开放科学仓储元数据采集与整理实践调查

开放研究信息巴塞罗那宣言 / COAR 联合调查
可用语言：English | Español | Português | 中文

引言

为什么这项调查很重要

开放科学仓储在学术传播体系中发挥着关键作用，通过保存并提供获取渠道，使各种有价值的研究产出得以长期利用。开放科学仓储收集并开放提供的元数据，为搜索和索引系统发现研究成果提供支持。研究人员、科研机构、资助机构、服务提供商和政策制定者也会利用这些元数据开展报告统计、监测、研究评价以及增值服务开发等工作。

《开放研究信息巴塞罗那宣言》(Barcelona Declaration on Open Research Information)与 COAR 都致力于提升开放科学仓储中的元数据质量。准确、细粒度的元数据对于提高仓储内容的可发现性和可复用性至关重要，也有助于构建可信的学术生态系统。

尽管开放科学仓储能够发挥重要作用，但不同仓储在元数据采集和整理实践方面存在显著差异，尤其体现在元数据的质量、完整性、一致性以及能否开放复用等方面。

调查目的

本调查旨在收集具有可比性的概括性信息，了解开放科学仓储如何采集、整理和提供研究信息(元数据)，并识别元数据整理过程中较为普遍的实践、面临的挑战以及行之有效的方法，以供更广泛的共同体参考。调查结果将有助于提高各方对高质量元数据价值的认识，并为开放科学仓储加强元数据实践提供思路和策略参考。

本调查所涉及的开放科学仓储范围

本调查主要面向机构型、国家级、学科型、政府型以及其他类型的开放科学仓储。这些仓储负责收集、整理、保存并提供学术资源及其相关研究产出元数据。

谁适合填写本调查

开放科学仓储管理人员、仓储馆员、技术负责人以及元数据整理人员。

理想情况下，问卷应由充分了解开放科学仓储工作流程、元数据采集和整理实践，以及研究信息系统和基础设施相关决策的人员填写。

注意：如果您的机构运营多个开放科学仓储，请针对每一个仓储分别填写一份问卷。

预计填写时间：20-30分钟

调查开放时间：2026年11月3日 - 2026年12月15日

数据使用和隐私

您的回答将仅用于本调查所述目的，并将按照保密原则进行处理和安全存储。调查结果可能被纳入汇总或匿名化分析。

部分调查结果可能公开发布，例如用于报告或演示，但不会披露任何个人信息或可识别身份的信息。

您的回答不会用于市场营销、广告或推广活动。我们可能会就您的回答提出简短的后续问题，以便进行核实或澄清。您可以在表单最后一个问题中选择是否同意我们后续联系您。

如有问题，请联系：survey@barcelona-declaration.org

第0部分 开放科学仓储及受访者基本信息

开放科学仓储信息

以下问题用于了解您的开放科学仓储的基本情况。

0.1 机构名称

0.2 国家

0.3 开放科学仓储公开名称

0.4 开放科学仓储主页 URL

（请填写仓储主页，而不是所使用的软件/平台网址）

受访者信息

以下问题用于了解答题人的基本情况，并使我们能够在对您的开放科学仓储存在疑问时与您联系。这些信息不会公开。

0.5 您的姓名

0.6 您的职务 / 职称

0.7 联系电子邮箱（可选）

第1部分 开放科学仓储的收录范围与关键实践

本部分旨在了解您的开放科学仓储的基本情况，包括其收录范围和所管理的研究产出类型，同时重点了解永久标识符和元数据方面的关键实践。

1.1 开放科学仓储的收录范围（可多选）

请指出您的开放科学仓储包含哪些类型的研究产出。如果以下选项未包含某种类型，请选择“其他”，并作简要说明。

□ 正式出版物（如期刊论文、图书、会议成果）
□ 学位论文
□ 研究数据
□ 软件/代码
□ 其他研究产出（如报告、研究方案、数据管理计划〔DMP〕、非文本型成果）
□ 其他（请说明）：_____

