doi:10.3772/j.issn.2095-915x.2020.01.001

# 日本国立信息研究所研究数据基础设施概述

开放科学（资源服务）标识码（OSID）

山地 一禎[1] 李颖[2]

1. 日本国立信息学研究所 东京 101-8430;
2. 中国科学技术信息研究所 北京 100038

**摘要：**自 2013 年发达国家八国集团 G8 科技部长发表联合声明以来，开放科学在全球越来越受欢迎。在日本文部科学省的资助下，国立信息学研究所 NII 开发了新型研究基础设施，建立了管理、存储及发现研究数据的有效模式，有力推进了日本全国性开放科学的发展。本文介绍开放科学的发展现状，并阐述作者见解。

**关键词：**开放科学；新型研究基础设施；研究数据

**中图分类号：**G25；TP39；G350

# Introduction of NII Research Data Cloud in Japan

YAMAJI Kazutsuna[1] LI Ying[2]

1. Research Center for Open Science and Data Platform, National Institute of Informatics,Tokyo, 101-8430, Japan;
2. Institute of Scientific and Technical Information of China, Beijing 100038, China

**Abstract:** The open science movement has been gaining traction internationally since the G8 Science Ministers issued a joint statement in 2013. Under the support from Ministry of Education, Culture, Sports, Science and Technology in Japan, National Institute of Informatics is now developing a new research infrastructure which can managed, stored, and discovered research data, and consequently facilitate open science in Japan. Current development and our perspective on open science is introduced in this article.

**Keywords:** Open science; new research infrastructure; research data

**基金项目：**日本国家科学研究助成事业“研究基础设施数字知识库的模型化及实证”（23300087）；文部科学省“作为新一代大学 ICT 环境的学术云”（962637）。

**作者简介：**山地 一禎，工学博士，教授，研究方向：信息学、媒体信息学与数据库等。山地教授将机构知识库系统以云服务形式向全日本大学及研究机构提供，从而使日本机构知识库构建数量跃居世界第一，为学术信息流通基础设施的开发、日本开放科学的发展做出了贡献，由此获得 2018 年度“文部科学省科技奖”。山地教授是 NII 开放科学基础设施研究中心现职主任；李颖，信息学博士，研究员，研究方向：开放科学，日本科研管理与政策，机器翻译，E-mail：liying@istic.ac.cn。

## 引言

科学研究正转向新范式。新范式下，学术界内外呈现出研究成果与数据开放合作和自由共享的态势。这种新范式通常称为“开放科学”[1-2]。开放科学可加速科学研究的进步，助力研究人员应对当今社会的各种挑战。

伴随全球开放科学运动的兴起，日本政府启动研究数据相关基础设施的规划。有关建议由日本科学委员会 2016 年提出[3]，成型于日本《科学技术五年基础规划》[4]。日本科学技术与创新委员会（CSTI, Council for Science, Technology and Innovation）为日本 2016-2020 年五年科学技术政策和实施指明了路径。其中，开发研究数据基础设施并向所有大学提供服务的期望寄托于 NII。NII 为此设立了开放科学和数据平台研究中心（RCOS, The Research Center for Open Science and Data Platform），目标设定为开发和管理研究数据的基础设施，为日本“开放科学”奠定基础。

## 1 NII 研究数据云

NII 研究数据云定位 e- 基础设施。以此管理、存储和发现研究数据及其他相关文档。图 1 为 NII 研究数据云的概念图。基础设施由三个特色平台组成：①研究数据管理平台 GakuNin RDM①；②知识库平台 WEKO3②；③发现平台 CiNii Research③。三个平台为不同研究阶段的研究人员服务：①研究人员或研究小组在研究项目开始时可在研究数据管理平台 GakuNin RDM 中建立项目，并在此平台上存储、管理和共享研究合作者之间的文档；②研究项目完成后，确定可公开的研究成果，并将相关文档复制到知识库平台 WEKO3；③知识库平台上的学术资源通过发现平台 CiNii Research 可被发现，随后可被其他研究人员用于新思路的构思及为自身研究寻求研究素材。

