

機関リポジトリでの 研究データ公開

情報・システム研究機構 国立情報学研究所
オープンサイエンス基盤研究センター

南山 泰之

minamiyama@nii.ac.jp

ORCID ID:0000-0002-7280-3342



自己紹介

経歴 8

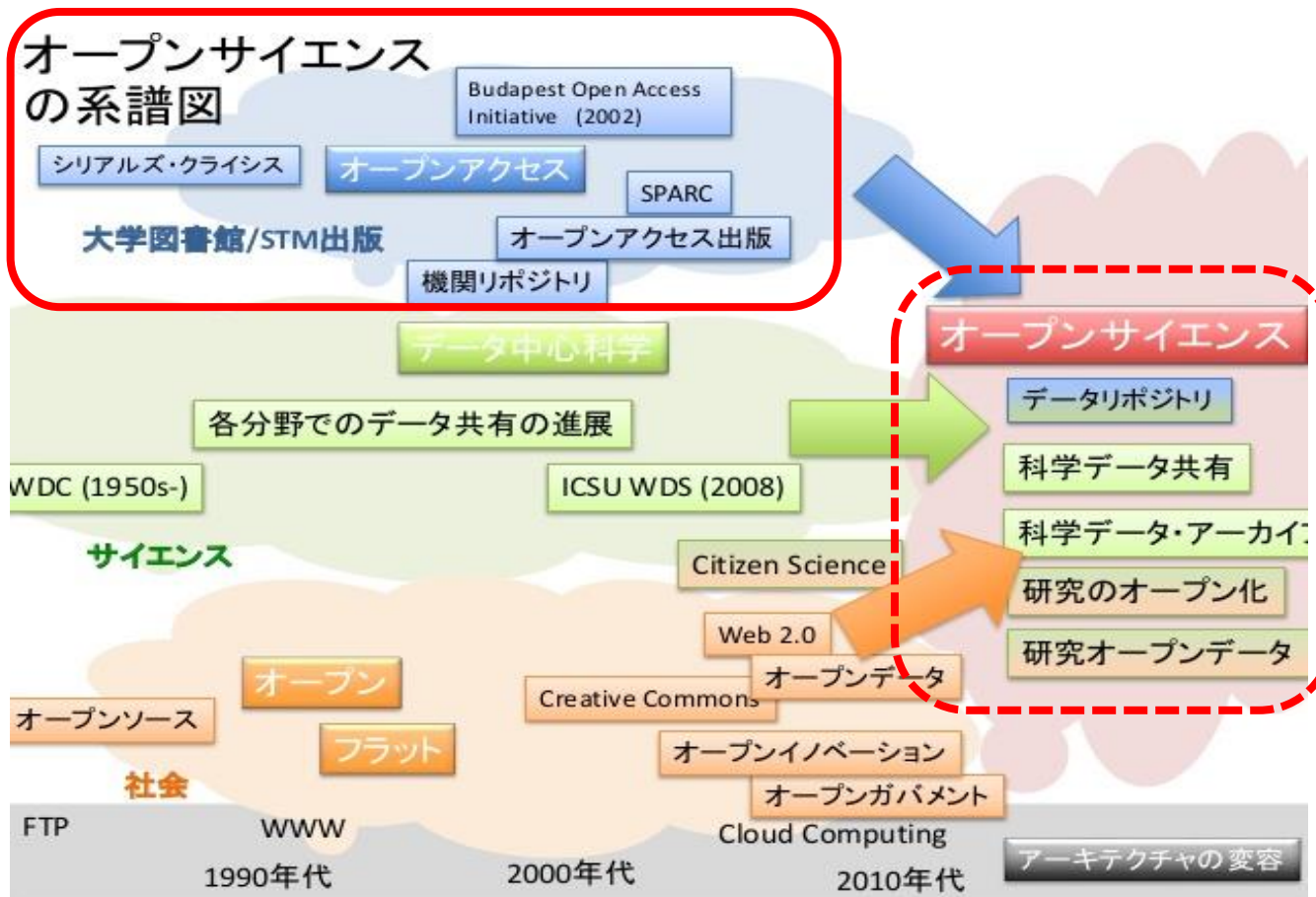
2022年9月 - 現在	国立情報学研究所 オープンサイエンス基盤研究センター データキュレーション基盤担当
2019年9月 - 2022年8月	国立情報学研究所 オープンサイエンス基盤研究センター 図書館戦略担当
2018年10月 - 2019年8月	東京財団政策研究所 政策データラボ データ・ライブラリアン
2017年4月 - 2018年9月	情報・システム研究機構国立極地研究所 アーカイブ室
2014年4月 - 2018年9月	情報・システム研究機構国立極地研究所 情報図書室
2011年4月 - 2014年3月	東京大学教養学部等 図書課（駒場図書館）
2005年9月 - 2011年3月	情報・システム研究機構国立極地研究所 情報図書室
2007年7月 - 2008年3月	第49次日本南極地域観測隊 夏隊 設営一般（庶務）

