

2022.11.11 RDUF公開シンポジウム

データ共有・公開制度検討部会 活動報告（2022）



研究データ利活用協議会（RDUF）
情報・システム研究機構 国立情報学研究所
南山 泰之

minamiyama@nii.ac.jp

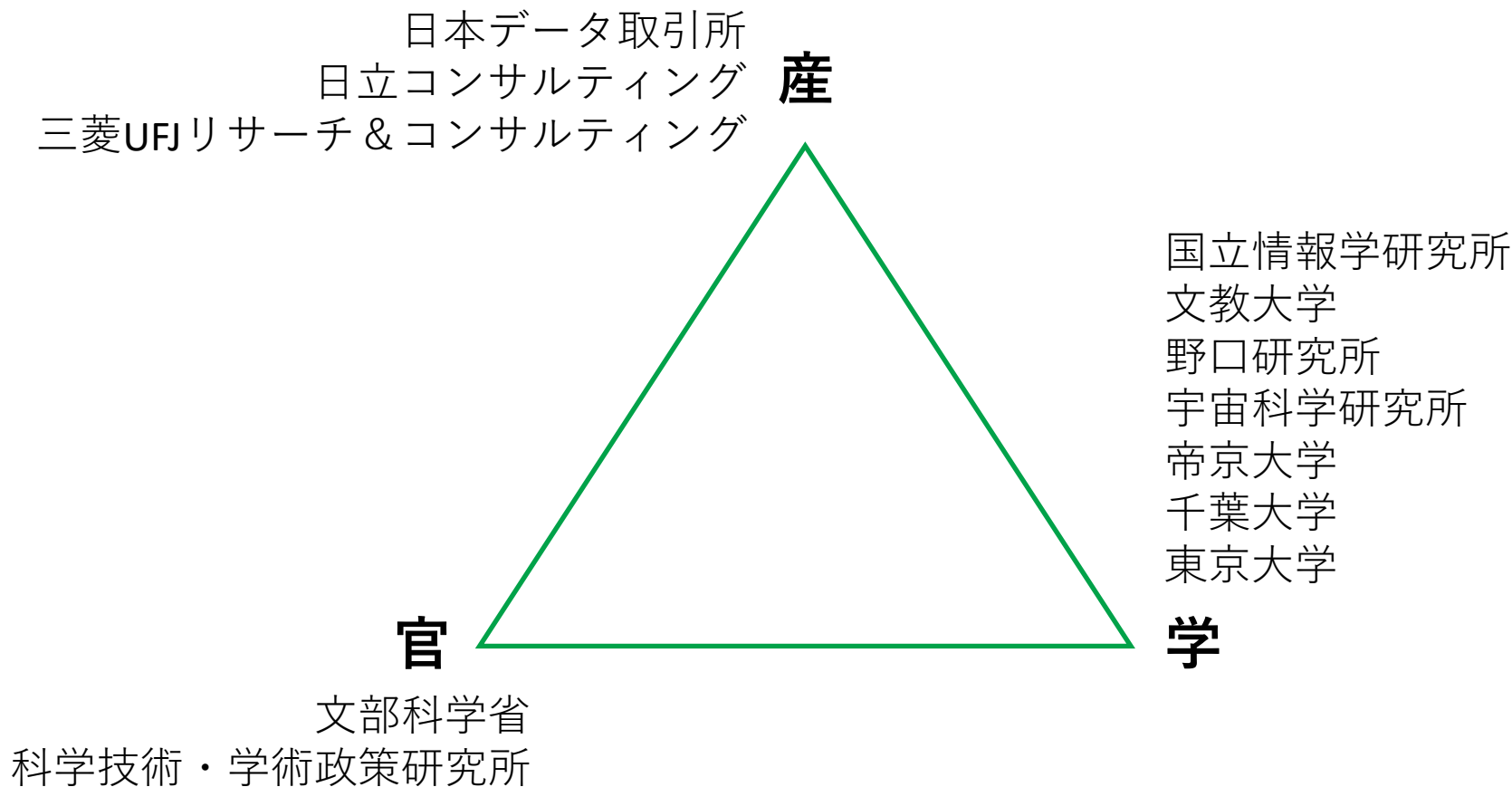
ORCiD ID:0000-0002-7280-3342

活動の概要

1. 「RDUF 研究データライセンス小委員会」による活動の継承と展開
 - 「研究データの公開・利用条件指定ガイドライン」の改訂
 - 英訳版の作成

2. 研究データ利活用にまつわる法的・制度的課題に関する論点の検討
 - データの品質保証・品質表示
 - データオーナーシップ

メンバー構成



https://japanlinkcenter.org/rduf/permanent_subcommittee/index.html

活動紹介①

「RDUF 研究データライセンス小委員会」による活動の
継承と展開

研究データの公開・利用条件指定ガイドライン (2020)