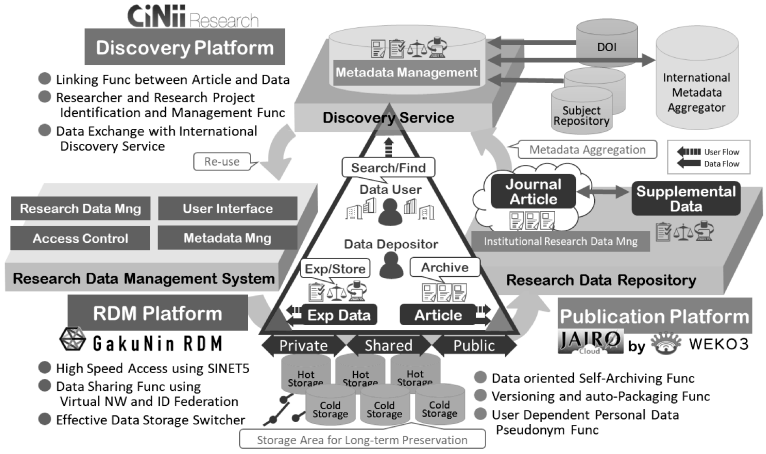


图 1 NII 研究数据云架构图

① GakuNin RDM：GakuNin 含义是学术认证联邦机制 , RDM 是 research data management 简称，研究数据管理。
② WEKO：非洲语言斯瓦希里语 repository（知识库，仓储）之意。
③ CiNii Research：CiNii 是 NII 学术信息导航系统，即 Citation Information by NII。

doi:10.3772/j.issn.2095-915x.2020.01.001

## 2 研究数据管理平台 GakuNin RDM

研究数据管理平台 GakuNin RDM 提供一种环境。此环境下，研究个体或研究团队可在研究项目实施期间管理研究数据及有关文档。在封闭环境中进行文档版本管理可实现项目成员间的访问控制，跟踪文档修改，提供长期保存，保障研究的完整性。

GakuNin RDM 基于开放科学框架（OSF, Open Science Framework）定制，以满足日本研究界需要。OSF 为一款开源软件，用于管理研究数据，由美国非营利组织开放科学中心（COS）开发提供[5]。图 2 是 GakuNin RDM 项目的首页。OSF 的原始版本允许研究人员通过简易操作发布研究结果，但 GakuNin RDM 不设发布文档功能，除非研究人员主动将其复制到知识库平台 WEKO3 并主动发布出版。通过将 GakuNin RDM 链接到 NII 的 GakuNin 服务，可以启用 Shibboleth④单点登录。

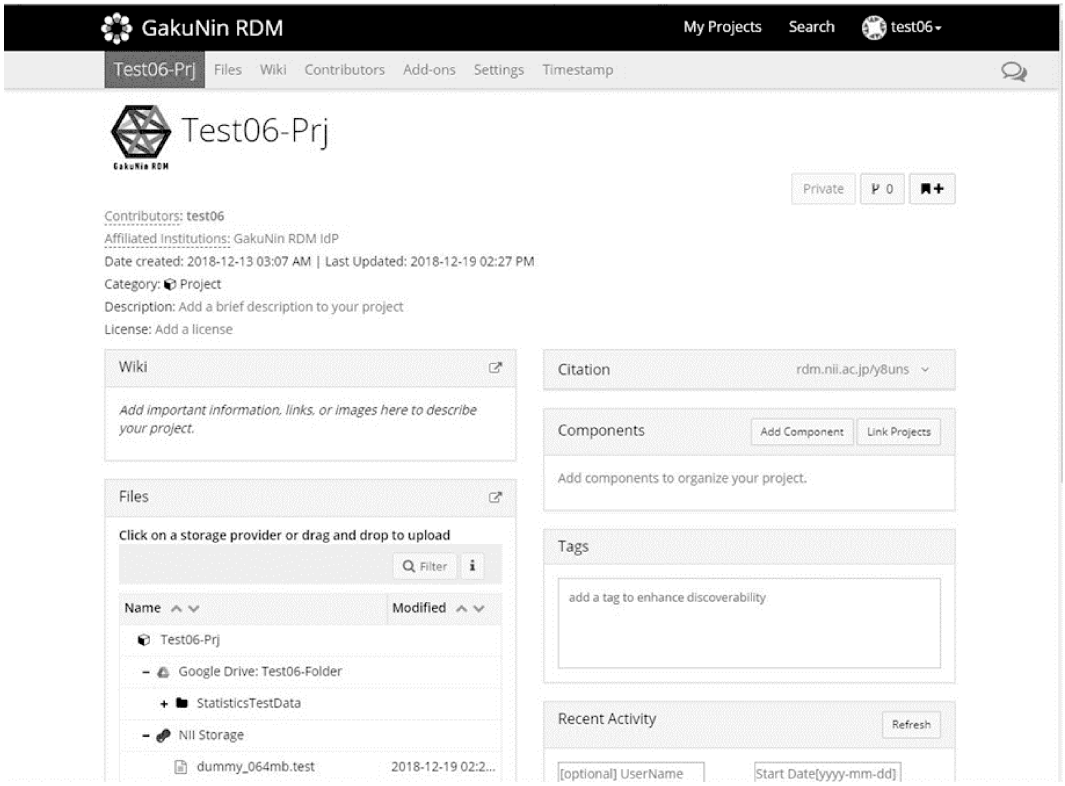


图 2 GakuNin RDM 首页截图

近年，伴随信息系统安全事件不断增多，系统工程师、数据库管理员，以及学术机构研究人员不得不在系统运行与维护方面花费越来越多的时间精力。通过安全层面的支撑，GakuNin RDM 允许研究人员专注于学术工作，同时减少系统管理员的工作负荷与成本。

