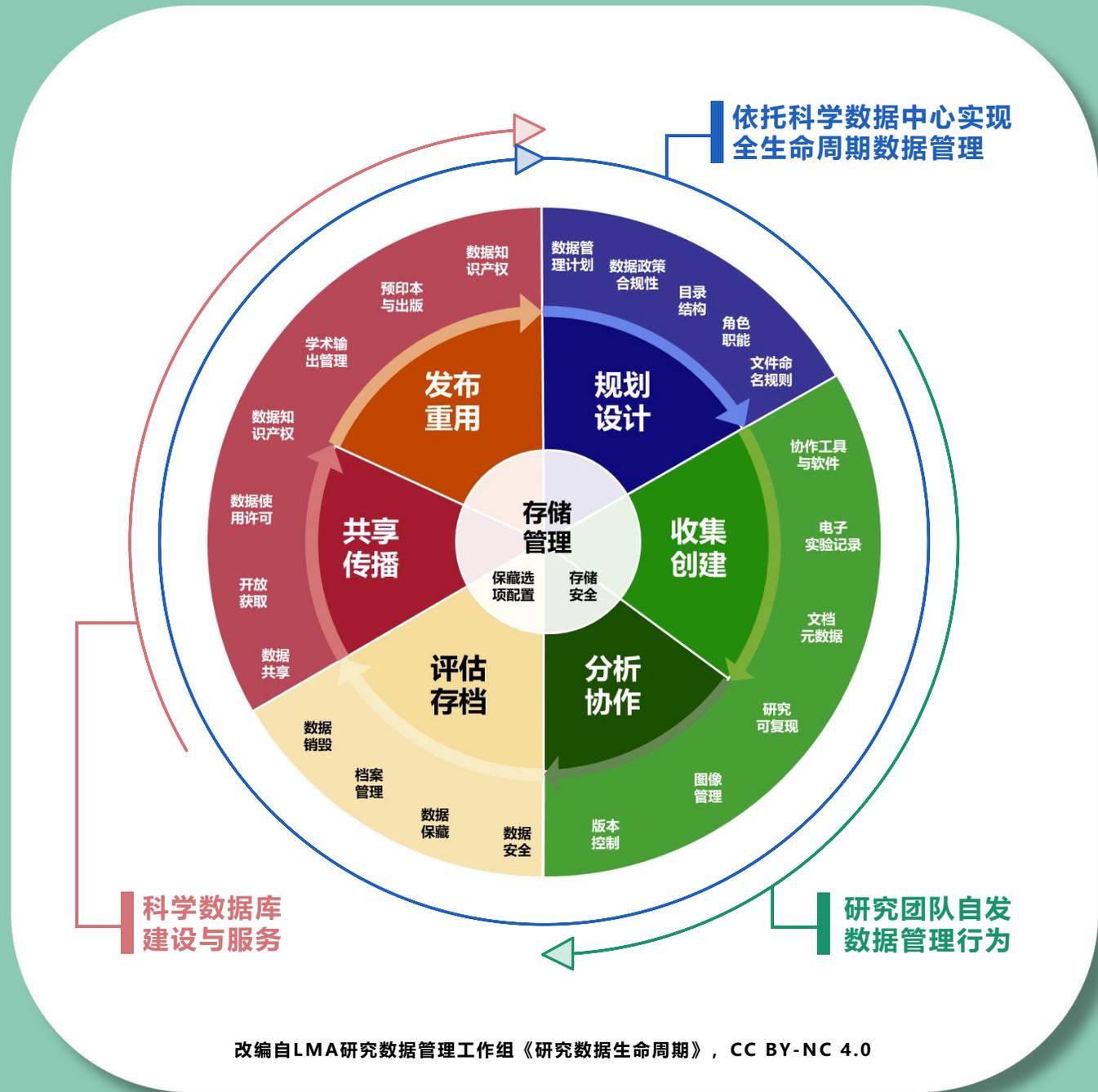
研究团队自发采集数据并分析结果，**缺少规划**，大多数年后**散失**，数据**不共享**流通或仅私下共享，**知识产权不明确**，**不可引用**，价值**不可评价**

科学数据库建设

依托优势资源或科学装置等整编加工和发布共享数据，受限于**基础领域**，反映研究的**周期长**，**权益信息损失**问题严重，**不可引用**，项目驱动**建设成本较高**，支持结束后亦从建设方手中逐渐散失

科学数据汇交管理

将研究团队产生的科学数据托管至科学数据中心保藏和共享，编制**数据管理计划**，**结合一线研究**，**作者授权**，托管**保藏**，有序开放**共享**，实现**全生命周期**科学数据管理，当前主要通过主管部门政策推动



先导专项数据汇交管理的政策环境

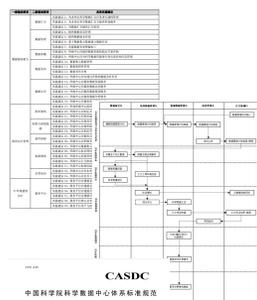


先导专项项目科学数据汇交的政策环境目前由国家、院和先导细则三级科学**数据共享管理政策、数据安全与个人信息保护法规**和**保密法规**共同构成，目前仍在快速建设健全

院科学数据中心体系汇交支撑保障能力建设

2021年，我院科学数据建设基础最好，能力最强的31个单位首批通过中国科学院科学数据中心认定，纳入中国科学院科学数据中心体系统一管理；针对数据汇交业务特点，体系国内较早提出**5项指南**，对各中心的汇交服务提出规定；面向数据中心通用IT能力建设需求，发布**科学数据中心软件栈12项支撑软件**，建成统一的全生命周期**科学数据汇交管理平台**

