

データ共有・公開制度検討部会 活動報告（2023）

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



研究データ利活用にあつわる法的・制度的課題 に関する論点の検討

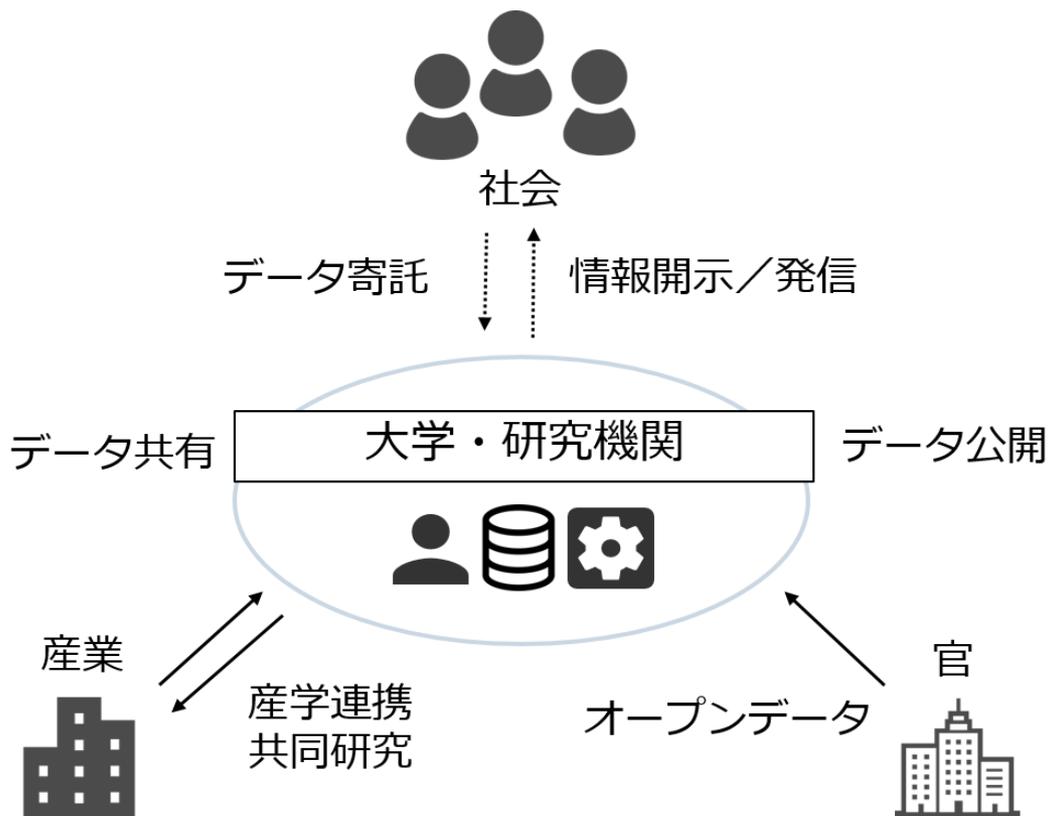
➤ 実務的な観点から

- 既存のガイドライン／基準のレビュー
- 利用シナリオの抽出
- ガイドライン草案の策定

➤ 理論的な考察

- 品質に関する相互認証のメカニズム
- 信頼性の判断基準

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？**

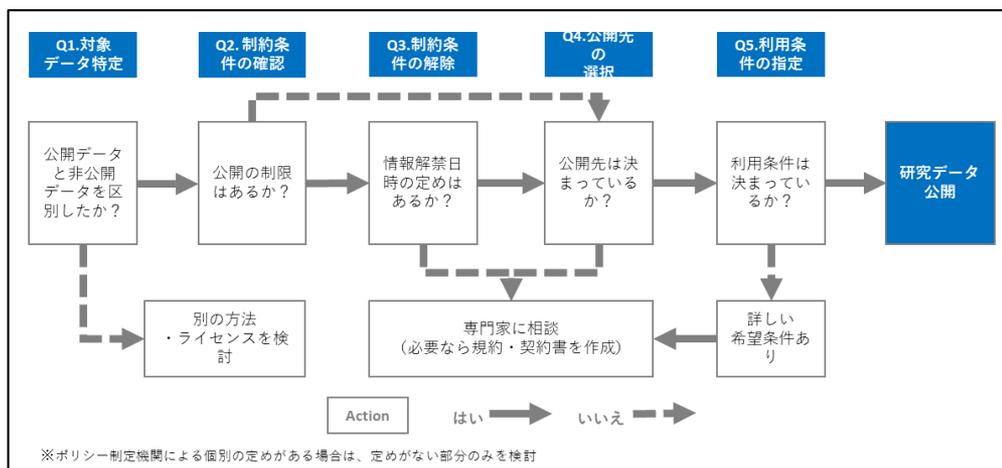
これまでの活動：研究データの公開・利用条件表示ガイドラインの策定

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定	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	9
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	9
所属機関（部署）、研究助成機関などによるポリシーが定められている場合	10
Q3. 公開制約条件の解除	11
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	11
個人情報を含む場合	12
国家安全保障、国際関係などに係る場合	12
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（選択表と表示例）	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



- 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

“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

普及状況

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):

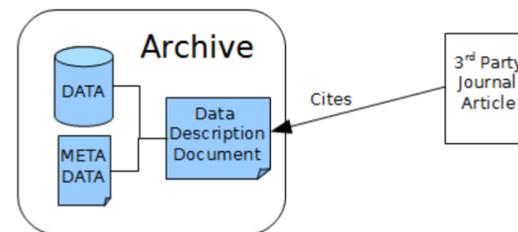
<http://hdl.handle.net/2433/262758>

- 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」:

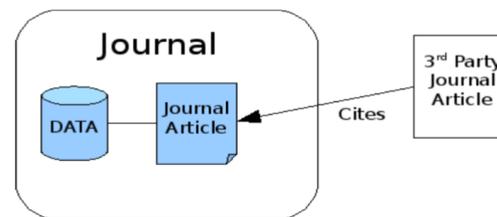
https://www.lib.hokudai.ac.jp/support/research_data/

研究データの共有・公開と品質管理