目次

1. 背景
2. 研究データ公開における課題：データキュレーション
3. 考察と実践
4. 今後の展開
5. まとめ

1. 背景

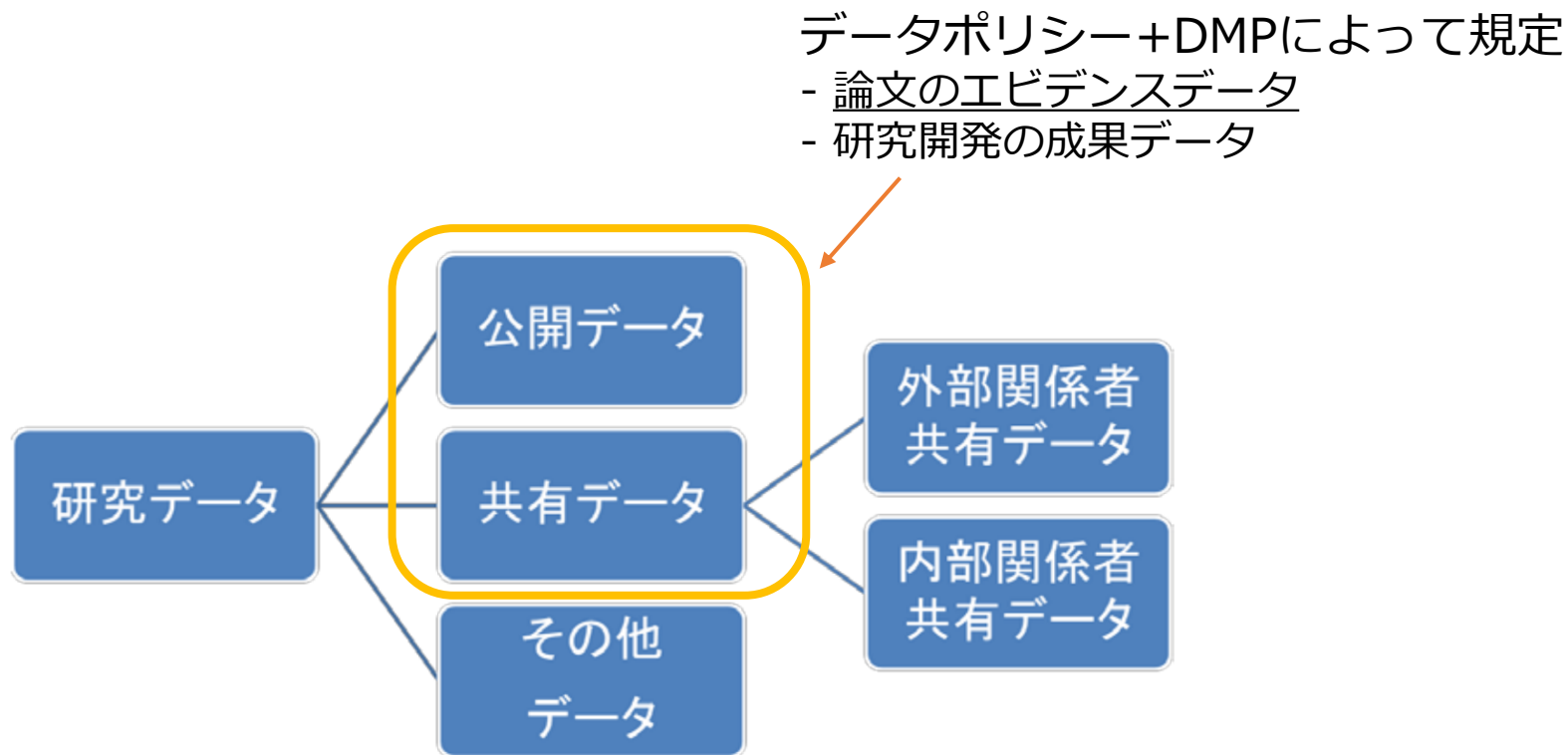
研究データの多様な来歴



機関リポジトリはどこまで扱うことが可能なのか？

研究データ管理に対する要請

研究データのうち、研究者の所属する大学、大学共同利用機関法人、国立研究開発法人等の研究開発を行う**機関や資金配分機関の基準等に基づいて**、管理・利活用の対象として、**研究者がその範囲を定めるもの**