2019年12月25日

研究データの公開・利用条件指定ガイドライン
研究データ活用協議会 研究データライセンス小委員会

目次

研究データの公開・利用条件指定ガイドライン	1
はじめに、研究データの公開と利用条件指定に関する5の質問	2
Q1. 公開対象とするデータの特長	5
データ公開が義務付けられている場合	6
Q2. データ公開の制約条件の確認	8
分野・研究コミュニティの慣習などで、公開制限が一般的な場合	8
個人情報を含む場合	8
国家安全保障、国際関係などに係る場合	8
共同研究契約や個別の契約により公開制限が定められている場合	8
所属機関（部署）、研究助成機関などによるポリシー	8
Q3. 公開制約条件の解除	13
分野・研究コミュニティの慣習などで、公開制限が一般的な場合	13
個人情報を含む場合	13
国家安全保障、国際関係などに係る場合	13
共同研究契約や個別の契約により公開制限が定められている場合	13
所属機関（部署）、研究助成機関などによるデータポリシーが定められている場合	13
公開できないと判断された場合	14
Q4. 公開先の選択	15
分野別リポジトリ	15
所属先のリポジトリ（機関リポジトリ）	16
Q5. 利用条件の指定（選択肢と表示例）	19
推奨される利用条件	20
（参考）利用規約の記載	29
参考文献一覧（本文中に含まれるものを除く）	31

**英訳版作成済；近日中に公開予定
情報の改訂は活動2に統合**

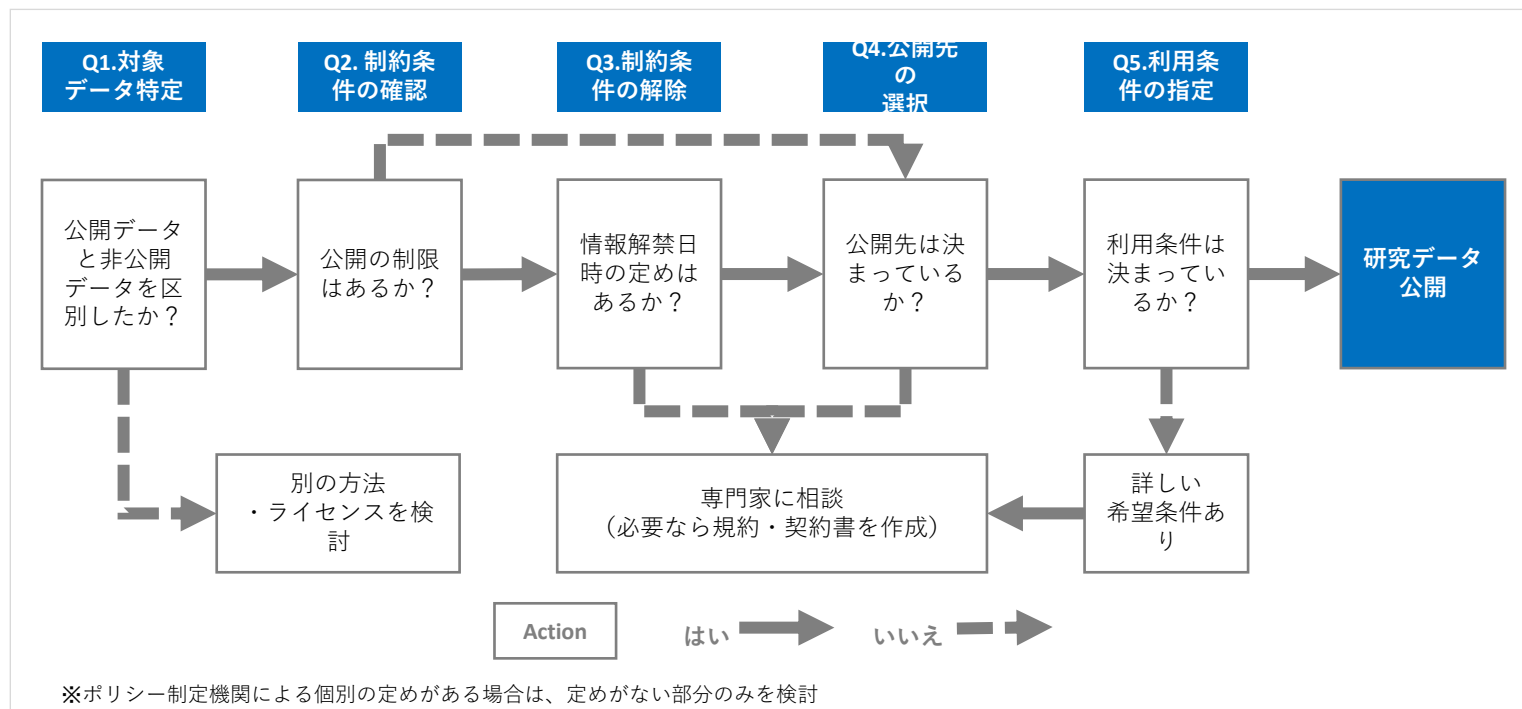
別添：研究データの公開・利用条件指定ガイドライン用語集			
No.	項目名	定義	出典
1	改竄（改ざん）	研究資料・機器・過程を変更する操作を行い、データ、研究活動によって得られた結果等を真正でないものに加工すること。	文部科学省、研究活動の不正行為等の定義。 https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu12/houkoku/attach/1334660.htm ,（参照2019-12-25）。
2	クリエイティブ・コモンズ	著作権者に自身の作品の使用条件を簡便なマークで意思表示する手段を提供し、著作物の円滑な流通、再使用を促進するプロジェクト、またはそれを推進する非営利団体の名称。著作権	日本図書館情報学会用語辞典編集委員会編、図書館情報学用語辞典 第4版、2014。
		ること。	DA-CODATA: Legal Interoperability Interest Group. Legal Interoperability of Research Data. Principles and Implementation Guidelines. Zenodo, 2016. https://doi.org/10.5281/zenodo.162241 , (accessed 2019-12-25).
4	研究データ	科学研究の情報源として利用されるデジタルデータ。数値、テキスト、画像、音声、動画など、さまざまな形式を含む。なお、本ガイドラインでは、研究ノート、論文、メタデータ、	1) 研究データ基盤整備と国際展開ワーキング・グループ報告書、研究データ基盤と国際展開に関する戦略、2019。