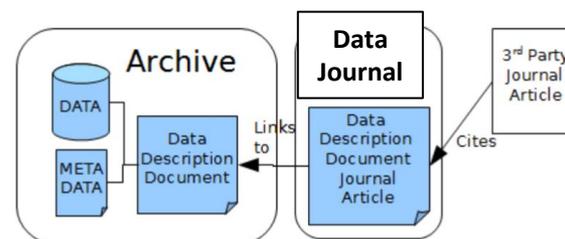
1) Data is published in a data repository



2) Data is published as 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

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）」
 - 内閣府地方創生推進事務局「データ連携基盤を通して提供されるデータの品質管理ガイドブック」←**New!**

- 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 / 2023での議論



【JOSS2022】 研究データの品質表示を考える
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

【JOSS2023】 コンテキストを説明するメタデータ
主な論点：「品質」に対する研究データ提供側と
利用側の視点の違い
ポイント：既存の「正確性」「再現性」に留まら
ない、工学的な評価情報をどう提供するか



開催報告：情報知識学会誌32巻3号（2022）／33巻3号（2023）

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。 <ul style="list-style-type: none"> •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。 	<ul style="list-style-type: none"> •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。 <ul style="list-style-type: none"> •用途に応じて必要な項目が網羅されているか。 	<ul style="list-style-type: none"> •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		<p>「信頼性」の観点から品質に関する考慮事項を整理 今後、ヒアリング調査／アンケート調査を通じて利用シナリオをブラッシュアップする予定</p>	
		<ul style="list-style-type: none"> •データセット間でデータに矛盾はないか。 	

研究データ利活用につわる法的・制度的課題 に関する論点の検討

➤ 実務的な観点から

- 既存のガイドライン／基準のレビュー (Done)
- 利用シナリオの抽出 (Ongoing)
- ガイドライン草案の策定 (Ongoing)

➤ 理論的な考察

- 品質に関する相互認証のメカニズム (Ongoing)
- 信頼性の判断基準 (Ongoing)