メタデータの共通項目

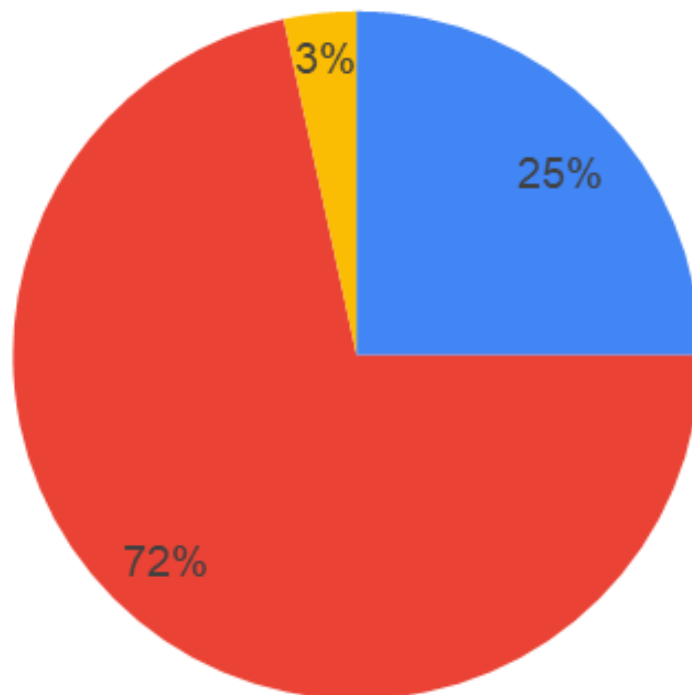
	項目	記入者	必須／任意	備考
1	資金配分機関情報	自動 (2と連動)	必須	JST、NEDO、BRAIN、AMEDのいずれか
2	e-Radの課題番号	PM以下	必須	e-Radに登録した課題番号
3	プロジェクト名	自動 (2と連動)	必須	PMが統括する研究開発の範囲の名称
4	データNo.	PM以下	必須	各FAが付与するデータNo.
5	データの名称	PM以下	必須	学会資料、報告資料、測定結果などの中身の分からない名称は避ける
6	掲載日・掲載更新日	PM以下 or自動	必須	メタデータの掲載日・掲載更新日
7	データの説明	PM以下	必須	端的かつ中身の分かる内容を記載
8	データの分野	自動 (2と連動)	必須	e-Radで登録した研究分野（主）の研究内容
9	データ種別	PM以下	必須 (選択式)	NIIの検索基盤に対応するために必須。ただし、デフォルトを「データセット」とし、異なるものについてのみ変更してもらう。
10	概略データ量	PM以下	任意 (選択式)	1GB未満、1GB以上10GB未満、10GB以上100GB未満、100GB以上
11	管理対象データの利活用・提供方針	PM以下	必須	無償／有償、ライセンス情報、その他条件（引用の仕方等）等を記載
	アクセス権	PM以下	必須 (選択式)	公開／共有／非共有・非公開／公開期間猶予から選択
	公開予定日	PM以下	必須	公開期間猶予を選択した場合、公開予定日を記載
12	リポジトリ情報	PM以下	必須	現在のリポジトリ情報、あるいはプロジェクト後のリポジトリ情報
	リポジトリURL・DOIリンク	PM以下	任意	情報があれば記載
13	データ作成者	自動 (*1と連動)	任意	管理対象データを生み出した研究者の名称
	データ作成者のe-Rad研究者番号 *1	PM以下	任意	管理対象データ作成者のe-Radの研究者番号
14	データ管理機関	PM以下 or 自動 (*2と連動)	必須	各管理対象データを管理する機関
	データ管理者	PM以下 or 自動 (*2と連動)	必須	データ管理組織において管理対象データを管理する担当者の名称
	データ管理者のe-Rad研究者番号 *2	PM以下	任意	e-Rad研究者番号がない管理者は記入不要、ある場合は必須
	データ管理者の連絡先	PM以下	必須	データ管理者の所属機関の住所や電話番号、メールアドレス等。管理対象データに興味を有する第三者が必ずデータ管理者にたどり着けるようにする
15	備考	PM以下	任意	

国内機関による 機関リポジトリでの研究データ公開経験

Q21. 貴機関の機関リポジトリで研究データを公開している事例はありますか。

データの個数 / Q21. 貴機関の機関リポジトリで研究データを公開している事例…

N=352



Q21. 貴機関の機関リポジトリ… ▼

- ある
- ない
- わからない

25%、88機関が研究データの公開経験あり

国内機関におけるRDMの現状と課題

- 大規模機関、研究開発法人が先行してRDM体制の構築が進んでいる
- RDM体制の構築においては、既存の機関リポジトリを用いた研究データ公開サービスが先行している
- 研究データ公開に当たってはデータキュレーションの実践が課題であり、かつ要望も高い

2. 研究データ公開における課題：データキュレーション

データキュレーションとは

伝統的な理解： Curation⊃Archiving⊃Preservation

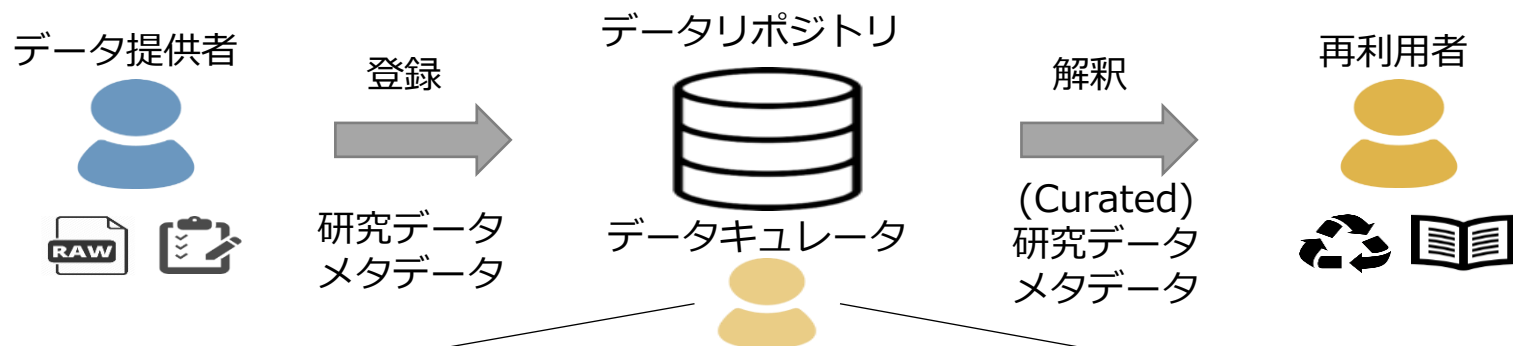
- **キュレーション**：データが目的に適合し、発見や再利用できるようにするために、データが作成された時点から、データの利用を管理し促進する活動。動的なデータセットの場合、これは目的に合ったデータを維持するための継続的なエンリッチメントまたは更新を意味する。より高度なレベルのキュレーションには、アノテーションや他の出版物とのリンクを維持することも含まれる
- **アーカイビング**：データが適切に選択され、保管され、アクセスできるようにし、セキュリティと真正性を含め、論理的および物理的な整合性が長期間にわたって維持されるようにするキュレーション活動
- **プリザベーション**：データの特定の項目が時間の経過とともに維持され、技術の変化によってもアクセスして理解できるようにするためのアーカイビングの活動