标准规范建设



科学数据中心建设指南

院体系科学数据中心建设要求，分为3方面，共48项实施建议

科技项目数据汇交指南

院项目科学数据汇交的职责、内容、流程与阶段性成果模板

科学数据中心安全管理

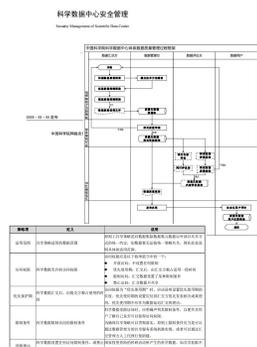
科学数据中心物理、通信、系统和数据等方面的建设要求

科学数据质量管理

项目汇交过程中数据质量控制的流程和主要质控点审核要求

科学数据共享管理

管理部门政策制订、汇交数据授权与许可、防止掠夺出版等



团队建设、培训、咨询服务、客户支持、审核质控规范

软件工具环境建设

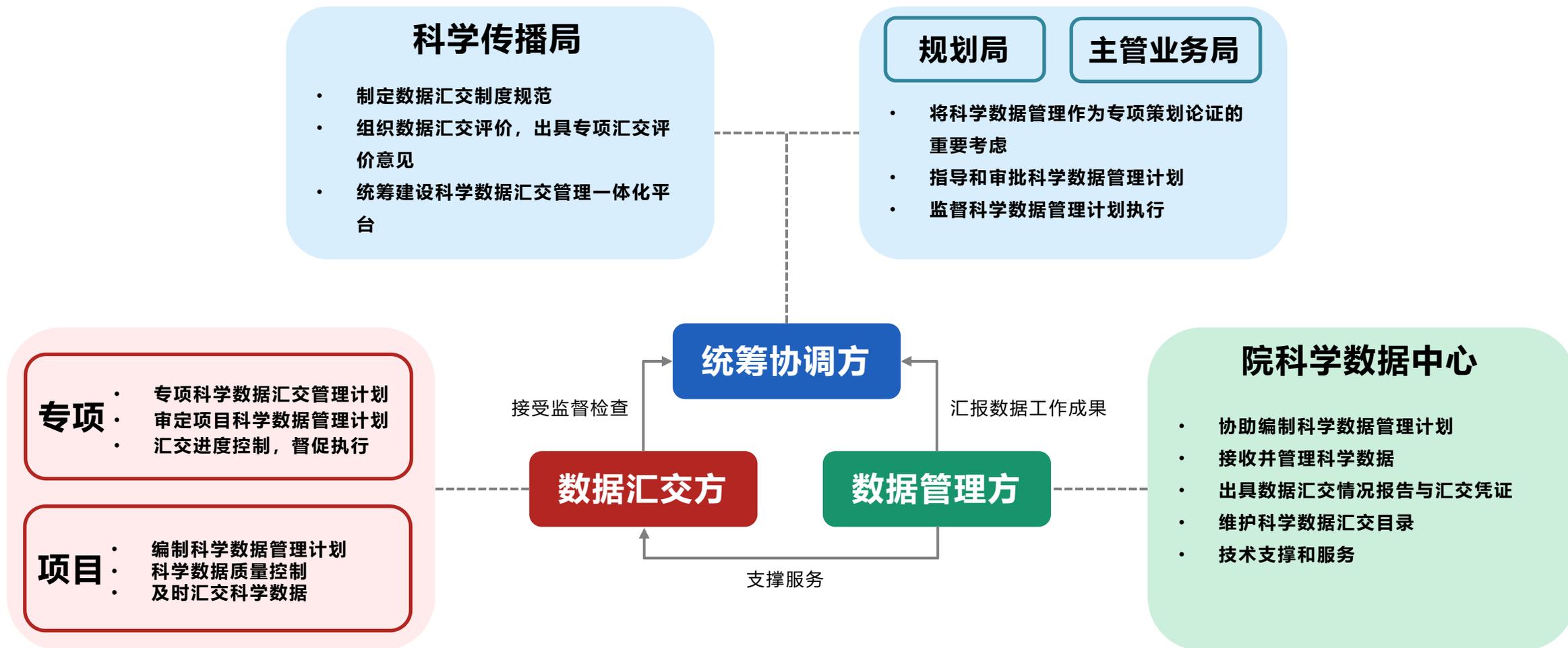
科学数据软件栈

科学数据中心全生命周期软件工具**12项**，覆盖体系**27/31个**数据中心，已支撑**7/13**中心开展先导项目数据汇交

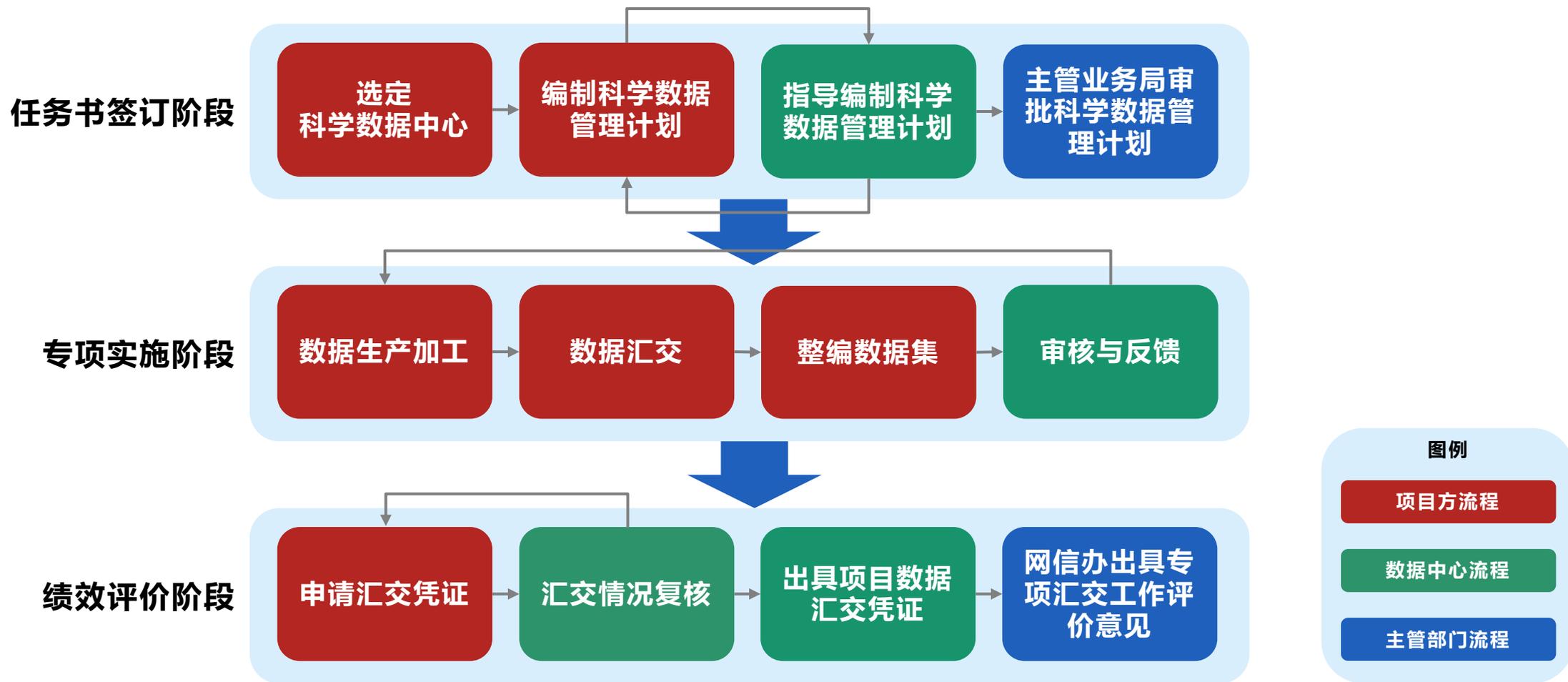
科学数据汇交管理平台

全球领先实现机器辅助科学数据管理计划制订与全体体系数据汇交情况统一追踪，机器辅助审核校验，和汇交凭证发放与验证

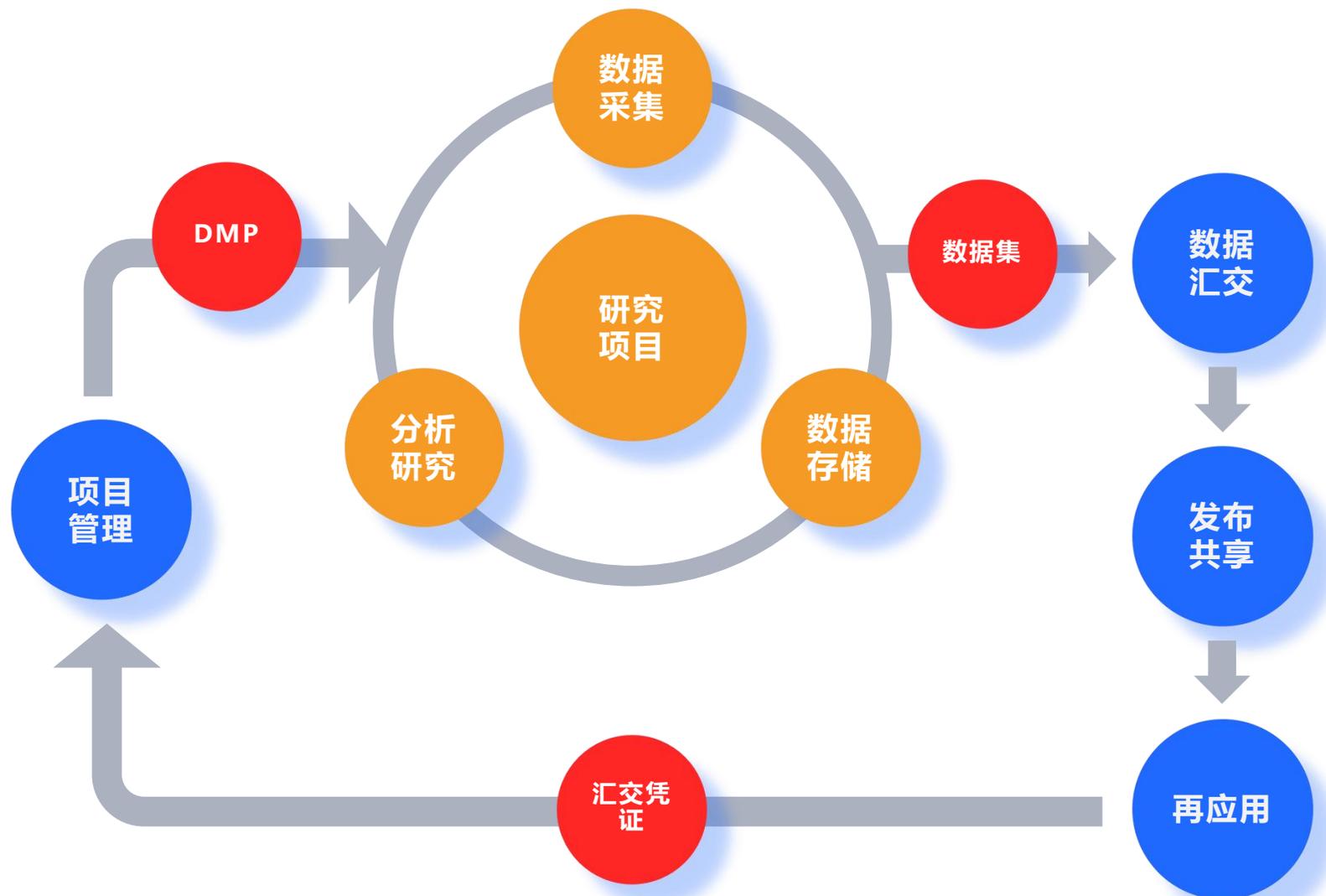


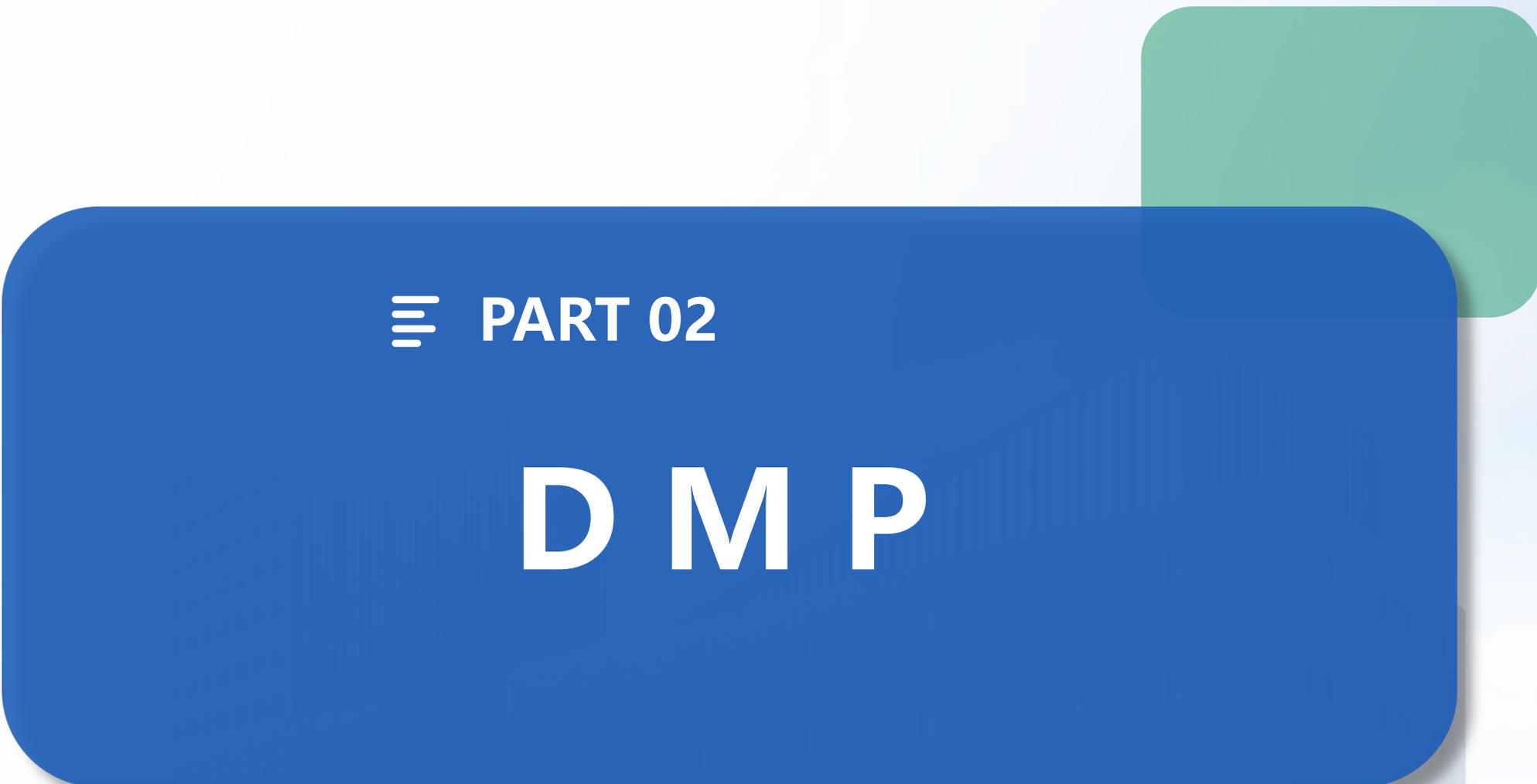


整体工作流程



项目数据汇交管理的时间节点和各阶段成果





☰ PART 02