1.2 您的开放科学仓储目前大约包含多少条记录？（单选）

○ 少于10,000条
○ 10,000–99,999条
○ 100,000–999,999条
○ 100万条或以上
○ 不确定

1.3 记录层面的永久标识符

您的开放科学仓储是否在单条资源层面分配或登记永久标识符（PID），例如DOI、Handle、ARK？（可多选）

□ 开放科学仓储为存入的资源分配新的 PID
□ 开放科学仓储记录存入资源已有的 PID（例如由论文出版商分配的 PID）
□ 否
□ 不确定

1.3.1 如果您的开放科学仓储为记录分配 PID，分配的是哪些类型？（可多选）

条件题：仅当第*1.3*题选择“开放科学仓储为存入的资源分配新的 *PID*”时显示。

□ DOI
□ Handle
□ ARK
□ 其他：_____

1.3.2 如果您的开放科学仓储不分配 PID，原因是什么？

条件题：仅当第*1.3*题回答“否”时显示。

□ 分配 PID 目前不是我们的优先事项

□ 成本因素
□ 技术能力有限
□ 平台功能限制
□ 其他:_____

1.4 实际使用的元数据元素

您的开放科学仓储在多大程度上实际填写以下元数据元素？

请重点考虑这些元素在实际记录中通常有多大比例被填写，而不是仓储系统是否设置了相应字段。

如果适用，您也可以在"其他"中填写一些使用频率较低的元素，例如贡献者 / 贡献声明（如CRediT角色）、数据可用性声明、撤稿或更正信息等。

元数据元素	必填	始终填写，但非必填	经常	有时	很少	从不
题名	□	□	□	□	□	□
作者	□	□	□	□	□	□
ORCID	□	□	□	□	□	□
作者所属机构	□	□	□	□	□	□
ROR	□	□	□	□	□	□
著作权人	□	□	□	□	□	□
许可	□	□	□	□	□	□
COAR资源类型词表	□	□	□	□	□	□
其他资源类型分类	□	□	□	□	□	□
主题词或关键词	□	□	□	□	□	□
项目	□	□	□	□	□	□
资助机构	□	□	□	□	□	□
原始发布日期	□	□	□	□	□	□
摘要	□	□	□	□	□	□
其他:_____	□	□	□	□	□	□

如果您选择了"其他"，请说明还使用了您的存储库中还填充了哪些其他元数据元素。

1.4.1 对于上述元数据填写实践，是否还有其他信息有助于解释您的回答？（选答）

例如，不同资源集合之间存在的差异，或者某些元数据元素仅用于特定类型的研究产出。

1.5 主要使用的元数据方案

请指出您的开放科学仓储主要使用哪些元数据模式描述资源。（可多选）

□ Dublin Core
□ 本地元数据方案
（如有，请具体说明其所依据的标准或元数据方案）:_____
□ DataCite

□ 不确定
□ 无
□ 其他:_____

第2部分　元数据采集与整理实践

本部分请说明您的开放科学仓储如何采集和管理元数据，包括主要采用哪些元数据整理和丰富的实践，以及这些实践在不同仓储记录中使用的一致性程度。

2.1 您的开放科学仓储通过哪些方式采集元数据？（可多选）

□ 通过 API 导入（从其他系统拉取或由其他系统推送）
□ 数据下载/导入（如批量文件、导出文件）
□ 由图书馆/开放科学仓储工作人员手工录入
□ 由作者本人或作者代理人手工录入（如学生或行政工作人员）
□ 其他：_____

2.2 元数据整理与丰富

您的开放科学仓储采用以下哪些元数据整理或丰富实践？
请说明这些实践在仓储记录中通常以多高频率开展。

实践活动	始终	经常	有时	很少	从不
完整性检查	□	□	□	□	□
元数据方案/应用规范验证（检查是否符合元数据方案要求）	□	□	□	□	□
PID 规范化（如检查、清理或规范 DOI、ORCID iD 等标识符）	□	□	□	□	□
元数据规范化（如统一名称、格式或文本字段）	□	□	□	□	□
去重	□	□	□	□	□
错误纠正	□	□	□	□	□
权利/许可与访问权限检查	□	□	□	□	□
元数据丰富（如增加主题分类）	□	□	□	□	□
其他（请说明）	□	□	□	□	□

如果您选择了“其他”，请说明还使用了您的开放科学仓储采用其他哪些元数据整理或丰富实践。

2.3 上述整理/丰富实践是否适用于所有新提交到开放科学仓储的资源？（单选）

○ 是，所有实践都适用于所有提交资源
○ 否，这些实践仅选择性地应用
○ 不确定

2.3.1 如果这些活动是选择性应用的，请简要说明。（简答）

2.4 记录公开之前，是否必须经过开放科学仓储工作人员人工审核批准？(单选)

○ 是
○ 否

2.5 平均而言，开放科学仓储工作人员整理每一条记录需要花费多少时间？(简答)