データを解釈可能かつ利用可能にするための一連の処理を
データキュレーションと呼ぶ

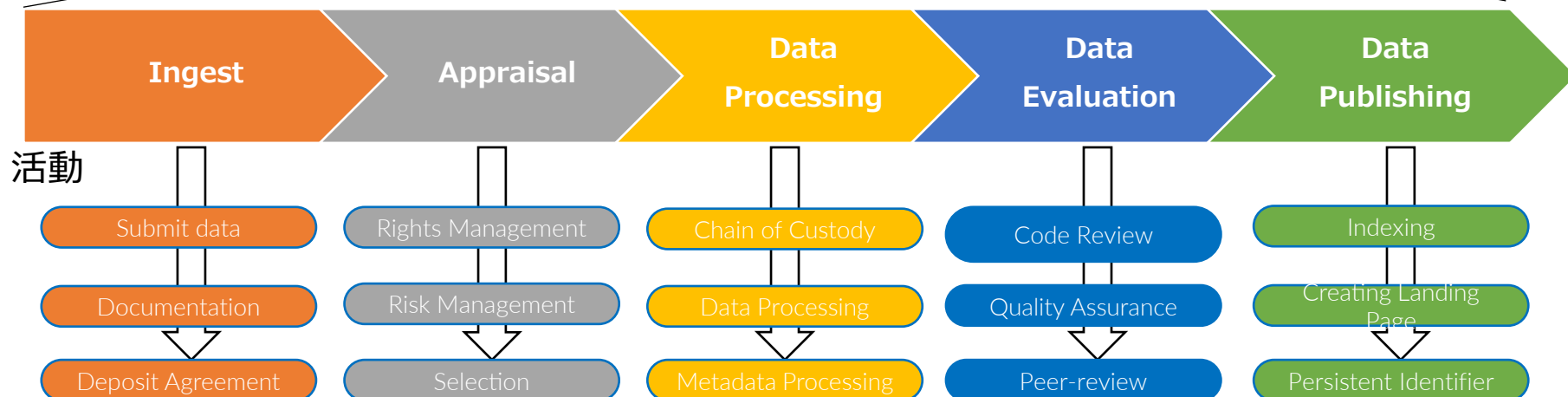
データキュレーションの構図

研究データの再利用にはデータキュレーションが不可欠

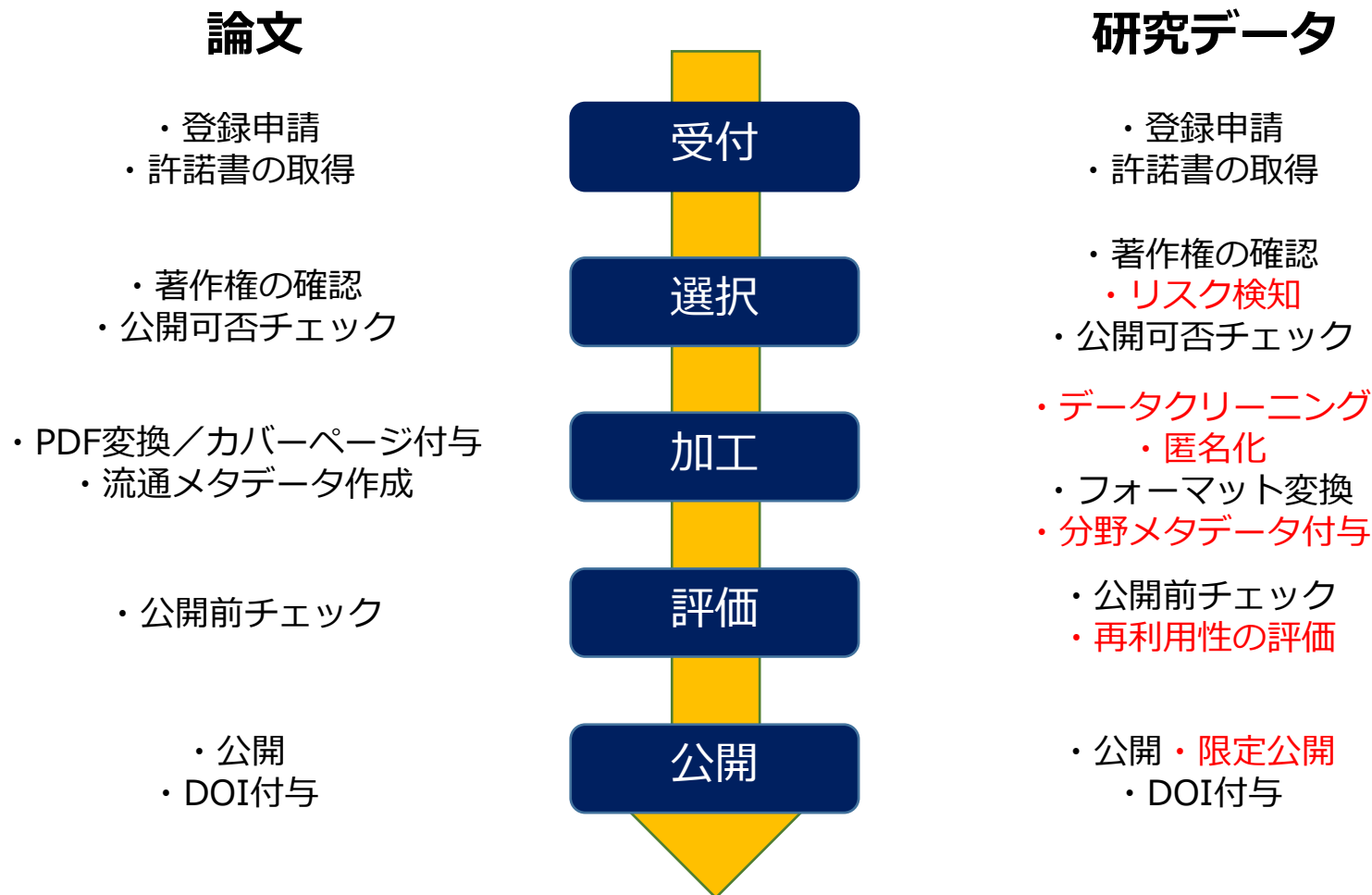


ステップ

データキュレーション



課題：リポジトリ登録時のタスクの違い



従来のリポジトリ登録と類似するタスク、異なるタスクが存在する

3. 考察と実践

メタデータ作成における課題

研究知識を表現するメタデータの例

Optical observations in Arctic and Antarctic stations

Home RT plots Tromsø Longyearbyen South pole McMurdo Link

Last update: 2015-8-11

- The observations at PRE building in Tromsø are carried out in corporation with Univ. of Tromsø.

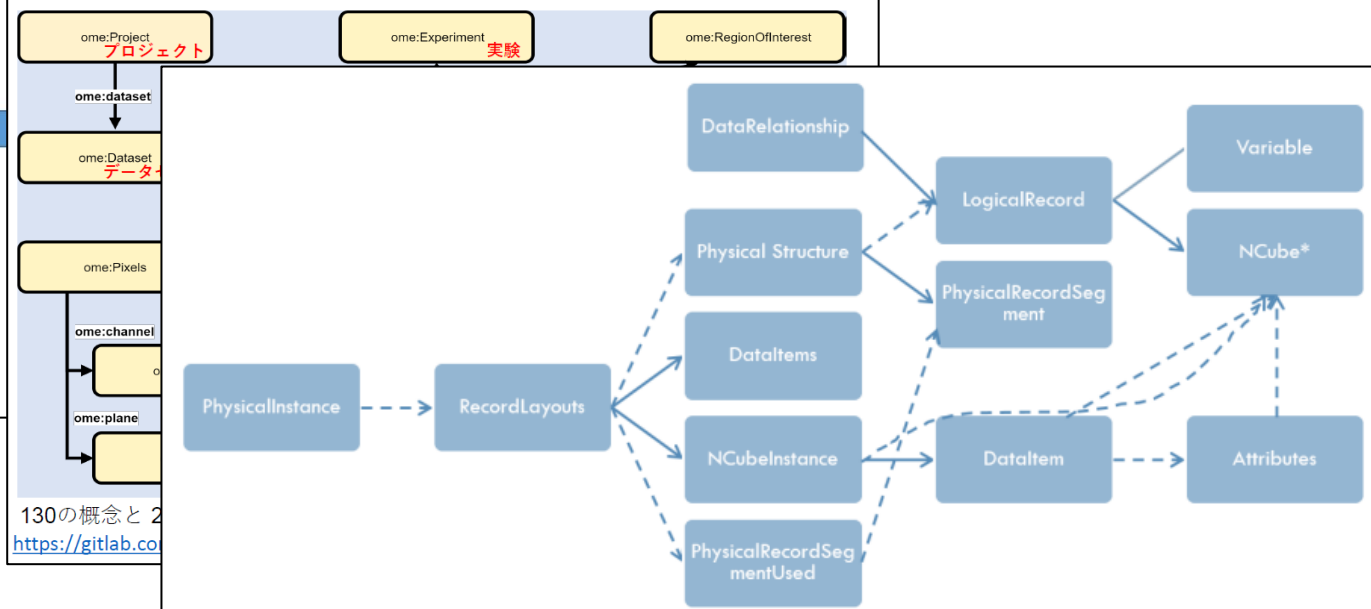
Operation notes are here

Optical instruments in Tromsø

- All-sky Color Digital camera (CDC)
 - PI: H. Miyaoka
 - Co-I: Y. Ogawa
 - Current archived page is here (supported by Y. Maezima (Maezima.net)).
 - Previous archived page is here (between 2003 and Dec. 2008).
- All-sky Watec imagers (AWI)
 - PI: Y. Ogawa
 - Co-Is: A. Kadokura, Y. Tanaka T. Motoba (APL/JHU), and K. Hosokawa (UEC)
 - Description of the instrument/data
 - Calendar for all the WHI archived data are here (since 2010)
 - Note: Data are processed only when aurorae were seen.
- Narrow field-of-view Watec imagers (NWI)
 - PI: Y. Ogawa