D M P



创建数据管理计划



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

数据管理计划 (Data Management Plan, DMP) 是指科研项目在开展科学研究时, 为规范科学数据的管理和使用, 明确数据的出处、著作权、使用权限、共享方式等, 制定的数据管理规划和实施方案。数据管理计划应当包括数据的收集、整理、归档、共享、保存、保护等环节, 以及数据的格式、元数据、文档等方面的详细规定。数据管理计划的编制和执行能够有效提高科学数据的管理水平和数据的可靠性, 促进科学知识的分享和交流, 为科学研究提供更好的支持和保障。

 **入口** <https://dmp.casdc.cn/>

 **执行人** 项目 / 课题数据汇交专员

 **时间** 原则上立项阶段

 **内容**

- 项目 / 课题信息
- 邀请汇交专员
- 数据汇交指标
- 对口数据中心



项目信息 *

* 项目/课题名称

项目/课题编号

* 实施主管部门

* 项目类型

* 项目/课题状态

* 项目持续时间 至

* 项目承担单位

项目参与单位

项目简介

上传背景材料

上传背景材料

请上传可供审核员理解项目任务的背景材料, 如“任务书”、“申报书”等。背景材料仅供审核员审核使用。最多上传5个文件, 单个文件大小不得大于30MB。
文件格式: doc, docx, pdf, zip.