本編 (32p.)

用語集 (7p.)

“研究データの公開・利用条件表示ガイドライン ver.1.0”。
https://doi.org/10.11502/rduf_license_guideline

研究データの公開・利用条件指定フロー



- 研究者の視点から、公開先や利用条件についてまとめたもの
- 5つの質問に回答していくことで、適切な公開場所と利用条件を選択できる

Minamiyama, Y., et al. Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan. Data Science Journal, 19(1), p.53. DOI: <http://doi.org/10.5334/dsj-2020-053>

普及活動

学協会

- J-STAGE Data データポリシー(利用機関向け)

https://www.jstage.jst.go.jp/static/files/ja/pub_JstageData_policy_for_members.pdf

- 情報科学技術協会「情報の科学と技術」データ共有・公開ポリシー

https://www.infosta.or.jp/wp/wp-content/uploads/2021/07/jkg_datapolicy_ver1.pdf

大学・研究機関

- 国立大学図書館協会. 「研究データのオープン化とそのメリット」

<https://www.janul.jp/ja/news/20200615>

- 京都大学の部局等における研究データ管理・公開に関する実施方針策定のためのガイドライン・ひな形(ガイドライン):

<http://hdl.handle.net/2433/262758>

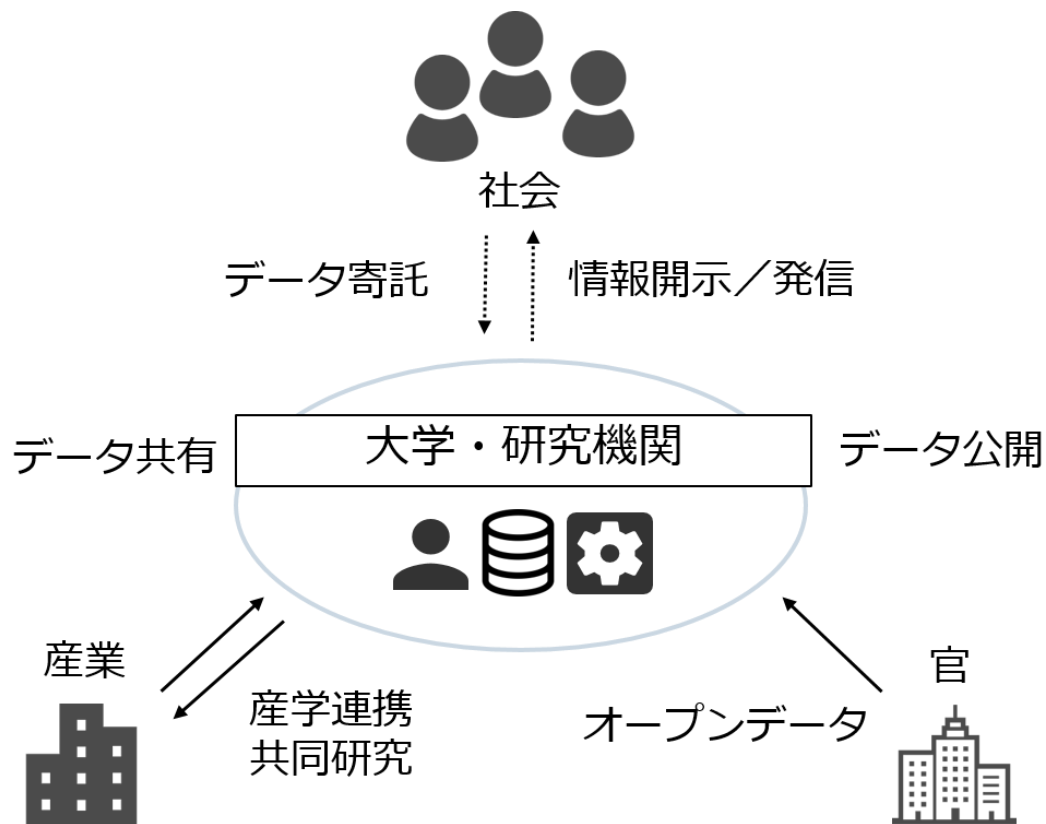
- 北海道大学附属図書館ウェブサイト「研究データの公開と管理」:

https://www.lib.hokudai.ac.jp/support/research_data/

活動紹介②

研究データ利活用にまつわる法的・制度的課題に関する
論点の検討

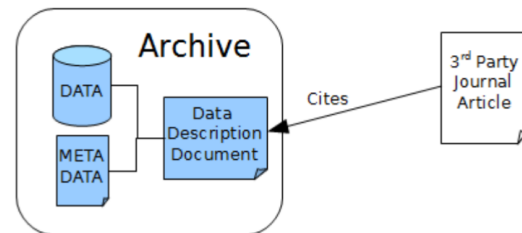
データ共有・公開と利活用の ギャップ



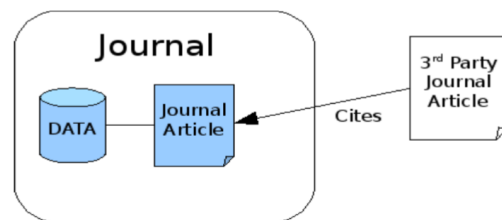
**研究データは大学・研究機関の外で
どのように利活用される可能性があるのか？**

研究データの共有・公開と品質管理

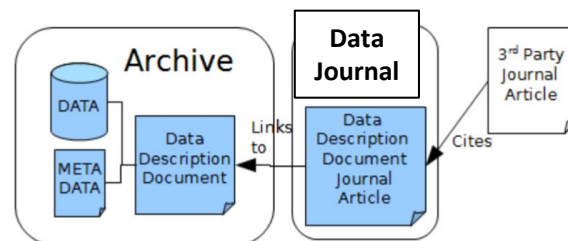
1) Data is published in a data repository



2) Data is published an appendix to a paper



3) Data is published as a data paper



- 品質管理方針はシナリオによって異なる
 - 標準的な品質表示基準を示せないか？

Ref: Mayernik, M. S., et al. (2015). Peer review of datasets: When, why, and how.