④ Shibboleth 是一个针对 SSO 的开源项目。Shibboleth 项目主要用于校园内 Web 资源共享，以及校园间的应用系统的用户身份联合认证。

另外，GakuNin RDM 还为日本研究人员提供了安全的存储方式。目前，不同国家间的多机构研究人员合作研究项目增多，使用外部公共云提供免费在线存储很流行，但应避免使用这类服务。因为研究素材很容易散落或丢失，且不能保证长期保存。GakuNin RDM 的内容存储区域，由每位用户所在学术机构提供，如果机构提供多个存储空间，则用户在保存文件之际可从 GakuNin RDM 上选择存储区域。RCOS 与日本各学术机构所属 ICT 中心合作，为每一机构提供定制服务。

在通用框架基础上，用户开发的插件可共享，利用多元化的工具可满足特定需求。

## 3 知识库平台 WEKO3

NII 知识库平台 WEKO3 是学术资源（例如研究出版物、灰色文献、论文、研究数据，以及研究人员决定开放访问的任何其他资源）存储和发布的平台。相关文档存放在研究人员所属机构运维的机构库中。

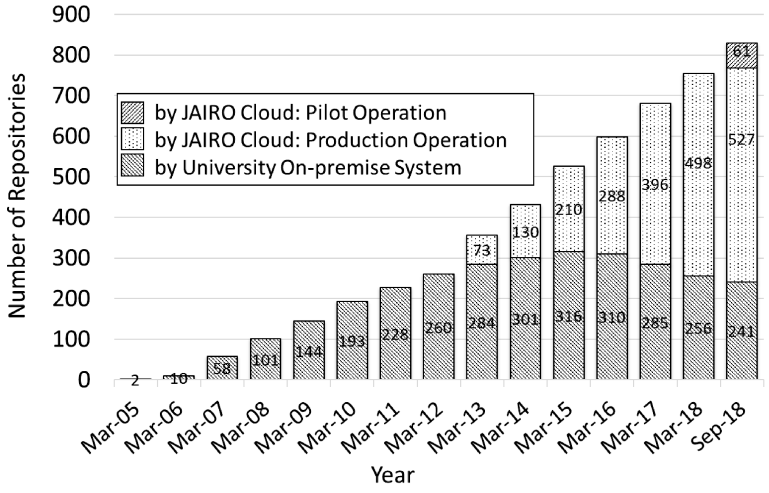


图 **3** 机构知识库在日本增长情况

WEKO3 之前的系统 WEKO2 已用于 NII 提供的机构知识库云服务“JAIRO⑤ cloud service”，即 JAIRO 云服务。如图 3 所示，2012 年 JAIRO 云服务启动后，日本机构知识库的数量急剧增加。通过将各自服务迁移到 JAIRO 云，自行设立机构知识库的机构数量正在减少。现有的 WEKO2 功能主要集中在促进开放存取，除此之外，WEKO3 还有容纳研究数据的功能。为使系统具有灵活与可扩展性，WEKO3 采用了可移植的体系结构。Invenio⑥是欧洲核研究组织（CERN）开发和使用的开放源码软件，用于管理机构知识库中的数字内容，用作基础系统[6]。平台提供公共和私人空间，以满足不同的内容共享需求。比如，一定期限后（embargo periods）提供、在应用基础上提供、或对部分数据进行匿名化后提供。

⑤ JAIRO：Japanese Institutional Repositories Online，日本机构知识库检索网站。
⑥ Invenio: https://invenio-software.org/

doi:10.3772/j.issn.2095-915x.2020.01.001

NII 知识库平台 WEKO3 确保在日本安全存储由日本资助机构支持的研究产出的研究数据。目前，学术期刊出版发行也要求研究论文与有关证据数据一并提交，为满足这类要求，许多研究人员通常采用学术期刊的补充形式或互联网免费知识库形式提供其研究数据。然而，这种自由的空间不能保证长期保存，有价值的研究数据还会转移到国外，造成出版商控制这些研究数据。WEKO3 为日本产出的研究数据提供安全的基础设施。