上传

创建 / 编辑项目/课题信息

数据汇交专员填写项目/课题或拟立项项目/课题基本信息



粒度

专项统一决定在项目/课题层级, 建议项目层级, 不支持多层级DMP



名称/编号

- 按所选层级认真填写项目/课题名称/编号
- 数据资源元数据的项目/课题信息或致谢等处必须与此严格对应



附件

一般应将任务书或其它关于项目/课题考核指标, 研究内容的证明材料作为附件提交, 以作为数据中心验证数据汇交指标完备性与合理性的参考依据

邀请 / 编辑项目数据 汇交专员

项目/课题委派至少一名具有在项目运行期间长期跟踪汇交进度、调度项目成员按要求汇交数据的汇交专员



汇交专员

至少安排一人负责跟踪进度，协调调度，不必本人进行汇交



负责人

建议项目/课题负责人在列表当中



联系信息

建议使用常用的工作邮箱，至少一人有联系电话



编辑

汇交专员团队在任何汇交阶段都可以编辑更新

项目数据汇交专员 *

邀请成员

姓名	电子邮件	电话	单位	项目角色	操作
陈昕	chx@cnic.cn			项目负责人	移除 编辑
刘宁 我	ln@cnic.cn			项目数据汇交专员	移除 编辑

共 2 条 < 1 >

创建数据管理计划

科学数据汇交指标 *

新增指标

DMP指标名称	数据类型	数量	数据量	文件数	汇交时间	访问权限	授权许可	保藏计划	操作
一种新型XXXX导体材料的计算数据	数据集	1	10.00MB	5	2023-08-31	有条件公开	CC BY 4.0	10年以上	编辑 移除
XXXX引擎人工智能大数据训练库	数据集	1	100.00GB	5300	2024-04-01	完全公开	CC BY 4.0	10年以上	编辑 移除
XX海域XXX数据集产品	数据集	1	1.00GB		2023-04-30	保护期	CC 0	5年以上	编辑 移除
XXX原理XXX计算软件 XXXX超算版	软件/代码	1	5.00MB		2023-12-31	完全公开	BSD	10年以上	编辑 移除
XDAXXX课题论文	论文	10	N/A	N/A	2024-04-30	不公开	N/A	10年以上	编辑 移除

共 7 条 5 条/页 < 1 2 >

创建/编辑数据汇交指标

参考任务书和研究内容，细化项目拟形成各类数据资源的内容和技术规格，指标都是下限



指标名称



汇交时间



资源类型



访问权限



数量



授权许可



数据量



保藏计划



文件数



数据质量保证

汇交指标统计分析

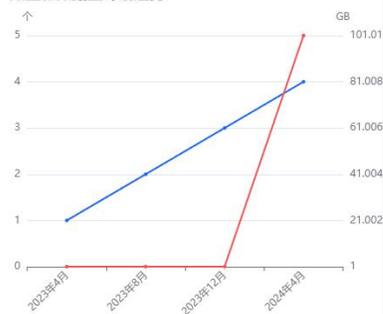
数据集计划汇交数

21

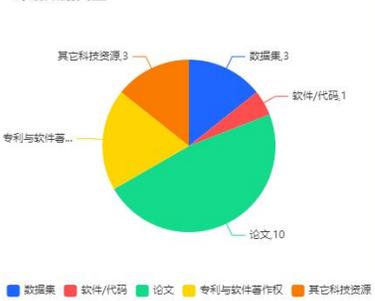
计划汇交数据量

101.01 GB

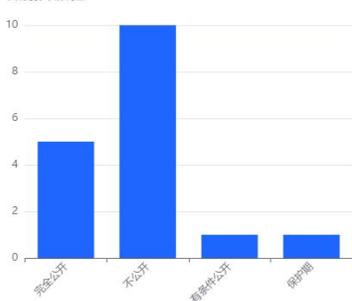
数量和数量计划趋势



计划数据类型



数据开放性



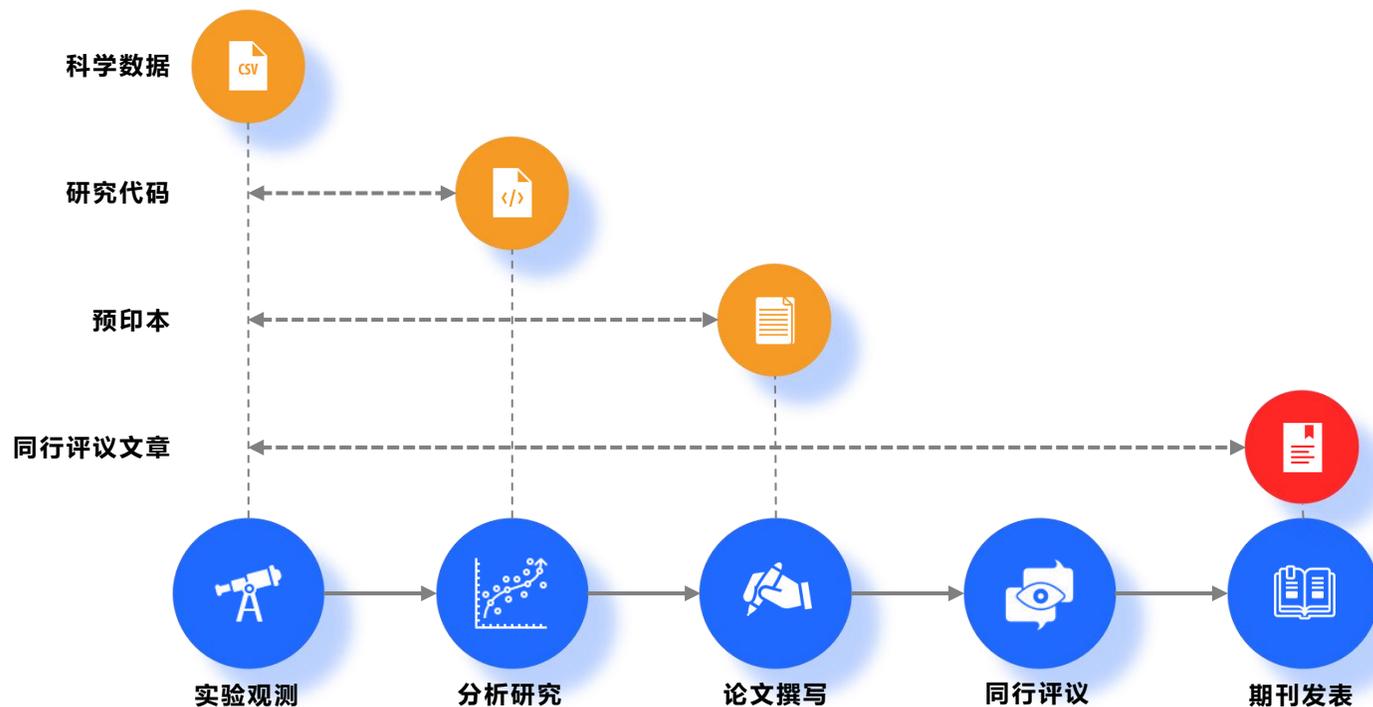
哪些数据资源应列入计划

应该列入的数据资源

- 任务书考核指标中明确的数据资源/软件
- 任务书研究内容叙述可产生的数据资源/软件
- 任务书主要成果的关联数据/软件
- 其他科技资源, 如论文, 专利, 软著