<http://pc115.seg20.nipr.ac.jp/www/opt/tro.html>

研究知識データの例 OME-OWL (OME-Core-201606) 光学顕微鏡撮像についてのデータスキーマ



リポジトリシステムの主たる役割はメタデータ管理
機関リポジトリで何を表現したいのか？

https://www.nii.ac.jp/sparc/event/2016/pdf/20161026_5.pdf

https://www.nii.ac.jp/event/upload/libfair2021_forum3_1.pdf

https://ddialliance.org/sites/default/files/DDI%203.2%20Best%20Practices_0.pdf 15

私見：機関リポジトリで表現したいこと



論文の根拠データを学術
成果物として扱いたい

学術成果物は引用可能に
なることで認知される

根拠データを引用できる
プラットフォームが必要

引用に必要なメタデータ

データの引用記述形式の一例

構造の表現に必要

作成者(出版年): タイトル. [バージョン.] 出版社. [データのタイプ.] 識別子

参考: DataCite, Cite Your Data <https://www.datacite.org/cite-your-data.html>

• 分野によって追加情報が必要な場合も

例)

- 経済・社会科学系: データがカバーする地理的情報、時間的範囲
- 生命科学分野の時系列データ: データベース名・アクセス番号
- 継続的に更新されるデータベースの場合: ダウンロード日

データの識別に必要

どのようにカバーするか?

Ref. “研究者のための研究データマネジメント - データの引用”
<https://jpcoar.repo.nii.ac.jp/records/294>

実践例

分野別の追加情報を記述したデータ文書を作成、公開

カタログのメタデータ

Data supporting “Investigation and Development of the workflow to clarify conditions of use for research data publishing in Japan” Version 1.0

<https://doi.org/10.20736/00001468>

The screenshot shows a data catalog interface. On the left, there is a list of files with columns for '名前 / ファイル' (Name / File), 'ライセンス' (License), and 'アクション' (Action). The first file is a folder named 'https://repository.nii.ac.jp/?action=repository_uri&item_id=1526&file_id=21&file_no=1'. The second file is 'Dataset (79.2 kB)' with a Creative Commons license. The third file is 'Codebook (562.3 kB)' with a Creative Commons license, which is highlighted with a red box. To the right of the list, there is a details view for the 'Codebook' file, showing '860 views' and a 'See details' link. A red arrow points from the 'Codebook' file in the list to the 'See details' link in the details view.

研究知識を表現するメタデータ

The screenshot shows a data document template. The top section is titled 'Data supporting “Investigation and Development of the workflow to clarify conditions of use for research data publishing in Japan” Version 1.0'. Below this, there is a section for '1. Author Information' with fields for Name, Institution, Email, and ORCID. The next section is '2. Date of data collection: 2018-02-13 to 2018-03-20'. The following section is '3. Geographic location of data collection: Japan'. The final section is '4. Data Description' with a paragraph of text. To the right of the data description, there is a section for '5. List of Variables' with a table of variables. The table has columns for 'Variable Name', 'Description', 'Type', 'Value Labels', and 'Value Labels (jp)'. The variables listed are 'SAMPLE ID', 'ANSWER DATE', 'CELLNAME', and 'Q1'.

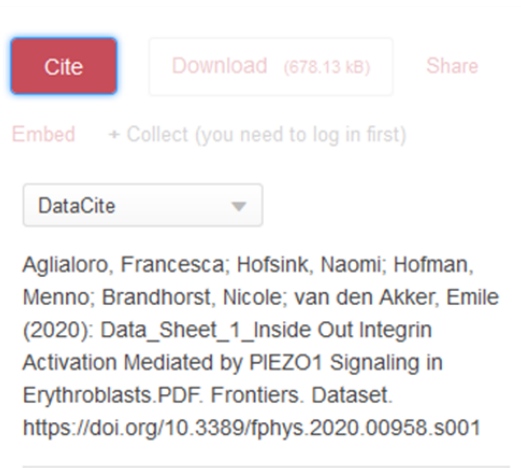
Variable Name	Description	Type	Value Labels	Value Labels (jp)
SAMPLE ID	respondents ID	Numeric		
ANSWER DATE	Date and time of data collection	Date, yyyy/mm/dd		
CELLNAME	Group number of respondents	Numeric		
Q1	Property of respondents	Numeric	1="I am not currently engaged in any research activities"	1="現在、研究活動は行っていない"
			2="Engineering"	2="工学"
			3="Astronomy"	3="天文学"
			4="Chemistry"	4="化学"

Data supporting “Investigation and Development of the workflow to clarify conditions of use for research data publishing in Japan” Version 1.0
<https://doi.org/10.20736/00001468>