研究人员可使用 WEKO3 发布数据。人们越来越认识到研究数据与研究出版物本身同样重要，甚至更重要。不同的研究人员从不同角度分析研究数据时，会带来更多的学术成果。因此，现在出现了“数据期刊”，如同研究出版物一样发布研究数据；针对每篇数据文章说明数据获取的条件、数据位置、文件格式，以及如何使用数据信息等。在数据期刊上发表的数据文章可供其他研究人员开展新的研究项目，并在相关出版物中得到引用，提高对数据成果的认知。

## 4 发现平台 CiNii Research

CiNii Research 旨在搜索和发现日本研究项目产生的出版物和数据集。如图 4 所示，CiNii Research 由三个部分组成：一是收集与日本研究项目相关的元数据；二是从聚合元数据中提取研究实体及其关系，构建大规模的学术知识图谱；三是通过对知识图谱节点索引为研究实体提供发现服务。另外，这些数据将被分配 DOI（唯一标识符），使其可检索和可重用。

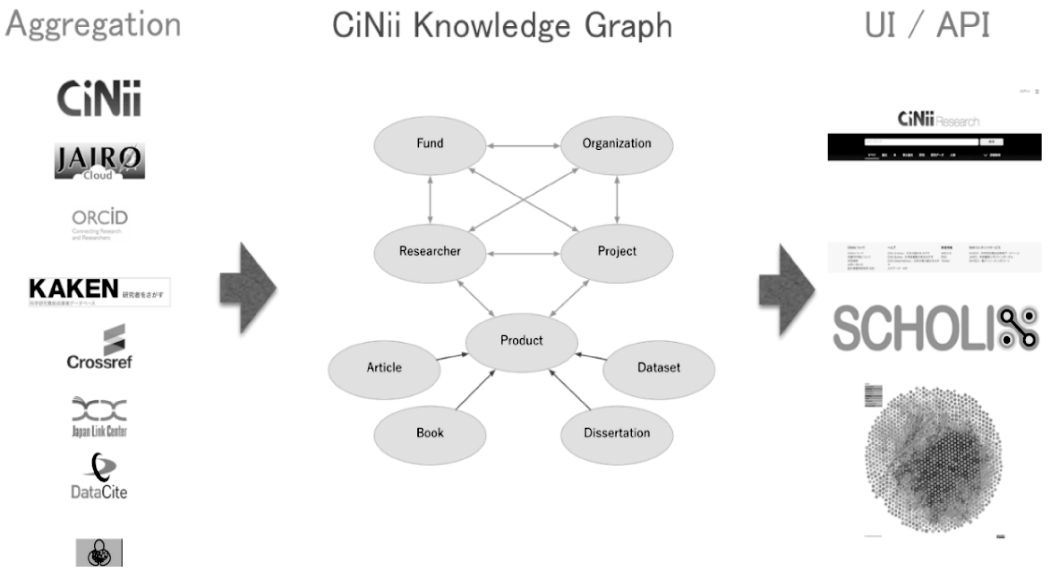


图 4 CiNii Research 构成主要三要素

第一部分，收集日本研究项目相关研究实体元数据。NII 与日本大学等机构合作，收集 NII 学术信息服务有关元数据。IRDB 是全国性机构知识库聚合器，截至 2018 年 9 月，合计从 685 个知识库 290 万条记录数据聚集了 5.5 万条（占总数的 2.5%）。CiNii 使用 IRDB 收集元数

据，并开发采用 JPCOAR Schema 1.0[7]，升级的 IRDB，JPCOAR Schema 1.0[7] 是日本机构知识库最新元数据模式，能支持永久性唯一标识符、开放获取策略等新功能。

另一聚合器是 KAKEN[8]，它负责收集科学研究资助（科研费）的成果报告，KAKEN 是日本政府的主要研究基金之一。NII 已开始收集像 DOI 这样的永久性标识符。日本链接中心（JaLC）[9] 是日本注册机构，NII 是 JaLC 的董事会成员之一。因此，JaLC 将是主要 DOI 来源，日本知识库使用 JaLC 将 DOI 分配给包括研究数据集在内的研究实体。

构建研究实体知识图谱是当今发现服务的重要组成部分，研究实体间的关联有助于找到更多的相关研究实体。图 4 描述了目标实体类型，包括成果、研究员、项目、组织和资金。成果定义为文章、书籍、论文和数据集的超集。CiNii Research 目前主要聚焦于成果、研究人员和项目，从元数据中提取研究实体，并使用永久标识符和名称消歧技术对其进行识别。知识图谱对与其他发现服务的全局协同也很重要，scholaix[10] 为学术文献和数据集间的互链提供了互操作框架，OpenAIRE 提供 LOD[11] 实现共享链接，本研究今后计划与国际性组织共享链接。