可不列入和不可列入的情形

- 管理维度的不列
- SM数据不得列入
- 非国家财政支持可不列
- 涉及个人敏感信息未脱敏不得列入
- 符合重要数据条件未获有关部门审批不得出境共享



学术成果可共享的时间延迟

汇交指标的命名

1 可以确定具体名称

使用具体名称，不必完全一致

- XX海域XXX数据集产品
- XX高原地形急变带XXXX要素数据集
- XXXX引擎人工智能大数据训练库

2 可以具体但不能确定名称

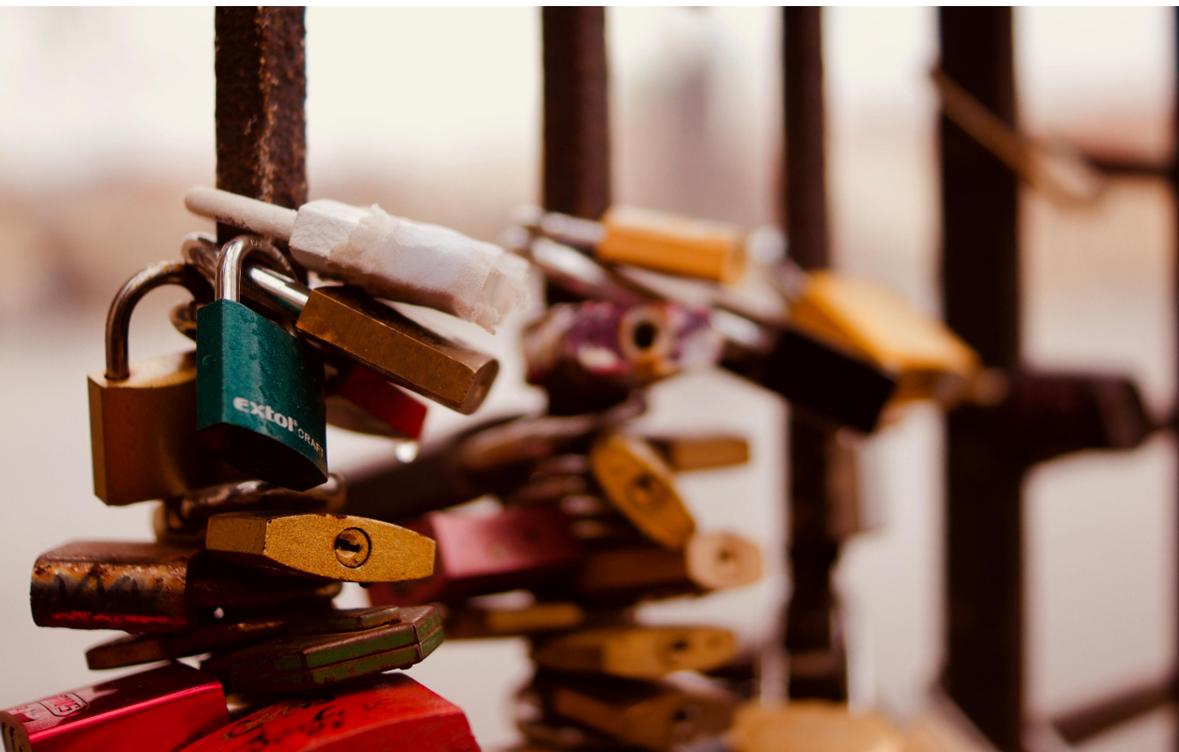
就具体数据资源的内容进行描述

- 蛋白质三维结构
- XXXX技术实验数据

3 仅能确定类型和数量

只写类型和数量

- XDAXXXX课题论文
- XXXXXX数据采集指南各1份



确定访问权限

数据资源是否或将以何种方式通过科学数据中心像用户开放，在以开放为常态的原则下，例外应有相应的理由

完全公开

没有任何保护诉求，数据提交后立即无条件向所有人开放共享

保护期

仅存在学术保护的诉求（如首发论文），在提交后保持不对他人公开一段时间（不超过24个月）

有条件共享

存在较强的商业利益保护诉求或按规定的其他保护需求，仅面向符合特定条件（如国内用户、审核通过等）的用户共享

不共享

极强的保护需求，在充分说明理由并经过数据中心和主管部门审核通过的前提下，仅提交不共享

在符合相关政策法规的前提下，不存在过度保护，以开放为常态

确定许可协议

- 许可协议是作者在适当保护自身著作权益的前提下，向用户渡让一部分权益，使用户可以在满足作者诉求的前提下使用数据
- 没有许可的数据 (all rights reserved)，即使公开也不可食用
- 建议开放许可协议，不建议自订协议，禁止掠夺性协议