4. 今後の展開

機関リポジトリを選ぶ理由

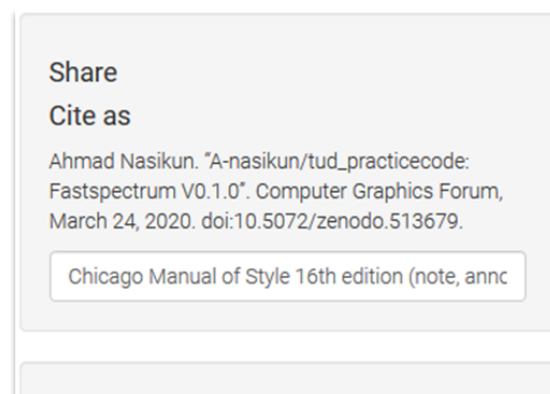
figshare



WEKO3



zenodo



<https://figshare.com/>

個々のデータの画面でCite
ボタンを押す



画面右のCite asというブロッ
クにスタイルを入力する



<https://zenodo.org/>

画面右のCite asというブロッ
クにスタイルを入力する

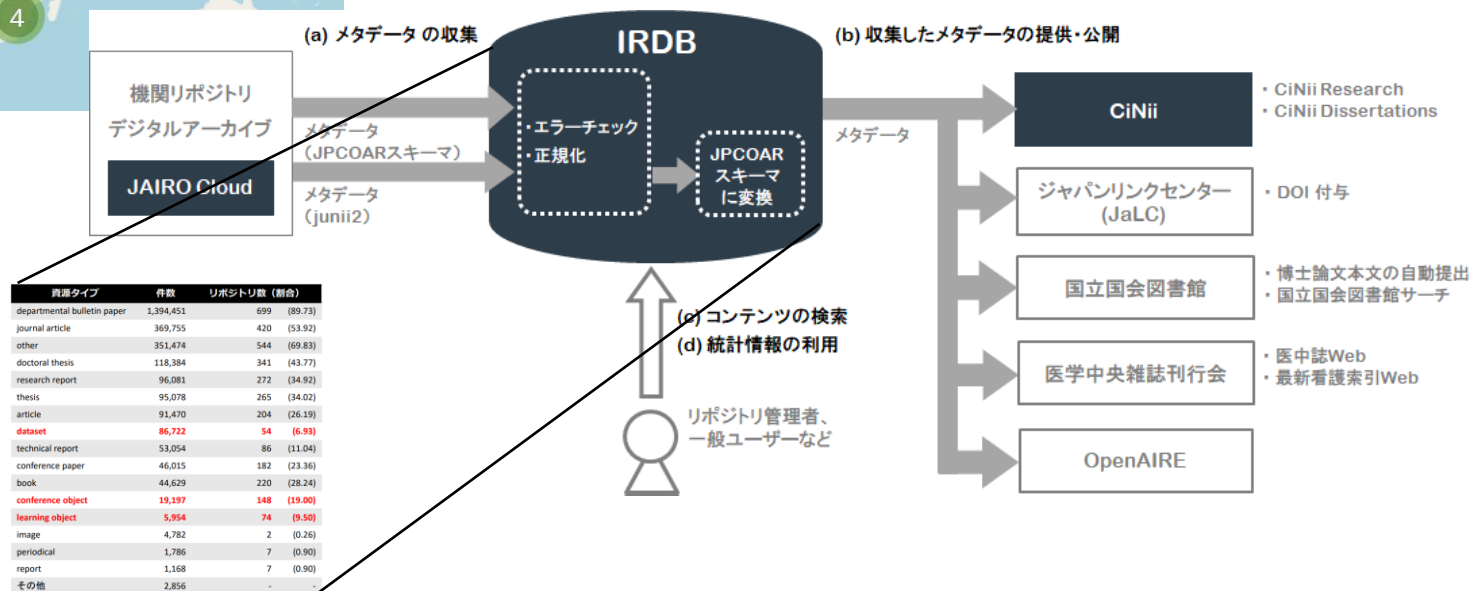
データ引用における技術的な違いはない

日本の機関リポジトリの特徴

<https://www.coar-repositories.org/>



<https://support.irdb.nii.ac.jp/ja>



国際コミュニティとの繋がり
IRDBによるメタデータ流通経路の拡張
多様なコンテンツ

課題

Strengths

Distribution of control of scholarly resources

Beyond journals - collect a variety of content types, across domains and regions (bibliodiversity)

Hosted by long-lived institutions and supported by teams of librarians (libraries/universities)

Weaknesses

Under-resourced and under-staffed

Silos across repositories

Not user friendly systems

Low value proposition for researchers / no prestige, lack of branding (no incentives)

Social emphasis on diversity, equity and inclusion
Open Science - Interest in research data and other research outputs

New models of scholarly communications coming onto the scene

Trend towards integration of university services and commercial consolidation

Commercial publisher narratives diminishing repositories

Pitted against the “gold route” of transformative agreements

Opportunities

Threats

SWOT Analysis for Repositories and the Global Repository Network (COAR - September 2022)より抜粋

国内での議論

図書館・博物館・大学でのデータ管理

6/10 (金) 14:00~15:30

C3

機関リポジトリでの研究データ公開に向けたグループディスカッション

BoF - Group discussion on publishing research data in institutional repository

南山 泰之 (国立情報学研究所)

