# 研究データ公開に向けたワークフローの検討

国立情報学研究所 学術情報基盤オープンフォーラム 2020

三上 絢子(北海道大学 附属図書館)

CC BY

# 本発表の背景

北海道大学 附属図書館→国立情報学研究所実務研修(2019/10/1～12/27)に参加

主なテーマ：**研究データ管理**について

→研究データ公開ワークフローの試案を作成(本発表のベース)

各地の会議で情報収集

所内での打ち合わせを経て試案作成

イベントのお手伝いもしました

研修の詳細はこちら：https://hrd.nii.ac.jp/jitsumu
本発表のワークフローの詳細なメタデータ表や研修の報告も公開しています。

# 本発表の背景

## 今回の想定：

大学や研究機関の図書館で
論文の根拠となる研究データを公開
（研究者から、論文採択に必要なため依頼があったケース）

**公開基盤：**次期JAIRO Cloudを想定

JPCOARスキーマベース
将来的にGakuNin RDMとの連携を予定

研究データ管理基盤
　バージョン管理
　作成者のプロフィール情報　など

# 全体の流れ

1. 研究データ公開の背景
2. 論文と研究データとの違い
3. ワークフロー
   公開に必要な事項の確認
   メタデータ付与
4. 今後の課題

# 研究データ公開の背景

## 研究公正などの必要性

出版社や助成機関の要請

- 論文投稿要件
  必須～推奨までのレベル
  公開リポジトリが指定
  （機関リポジトリなど）
- 助成採択の参考or要件

## 公開によるメリット

データ公開・利活用への期待

- ICT技術の進歩
  取り扱えるデータ量が飛躍的に増えた
  大量かつ高品質のデータの利活用が課題
- 研究を広く周知したい

# 論文と研究データの異なる特徴

## 論文

- ✓元々公開を前提に作成
- ✓基本的に、既に公表済の資料を機関リポジトリで再公開する
- ✓著作権は出版社に移転または著者が保持（OA論文の多く）
- ✓大規模出版社の多くは再公開のための許諾条件を明記
- ✓(一般に)引用による再利用＝出典情報を明記する
- ✓既に国内の多くの学術機関に公開用リポジトリが存在

## 研究データ

- ✓必ずしも公開が前提ではない
- ✓査読前に、論文に関するデータの公開を求められる場合がある
- ✓権利保持者が明確ではない
- ✓再利用の慣習は定まっていない
- ✓研究データに関するポリシーを定めていない学術機関も多い
  (国立研究開発法人中心に策定済の機関も
  2020/03/09 京都大学がポリシー策定)

# 公開までの流れ

公開依頼

## 論文(現状)

公開可否の確認

権利者＝出版社のポリシーを確認
再公開（アーカイブ）条件チェック
Web公開は可能か
バージョンの指定（著者版、出版社版）
日時の指定（エンバーゴ）

メタデータの付与

- 検索性を高めるための情報付加
  *タイトル・著者名・掲載雑誌…*
- 希望に応じてDOIを取得

## 研究データ

権利者＝研究者など？ 明確ではない場合も
**法倫理問題、所属機関等のポリシー**確認
公開範囲の指定があるか
公開日時の指定はあるか

- 検索性を高めるための情報付加
  *タイトル・**権利者名・説明・ライセンス**…*
- 希望に応じてDOIを取得

公開（機関リポジトリ掲載）

# 研究データ公開ワークフローの概要

## 論文公開のワークフローをベースに作成

研究データ

↓

公開可否の確認（法倫理的問題や契約など）

公開条件の指定（公開範囲、公開日時、ライセンス…）

メタデータの付与

研究データ公開へ

# ワークフロー前半：公開可否

## 論文公開のワークフローをベースに作成

研究データ

**公開可否の確認（法倫理的問題や契約など）**

**公開条件の指定（公開範囲、公開日時、ライセンス…）**

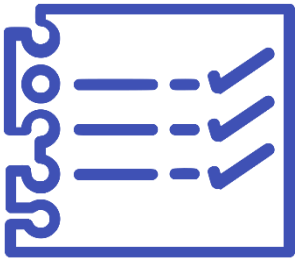
メタデータの付与

研究データ公開へ

# ワークフロー（前半）の内容

公開可否の確認①
研究者自身あるいは権利者が確認

公開可否の確認②
研究コミュニティや法務部門等で確認



公開条件の確認

参考：
研究データ利活用協議会(RDUF) 研究データライセンス小委員作成,
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）