第3部分　元数据来源与填写方式

本部分请说明您的开放科学仓储是否从本地/内部来源，和/或外部来源获取元数据；如果获取，请说明这些来源中的元数据如何添加到开放科学仓储记录中。作者本人或作者代理人直接录入元数据的情况已经在第2.1题中涉及，因此本部分不再包含。

3.1 元数据来源类别（可多选）

□ 从本地 / 内部来源导入（如 CRIS）
□ 从外部来源导入（如 OpenAlex、OpenAIRE）

3.2 填写或丰富元数据记录的本地元数据来源

对于您的开放科学仓储使用的每一种本地（内部）来源，请指出元数据通常如何导入或添加到仓储记录中。可多选。

如果某一来源未使用，请将该行留空。

本地来源	手工录入	批量上传（如CSV/XML）	API	收割（如OAI-PMH、OpenAIRE）	基于标识符导入	其他
机构 CRIS/科研人员信息系统	□	□	□	□	□	□
图书馆目录	□	□	□	□	□	□
研究生院（用于学位论文）	□	□	□	□	□	□
机构出版平台/出版机构	□	□	□	□	□	□
其他内部来源（请说明）：_____	□	□	□	□	□	□

3.2.1 关于贵仓储库使用的外部元数据来源或导入方式，您是否还有其他需要补充的信息？（选填）

例如，如果贵仓储库对某一数据来源采用了多种导入方式，请注明该数据来源，并列出题项中未选择的其他导入方式。

3.3 填写或丰富元数据记录的外部元数据来源

对于以下每一种来源，请说明元数据通常如何添加到开放科学仓储记录中。可多选。

如果某一来源未使用，请将该行留空。

外部来源	手工录入	批量上传（CSV/XML）	API	收割（OAI-PMH、OpenAIRE）	基于标识符导入	其他
OpenAlex	□	□	□	□	□	□
Crossref	□	□	□	□	□	□

DataCite	□	□	□	□	□	□
ORCID	□	□	□	□	□	□
国家或国际规范控制文件	□	□	□	□	□	□
国家开放科学聚合平台（包括国家级CRIS）	□	□	□	□	□	□
出版商平台	□	□	□	□	□	□
Scopus	□	□	□	□	□	□
Web of Science	□	□	□	□	□	□
Dimensions	□	□	□	□	□	□
Google Scholar	□	□	□	□	□	□
其他外部来源（请说明）：_____	□	□	□	□	□	□

3.3.1 关于贵仓储库使用的外部元数据来源或导入方式，您是否还有其他需要补充的信息？（选填）

例如，如果贵仓储库对某一数据来源采用了多种导入方式，请注明该数据来源，并列出题项中未选择的其他导入方式。

3.4 您的开放科学仓储中，哪些元数据来源通常需要投入最多的整理工作？（开放回答）

请指出主要来源，并尽可能简要说明原因。

第4部分　元数据开放与复用

本部分请说明您的开放科学仓储中的元数据记录是否允许复用，以及其他用户可以通过哪些方式访问这些元数据，例如接口或下载方式。

4.1 您的开放科学仓储元数据是否公开提供？（单选）

○ 是
○ 部分公开
○ 否
○ 不确定

4.2 用户可以通过哪些接口或方式访问您的元数据？（可多选）

□ OAI-PMH 接口
□ 其他 API（有公开文档的接口端点）
□ 手动/批量下载（如 CSV/XML 导出）
□ 公共 Web界面（UI）
□ 其他：_____
□ 无

4.3 您的开放科学仓储元数据是否采用特定许可协议的方式开放？（单选）

○ 是
○ 否
○ 不确定

4.3.1 如果是，使用的是哪些特定许可协议？（可多选）

□ CC0 /公共领域
□ CC BY
□ 其他 CC 许可
□ 其他：_____

4.3.2 如果是，通过何种方式告知用户相关许可信息？

第5部分　补充信息

本部分邀请您补充说明开放科学仓储中元数据的使用情况，以及元数据导入和整理实践中面临的挑战或存在的机会。

5.1 主要使用场景和用途(开放回答)

请描述贵机构使用开放科学仓储元数据的主要方式，以及这些元数据如何支持机构目标的实现。(例如:资源发现，报告统计，开放获取监测，政策合规，互操作性)

5.2 关于元数据导入和整理实践，您是否还有其他希望补充的内容？(开放回答)