Yasuyuki Minamiyama (National Institute of Informatics)

研究データの公開に際しては、データ公開可否の判断、公開のためのインフラ選択、保存付与といったいくつかのステップが存在する。機関として適切な研究データ公開を行うために、機関間で合意していく必要があるものの、国内大学・研究機関における標準的なサービスモデルは、本セッションでは、研究データを公開するステップを具体的に設定し、各ステップで実施

か、AXIES-JPCOAR 研究データ連絡会

応

ー

RDM事例形成プロジェクト

目的・概要

- 国内大学・研究機関において、研究データ管理 (RDM: Research Data Management) の取組み事例を形成し、その成果を「事例集」として取りまとめることにより、他大学・研究機関におけるRDM取組み拡大を図る。
- 事例集には、取組み事例を列挙するに留まらず、実際にRDMを実現するにあたり課題となる点や考慮すべき点 (体制、予算、規定、基盤、その他) などの分析を含める。

機関における研究データ公開サービスへの期待と展望

日時

11/14 (月) 10:30-12:00

- 実施方法: Live
- 参加無料、要申し込み

home · event ▼ · project ▼

情報学研究所オープンサイエンス基盤研究センター特任助教)
科学省科学技術・学術政策研究所上席フェロー / 内閣府科学技術政策局参事官)
大学文学部英米語英米文学科准教授)
大学附属図書館事務部長 / 国立大学図書館協会資料委員会オー
学附属図書館学術支援課研究支援掛長 / 国立大学図書館協会資
エンス小委員会)

https://joss.rcos.nii.ac.jp/session/overview/?id=se_118
<https://www.nii.ac.jp/event/other/libraryfair/2022/>
<https://sites.google.com/view/axies-jpcoar/>

NII RDCにおける取り組み

汎用的なプロセスをアプリケーションとして実装し、
大学・研究機関へデータキュレーション支援サービスを提供

活用

コード付帯機能

データ・プログラム・解析環境のパッケージ化と流通機能を提供し、研究成果の再現性を飛躍的に向上

信頼

データプロビナンス機能

データの来歴情報の管理から利用状況を把握でき、データ公開へのインセンティブモデルを提供

蓄積

セキュア蓄積環境

安全で強固なデータの保存・保護機能を有する超鉄壁ストレージを提供し、機微な情報も安心して保全

セキュア
蓄積環境

管理

データガバナンス機能

計画に基づきデータ管理等を機械的に支援し、DMPをプロジェクト管理に不可欠な仕組みへと変革

流通

キュレーション機能

専門的なキュレーションを実践できるエコシステムを構築し、データ再利用の促進に寄与

保護

秘匿解析機能

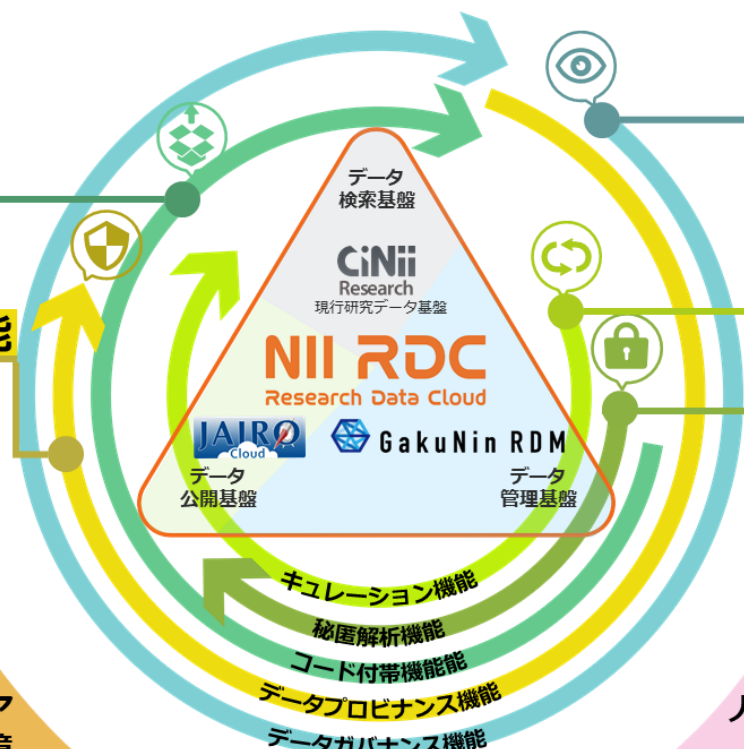
秘密計算技術で機微な情報も安心して解析できる環境の提供で、新しいデータ駆動型研究の世界を開拓

育成

人材育成
基盤

人材育成基盤

RDMに必要なスキルを学ぶ環境を提供し、
全ての研究者を新しい科学の実践者へと育成



5. まとめ

まとめ

1. 背景

- 機関リポジトリでの研究データ公開が政策的に求められている

2. 研究データ公開における課題：データキュレーション

- 論文と研究データではデータキュレーションで実施すべきタスクが異なる

3. 考察と実践

- データ引用の観点から機関リポジトリでのデータキュレーションを実践した

4. 今後の展開

- 機関リポジトリの強み／弱みを分析しながら議論が続けられている