数据集

推荐CC-BY 4.0，对方使用数据时应适当署名，如引用、致谢

软件

建议使用开源软件常用协议，如MIT、BSD、Apache、GPL等

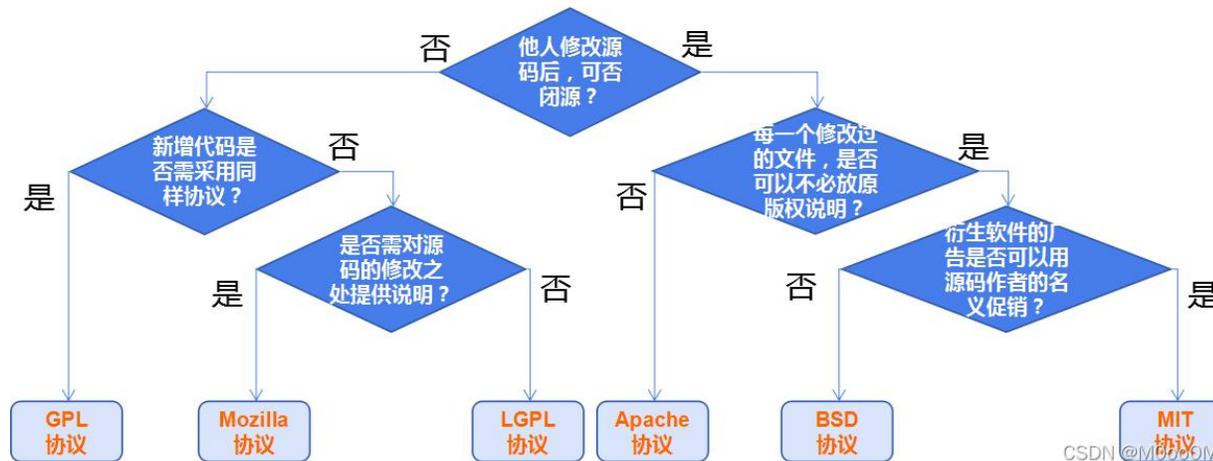
已发表论文

不得改变原文的许可，出版社协议不允许再传播的也不得公开

专利、软著

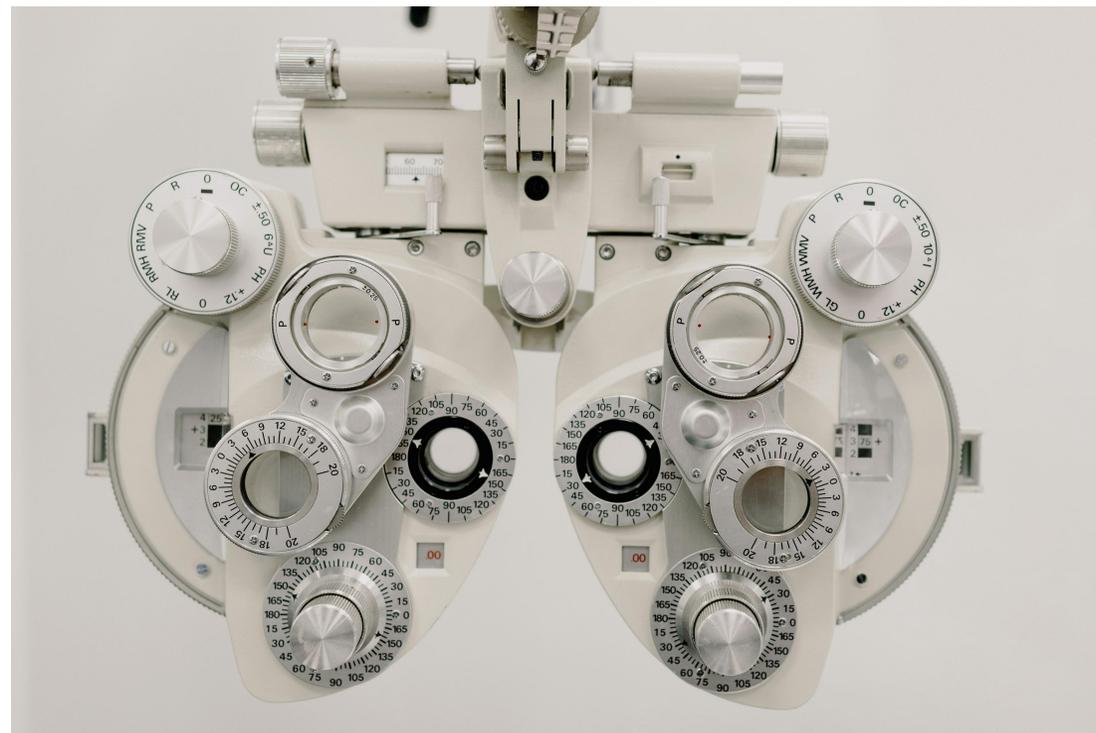
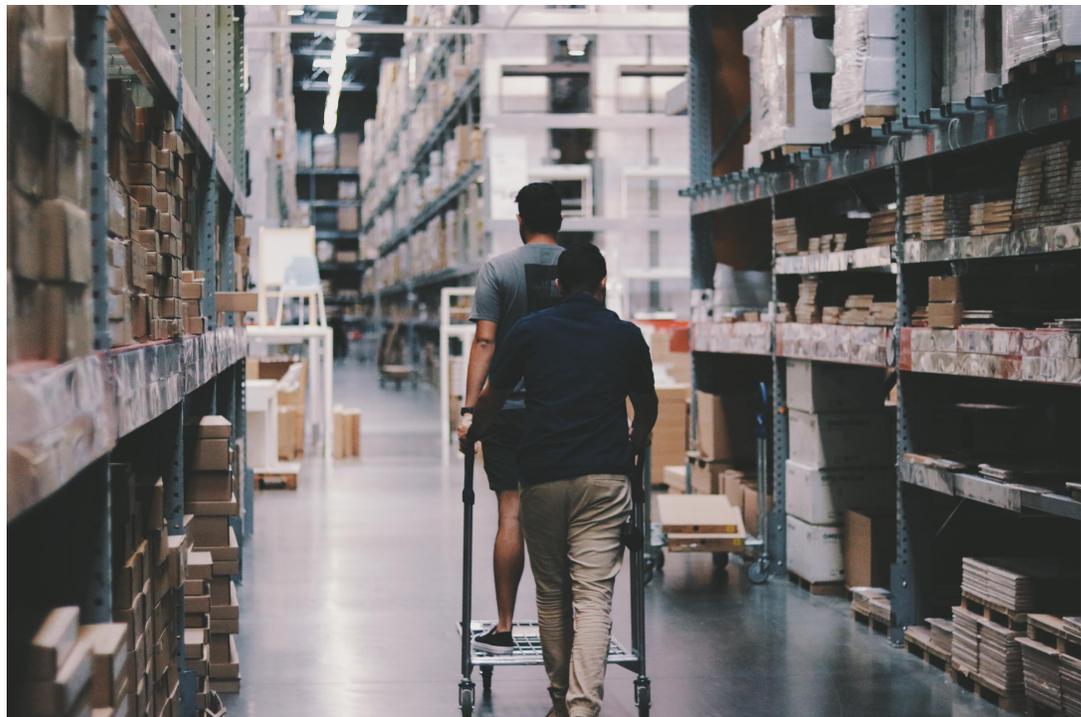
不涉及通过科学数据中心授权许可

图标	说明	缩写	署名要求	混合作品	允许商业性使用	符合自由文化作品	符合OKI的“开放定义”
	不受限制地在全球范围内发布内容	CC0	否	是	是	是	是
	署名 (BY)	BY	是	是	是	是	是
	署名 (BY) - 相同方式共享 (SA)	BY-SA	是	是	是	是	是
	署名 (BY) - 非商业性使用 (NC)	BY-NC	是	是	否	否	否
	署名 (BY) - 非商业性使用 (NC) - 相同方式共享 (SA)	BY-NC-SA	是	是	否	否	否
	署名 (BY) - 禁止演绎 (ND)	BY-ND	是	否	是	否	否
	署名 (BY) - 非商业性使用 (NC) - 禁止演绎 (ND)	BY-NC-ND	是	否	否	否	否



确定保藏计划

数据中心根据本领域不同数据的重要程度，制订数据保藏分级策略，越重要的数据应在数据中心保藏越长时间，期间数据中心应保障数据安全性和可用性



数据质量保证

质量是造出来的，关于数据形成过程中的实验设计，装置精度，数据源选择，清洗方法，字段命名，采用相关标准等方面计划的叙述

创建数据管理计划



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

选择数据中心

1 学科就近，自主选择

优先选择具有相关领域数据治理能力的数据中心

2 学科就近，自主选择

以便数据中心根据专项办的需求实施汇交管理

3 适当兼顾其他管理诉求

数据管理计划摘要

项目类型 中国科学院战略性先导科技专项 (B类)