CiNii Research 提供按关键字搜索的简单输入表单，用户可在搜索关键字之前从选项卡中选择类型，如果用户选择“数据集”选项卡，则仅对数据集过滤搜索。CiNii Research 不支持典型的分面检索以保持搜索结果尽可能简单。

发现服务链接到研究数据管理平台 GakuNin RDM，从搜索结果中找到特定的研究数据后可导入。

## 5 全球研究数据基础设施网络

如图 3 所示，日本已经拥有 800 多个知识库，可在学术流通网络中形成超分布式体系结构。所有元数据聚合到机构知识库数据库（IRDB）中，IRDB 是 CiNii Research 的核心后端服务之一。CiNii Research 和 OpenAIRE 之间的国际合作使本系统从日本聚合节点（IRDB）向全球发现服务（OpenAIRE）提供数据。CERN 和 NII 之间存储系统的国际发展也影响 CERN 运营的大型开放数据知识库 Zenodo 及本研究数据基础设施之间的无边界交互。

NII 研究数据云将作为基础设施的顶层服务运行。作为日本国家研究和教育网络（NREN），NII 为 800 多所大学和研究机构提供名为 SINET 的学术骨干网。在主干网之上 GakuNin 允许为联邦身份和访问管理建立可信网络，GakuNin 云支持大学与研究机构使用云，NII 研究数据云存在这些中间件服务之上，类似架构在欧洲开放科学云中看到，由欧盟的地平线 2020 基金[12]、非洲开放科学平台[13] 和澳大利亚研究数据共享平台[14] 推广。基于这些国家云国际一级的框架，本开发将促进研究数据的全球性获取。

## 6 结论

日本有责任向公众提供国家公共资金资助的研究成果，国家还应负责为国内产出学术资源提供安全的基础设施，以确保这些资源不会随着时间的推移而分散丢失。NII 研究数据云提供相关基础设施，在此基础设施上，国内产出的学术资源将被安全存储、长期保存。从而可

doi:10.3772/j.issn.2095-915x.2020.01.001

向公众提供日本的学术资源，并向国内外有关各方展示日本的研究成果。

## 参考文献


[1] Pontika N, Pearce S, Knoth P. Fostering Open Science to Research using a Taxonomy and an eLearning Portal [C]. iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business, October 21-22, 2015, Graz, Austria. 2015.

[2] OECD. Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25[R]. OECD Publishing, Paris, 2015.

[3] Science Council of Japan (SCJ). Recommendations Concerning an Approach to Open Science that Will Contributes to Open Innovation [EB/OL]. [2018-12-15]. http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230-en.pdf

[5] The Council for Science. Technology and Innovation (CSTI), The 5th Science and Technology Basic Plan (in Japanese)[EB/OL]. [2018-12-15]. https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html

[6] Nosek B A. Center for Open Science: Strategic Plan, Open Science Framework, 2017. [EB/OL]. [2018-12-15]. https://osf.io/ymu94/

[7] CERN. CERN collaborates with Japan's NII on digital libraries [EB/OL]. [2018-12-15]. https://home.cern/news/news/knowledge-sharing/cern-collaborates-japans-nii-digital-libraries.

[8] JPCOAR Schema [EB/OL]. [2018-12-15]. https://schema.irdb.nii.ac.jp/en

[9] Database of Grants-in-Aid for Scientific Research[EB/OL]. [2018-12-15]. https://kaken.nii.ac.jp/en/

[10] Japan Link Center [EB/OL]. [2018-12-15]. https://japanlinkcenter.org/top/english.html

[11] Burton A, Koers H, Manghi P, et al. The Scholix Framework for Interoperability in Data-Literature Information Exchange. D-Lib, 2017, 23(1).

[12] Alexiou G, Vahdati S, Lange C, Papastefanatos G, Lohmann S. OpenAIRE LOD Services: Scholarly Communication Data as Linked Data[A]. In: González-Beltrán A, Osborne F, Peroni S. Semantics, Analytics, Visualization. Enhancing Scholarly Data[M]. Berlin: Springer Cham, 2016.

[13] European Open Science Cloud (EOSC) [EB/OL]. [2018-12-15]. https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[14] African Open Science Platform [EB/OL]. [2018-12-15]. http://africanopenscience.org.za/

[15] Australian Research Data Commons [EB/OL]. [2018-12-15]. https://ardc.edu.au/