# 公開ワークフロー：公開可否の確認①

- 公開不可能な事由がないか確認
  共同研究者の合意を得ているか
  企業等の共同研究上の契約

- 倫理規定確認
  （研究計画の段階でクリアしているはず）
  →もし変更があれば確認

# 公開ワークフロー：公開可否の確認②

- 研究コミュニティの慣習において公開が妥当か

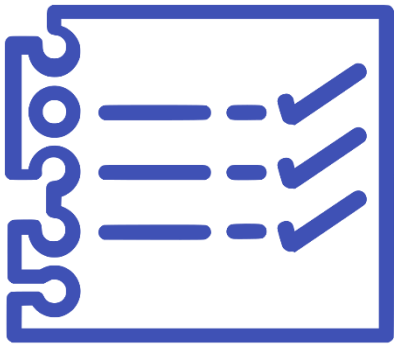
  治験データ
  絶滅危惧種の生息地域
  文学研究のインタビュー　など

- 法務部門等に確認する必要のある問題

  主に法的問題
  ・個人情報
  ・出願予定の知的財産権
  ・国家安全保障

当てはまる項目があれば各担当部署に相談

# 公開ワークフロー：公開条件の設定



- 公開範囲
  一般公開
  研究者間に共有　など
- 日時
  即時
  論文公開後速やかに　など
- ライセンス
  CC0　など
- その他（DOI登録など）

機関や出版社などのポリシーに指定がある場合は従う

共同研究の合意や契約に沿った形で

# ワークフロー後半：メタデータ付与

## 論文公開のワークフローをベースに作成

研究データ

公開可否の確認（法倫理的問題や契約など）

公開条件の指定（公開範囲、公開日時、ライセンス…）

**メタデータの付与**

研究データ公開へ

# メタデータについて

次期JAIRO Cloudで使用するJPCOARスキーマをベースに検討

JPCOARスキーマに基づくメタデータの例：（**太字**が要素名）

```
<dc:title xml:lang="ja">情報爆発時代の研究基盤構想</dc:title>
<jpcoar:creator>
    <jpcoar:creatorName xml:lang="ja">安達, 淳</jpcoar:creatorName>
    <jpcoar:affiliation>
        <jpcoar:affiliationName xml:lang="ja">東京大学
</jpcoar:affiliationName>
    </jpcoar:affiliation>
</jpcoar:creator>
```

JPCOARスキーマサンプル
https://github.com/JPCOAR/schema/tree/master/1.0/samples

Q. 研究データにはどのような
メタデータを付与すればよいか？

# メタデータの整理

再利用の観点で必須

再利用には必須ではない
（検索に有用、あればよい）

データ提供者*のみ
知っている内容
(*研究者など)

タイトル
作成者姓名
寄与者/権利者
アクセス権
権利情報
内容記述
日付
資源タイプ
バージョン情報

作成者姓/作成者名
所属機関名
APC
主題
言語
時間的範囲
位置情報
助成情報

メタデータ入力者*
が後から付与
できる内容
(*図書館員など)

出版者（リポジトリ名）
ファイルのURL
フォーマット

所属機関識別子
助成機関識別子
識別子スキーマのURL
ファイルサイズ

※JPCOARスキーマより、データ利用の観点で必須要素を整理
※JPCOARスキーマの記入レベル（M:必須 R:推奨…）とは必ずしも一致しない

# 再利用のために必要なメタデータ

| グループ | 要素名（第1階層） | 項目名（日本語） | 要素名（第2階層）または属性 | データ入力者 | 自動入力 |
|---|---|---|---|---|---|
| タイトル | dc:title | タイトル | × | データ提供者 | |
| 作成者情報 | jpcoar:creator | 作成者姓名 | jpcoar:creatorName | データ提供者 | 可（GakuNin RDMより） |
| 寄与者情報 | jpcoar:contributor | 寄与者 | contributorType | データ提供者 | |
| | | 寄与者姓名 | jpcoar:contributorName | データ提供者 | 可（GakuNin RDMより） |
| アクセス権 | dcterms:accessRights | アクセス権 | × | データ提供者 | |
| 権利情報 | dc:rights | 権利情報 | × | データ提供者 | |
| | jpcoar:rightsHolder | 権利者名 | jpcoar:rightsHolderName | データ提供者 | 可（GakuNin RDMより） |
| 内容記述 | datacite:description | 内容記述 | × | データ提供者 | |
| 公開場所 | dc:publisher | 出版者 | × | メタデータ入力者 | 可 |
| 日付 | datacite:date | 日付 | × | データ提供者 | 可（GakuNin RDMより） |
| バージョン情報 | datacite:version | バージョン情報 | × | データ提供者 | |
| DOI関連 | jpcoar:identifier | 識別子 | × | データ提供者 | |
| | jpcoar:identifierRegistration | ID登録 | × | メタデータ入力者 | |
| ファイル情報 | jpcoar:file | 本文URL | jpcoar:URI | メタデータ入力者 | 可 |
| | | フォーマット | jpcoar:mimeType | メタデータ入力者 | 可 |
| | | 日付 | datacite:date | データ提供者 | 可 |
| | | バージョン情報 | datacite:version | データ提供者 | |