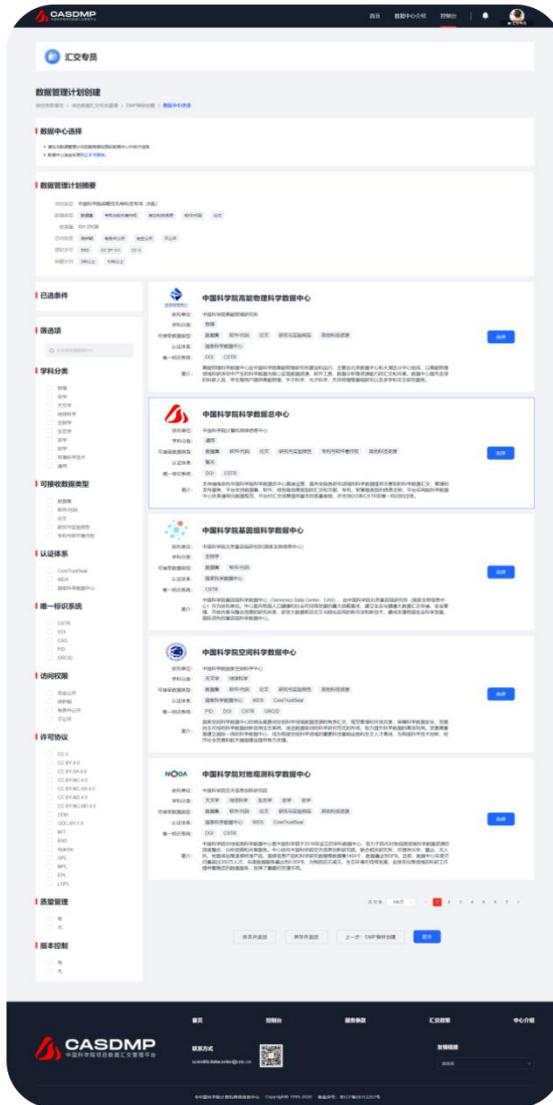
数据类型 **数据集** 专利与软件著作权 其它科技资源 软件/代码 论文

数据量 101.01GB

访问权限 **保护期** 有条件公开 完全公开 不公开

授权许可 BSD **CC BY 4.0** CC 0

保藏计划 5年以上 10年以上



1

下载待盖章DMP

中国科学院科学数据管理计划



中国科学院 中国科学院战略性先导科技专项（B类） 科学数据管理计划

1. 汇交项目信息

简要描述汇交项目的基本信息，已立项项目严格按实际立项情况填写，申报中的项目按拟申报情况填写，并在立项成功后在系统中按实际情况更新。

表1 汇交项目信息

项目/课题名称	4月4日演示案例		
项目/课题编号	TEST0404	申报状态	已批准
开始时间	2022-04-01	结束时间	2024-04-01
承担单位	中国科学院计算机网络信息中心		
参与单位	中国科学院大连化学物理研究所		
简介	该计划仅用于演示汇交的操作流程		

2. 项目成员信息

请填写项目汇交主要负责人及相应的联系信息，如有多个请添加行。

表2 项目成员信息

项目角色	姓名	电子邮件	单位名称
项目负责人	陈昕	chs@cnic.cn	/
项目数据汇交专员	刘宁	ln@cnic.cn	/

3. 拟汇交的院科学数据中心

中国科学院科学数据总中心

2

上传已盖章扫描件

中国科学院科学数据管理计划

5. 相关单位签章

项目负责人承诺:
项目负责人(签字) _____ 年 月 日
课题负责人承诺:
课题负责人(签字) _____ 年 月 日
项目承担单位意见:
法人代表(签字) _____ 承担单位(公章) _____ 年 月 日
院科学数据中心意见:
主任(签字) _____ 承担单位/数据中心(公章) _____ 年 月 日

本科学数据管理计划由中国科学院项目科学数据汇交管理平台导出
(<https://dmg.casdc.cn>)



☰ PART 03

数据集

数据汇交

数据汇交，由数据作者或其授权人逐项进行，汇交内容包括实体数据、元数据、必要时还应提交与数据使用相关的必要附件，全部数据汇交完成后由相关负责人按项目任务要求整编形成数据集

-  **入口** 对口数据中心
-  **执行人** 具体数据集作者
-  **时间** 早于数据管理计划约定时间
-  **内容**
 - 提交实体数据
 - 填写数据集元数据
 - 提交必要的附件



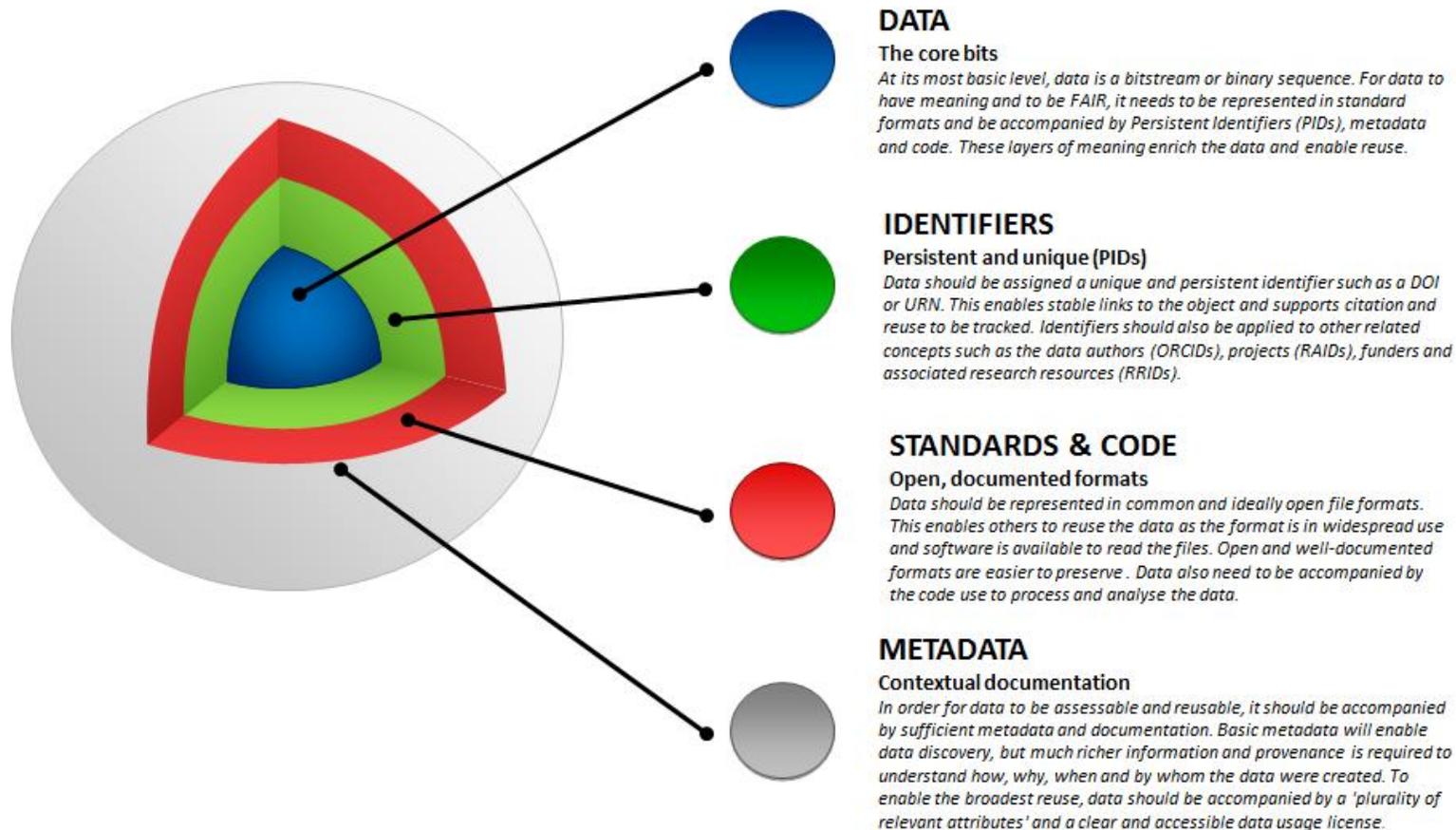
中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



The screenshot displays the National Astronomical Data Center (NADC) website interface. At the top, there are navigation menus for 'Home', 'About NADC', 'Data Submission', 'Data Access', 'Services', and 'Contact Us'. Below the navigation, there are several main sections:

- 新闻动态 (News):** A list of recent news items with dates and titles.
- 编辑数据 (Data Editing):** A section for editing data, including options for 'Data Entry', 'Data Review', and 'Data Approval'.
- 数据汇交 (Data Submission):** A central section for data submission, featuring a 'Data Submission' button and a 'Data Submission' form.
- 公共服务 (Public Services):** A section for public services, including 'Data Access', 'Data Download', and 'Data Query'.
- 平台统计 (Platform Statistics):** A table showing statistics for different data submission platforms.
- 联系我们 (Contact Us):** A section for contact information, including phone numbers, email addresses, and a QR code.