研究データの品質表示？

多くのデータ論文に含まれる記述様式

No.	項目	説明
1	基本情報	所有者、名前、DOI/URIなど
2	利害関係	データセットに関係する個人的/組織的な関係性
3	範囲	空間的/時間的
4	形式	フォーマット、エンコード、言語など
5	ライセンス	
6	詳細な帰属	各著者ごとの貢献度
7	プロジェクト	
8	データセットの来歴	取得方法、機材などを含む
9	品質	データの限界値・異常値などの情報
10	再利用	潜在的な価値

全分野に共通した表示様式は議論中

Ref: <https://current.ndl.go.jp/ca1858>

関連標準／指針

➤日本の省庁によるガイドライン・規則

- 総務省統計局「公的統計の品質保証に関するガイドライン」
- 国土地理院「品質の要求，評価及び報告のための規則」
- デジタル庁「データ品質管理ガイドブック（β版）」

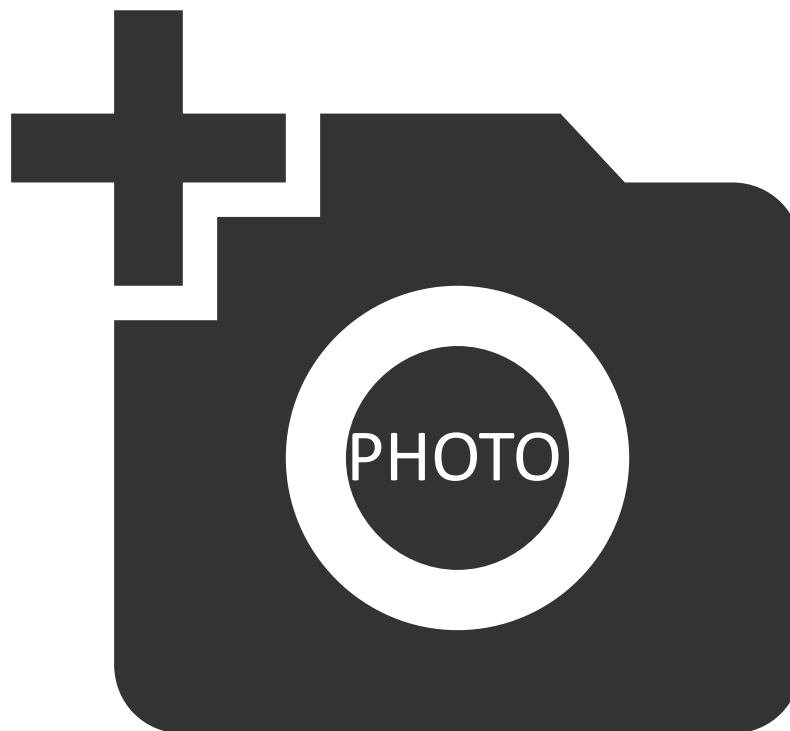
➤ビジネスデータ活用の知識体系

- DAMA-DMBOK（データ品質の項目）
- データ管理成熟度モデル（DMM）

➤学術コミュニティ

- 明確な標準なし（分野別の規定はある）

JOSS2022での議論



主な論点：研究データの管理責任／及びデータ利活用の拡大
ポイント：事前に定めたプロセスは適切か／定められたプロセスを
たどることで結果が再現可能か

当日資料：https://joss.rcos.nii.ac.jp/session/overview/?id=se_110

開催報告：https://doi.org/10.2964/jsik_2022_027

案：研究データの品質表示ガイドライン（仮称）の策定

データの品質

サービスの品質

プロセスの品質

#	データ品質の指標	評価項目	問題となる例
1	正確性 (Accuracy)	データの基本は正確であること。データの正しさは、以下の点に着目して評価する。 •書式が正しいか。 •誤字脱字などはないか。 •意味的な誤りがないか。 •データに誤りはないか。	•同上、〃などの記述がある。 •日付や数字が記述されるべき欄に「不明」など数字以外の文字列が記述されている。 •住所が記述されるべき欄に電話番号が記述されている。 •フリガナ欄にカタカナとひらがなが混在している。
2	完全性 (Completeness)	データは目的に応じて抜け漏れなく存在することで、詳細な分析をすることができるようになる。データが完全であることを以下の点に着目して評価する。 用途に応じて必要な項目が網羅されている	•重要なデータ項目が定義されていない。 •データが取得できないという理由で必須項目に空欄がある。
3	品質表示に関する仕様と考慮事項をまとめたガイドライン（案）を執筆中		
		下の点に着目して評価する。 •データセット内でデータに矛盾はないか。 •データセット間でデータに矛盾はないか。	データに含まれていた合計値が一致しない。

まとめ

1. 「RDUF 研究データライセンス小委員会」による活動の継承と展開
 - 「研究データの公開・利用条件指定ガイドライン」の改訂
 - 活動2に統合
 - 英訳版の作成
 - 近日中に公開予定（GitLab / GitHub Pages?）
2. 研究データ利活用に関わる法的・制度的課題に関する論点の検討
 - データの品質保証・品質表示
 - データオーナーシップ
 - ガイドライン草案の執筆を通じ、論点を整理中