※JPCOARスキーマより、データ利用の観点で必須要素を整理
※JPCOARスキーマの記入レベル（M:必須 R:推奨…）とは必ずしも一致しない

# 要素をどのように入力するか？

## 例：作成者のメタデータの入力

*作成者識別子:ORCIDやResearcherIDなど

| JPCOARスキーマの要素 | | | 作成者識別子がある場合 | 作成者識別子がない場合 |
|---|---|---|---|---|
| 作成者（寄与者、権利者） | 作成者識別子 | | 手入力 | - |
| | 姓名 | | 自動補完 | 手入力 |
| | 姓 | | | 自動補完 |
| | 名 | | | |
| | 所属 | 所属機関識別子 | 各スキーマごとのDBを所属機関名で検索<br>運用によってあらかじめ設定<br>（ISNIや科研費の機関番号など）<br>自動的に決まる | |
| | | スキーマ種別 | | |
| | | スキーマを示すＵＲＬ | | |
| | | 所属機関名 | 自動補完 | 手入力 |

データ提供者（研究者など）しかわからない内容

メタデータ入力者（図書館員など）が後から入力できる内容

メタデータ入力①:

# データ提供者が入力する項目

データ提供者の例：

研究者

- GakuNin RDM*から流用できるもの
  - 作成者の情報
  - 寄与者の情報
  - ファイルの更新日
- 手入力する必要があるもの
  - タイトル
  - 内容記述
  - 助成情報
  - 権利情報
- この後の工程で設定するもの
  - アクセス権（公開範囲、公開日）
    - 一定期間後の公開
    - 特定の研究者のみアクセスできる　など

研究者の条件に沿うように、
メタデータ入力者がメタデータを設定・入力

* GakuNin RDMはNII開発中の
研究データ管理基盤
(https://rcos.nii.ac.jp/service/rdm/)
次期JAIRO Cloud(WEKO3)と連携予定

メタデータ入力②：

# メタデータ入力者が入力する項目

メタデータ入力者の例：

図書館員

データベース管理者

- 機械的に補完できるもの
  - 各種識別子（存在する場合）
  - 識別子スキーマのURL
  - ファイルのURL
  - ファイルサイズ
- 手入力する必要があるもの
  - アクセス権（公開範囲、公開日）
    - データ提供者の要望に応じて設定
  - DOI登録を行うための情報
  - その他、統制語彙の確認

メタデータ入力者が項目入力後、公開

（公開日の指定がある場合は公開日になったら）

# ワークフローのまとめ

研究データ公開に際して確認＆入力すべき事項

- 公開できる条件

  研究者が確認すべき要点

  他の部署に確認する必要のある問題

- メタデータ入力

  データ本体を見てわからない要素は研究者が入力

  自動で入力できる要素は自動化したい

  研究データの再利用を考えてメタデータを作成

# 今後の課題

## 異なる想定ケースへの対応

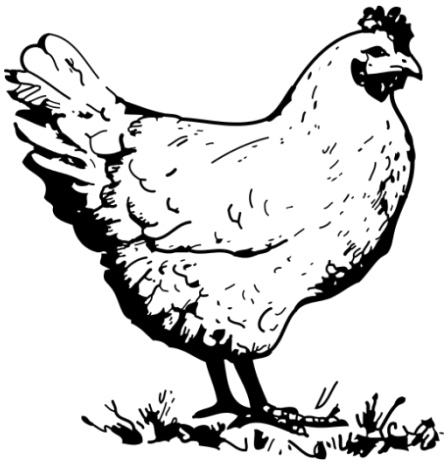
- 助成機関による公開義務化
- (査読後の)論文の根拠データ公開(義務ではない場合も含む)
- 社会的な要請が高い場合（大学以外では国の研究所？）
- 自主的な公開

## 図書館内や研究者の協力を得た試行

発表者の所属機関においても、現時点までで
(論文採択に必要なため)データの公開依頼が数件来ている
→実際に対応する準備
(マニュアル整備、フィードバック、他部署との連携…)

# おわりに

**卵が先か　にわとりが先か**

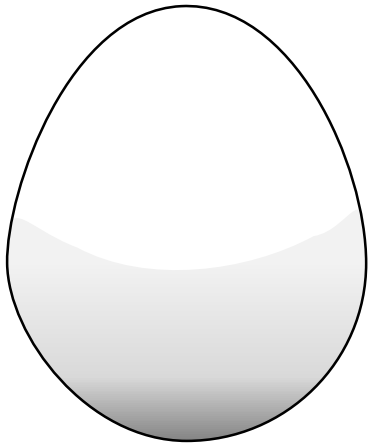


今回の検討は図書館員の視点から

- 勿論、研究データ管理は図書館だけでできることではない
- 一方で研究者へ向けたサービスの整備・発信の継続は大事

すぐに完成させることは難しいが、
手を付けない訳にもいかない

- 本発表も方法の模索
- 関心を持って少しずつ進んでいければ…



**ご清聴ありがとうございました**