平台名称	平台访问量	数据量 (TB)
FTP	35194	462.63
Web	3651.50	436.72TB



RDA FAIR数据对象模型

实体数据

实体数据应该采用同类数据通用格式，直接或在说明文档和相关工具的支持下具有可用性，数据内容与数据管理计划和元数据说明一致，且不存在涉及国家安全与伦理道德等方面的问题

元数据

提供丰富的元数据信息，包括关于数据内容的描述、著作权声明、共享方式、许可协议、资助项目、关联论文、参考引用来源等方面

附件

必要时，数据作者还应提交：

- 数据字典
- 专用格式数据解析/处理工具
- 数据质量说明文档/数据论文
- 特定数据所需的资质证明

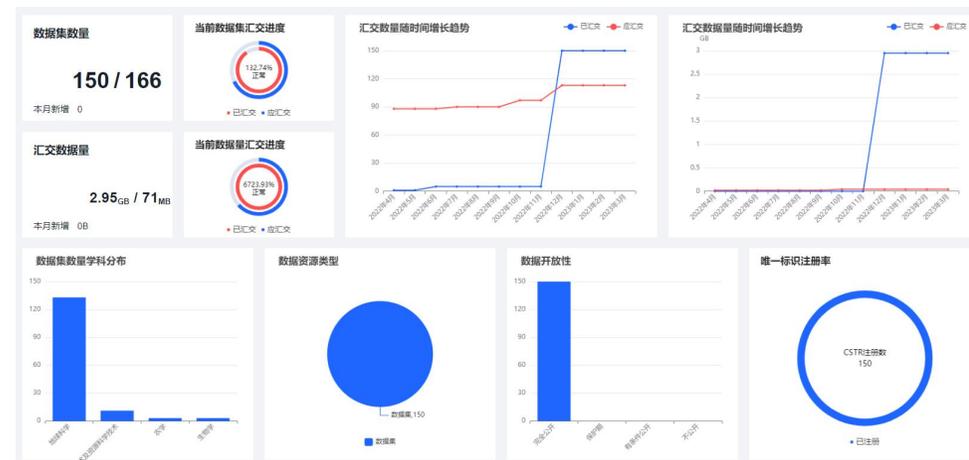
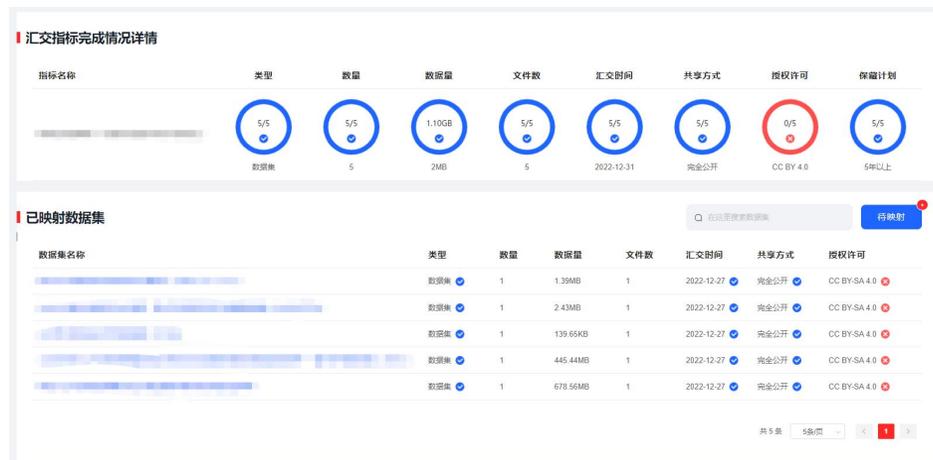
数据汇交进度跟踪

数据汇交专员定期跟踪数据管理计划的执行情况，调度相关项目成员提前前往数据中心完成对应数据集的提交

DMP指标名称	数据类型	数量	数据量	文件数	汇交时间	访问权限	授权许可	保密计划	操作
一种新型XXXX导体材料的计算数据	数据集	1	10.00MB	5	2023-08-31	有条件公开	CC BY 4.0	10年以上	编辑 移除
XXXX引擎人工智能大数据训练库	数据集	1	100.00GB	5300	2024-04-01	完全公开	CC BY 4.0	10年以上	编辑 移除
XX海域XXX数据集产品	数据集	1	1.00GB		2023-04-30	保护期	CC 0	5年以上	编辑 移除
XXX原理XXX计算软件 XXXX超算版	软件/代码	1	5.00MB		2023-12-31	完全公开	BSD	10年以上	编辑 移除
XDAXXXX课题论文	论文	10	N/A	N/A	2024-04-30	不公开	N/A	10年以上	编辑 移除

数据汇交指标映射

项目成员在对口科学数据中心完成汇交，并审核通过后，数据集信息通过互操作接口同步至数据汇交管理平台，



≡ PART 04

汇交凭证



综合评审并开具汇交凭证

申请阶段性汇交证明

数据汇交专员确认当前阶段汇交任务完成后，通过平台向数据中心提出申请

申请汇交凭证

数据汇交专员确认全部计划指标汇交完成后，通过平台向数据中心提出申请

数据中心综合评审

数据中心检查

- 汇交文档规范性
- 汇交数据形式质量
- 汇交指标完成度与一致性

总中心复核

数据中心工作情况

开具汇交凭证

开具凭证

- 注册CSTR，可验证
- 离线打印盖章寄送

开具阶段汇交证明

开具阶段证明凭证

- 离线打印盖章寄送



数据中心审核数据管理计划

项目管理 流程阶段	名称	说明	工作	审核因素
任务书签订阶段	科学数据管理计划	关于项目拟汇交数据资源的类型、体量、时间、共享权限、保藏计划和对口数据中心等方面的计划	管理方对接 宣讲培训 计划审核	任务书相关指标 支撑任务书主要成果（如论文）的数据 研究方法明确提到的有价值数据 其他有学术价值的的数据，代码，专利，论文等 是否符合年度任务分解 保护条件的必要性，理由是否充分
专项实施阶段	实体数据	科学数据主要包括在自然科学、工程技术科学等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据	技术存储支持 数据、元数据、 附件审核	体量 内容一致性 格式规范性
	元数据	关于数据集的描述信息，如摘要，关键词 关于数据集的著作权属，如作者，贡献者，发布机构 关于数据集的共享权限，如共享权限，许可协议 关于数据集的项目致谢，参考引用等		著录规范性 内容丰富程度 学术道德 与指标一致性
	其他附件	如必要的处理、加工和解析软件工具		
绩效评价阶段	阶段性汇交证明	面向中期评审，年度评审等场景，关于截止当前应完成指标完成情况的证明文件	完备性审核 保藏管理	当前阶段体量 质量
	汇交凭证	对数据管理计划各方面完成情况的综合评审结果	共享发布	体量、质量、流程文件规范性等



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

感谢聆听！ 请批评指正！

