

令和元年度 国立情報学研究所 実務研修成果報告会

北海道大学附属図書館 三上 絢子
令和元年(2019年) 12月25日

アジェンダ

1. 研修の概要
2. 研修テーマについて
3. 研修成果と今後の展望
4. 研修を振り返って

研修の概要

自己紹介

北海道大学 附属図書館

- 学習支援企画担当 3年
主にセミナーや講習会の企画、学生向けの情報収集ガイドも
 - 研究支援企画担当 半年
機関リポジトリの管理やWebページ周りがメイン
(昨年、システム管理担当から名前が変わった)
- 今年10月からNII実務研修

実務研修概要

- 主なテーマ：
研究データ管理・公開について
(現在、機関リポジトリ & システム周りの担当
将来的に北大でも研究データを扱うことを想定して)
- 研修先：
国立情報学研究所 学術コンテンツ課
及びオープンサイエンス基盤研究センター
- 研修期間：10/1～12/27 (三か月間)

研修中のイベント

10月 主に情報収集

- 10/1 研修開始、オリエンテーション
- 10/23～10/25 RDA@フィンランド

11月 主に課題の検討

- 11/12～11/14 図書館総合展@横浜
- 11/23 情報知識学フォーラム@金沢
- 11/26 RDA報告@NII

12月 主に成果のまとめ

- 12/3 中間報告@NII
- 12/12～12/14 AXIES年次総会@福岡
- 12/25 成果報告（本日）@NII

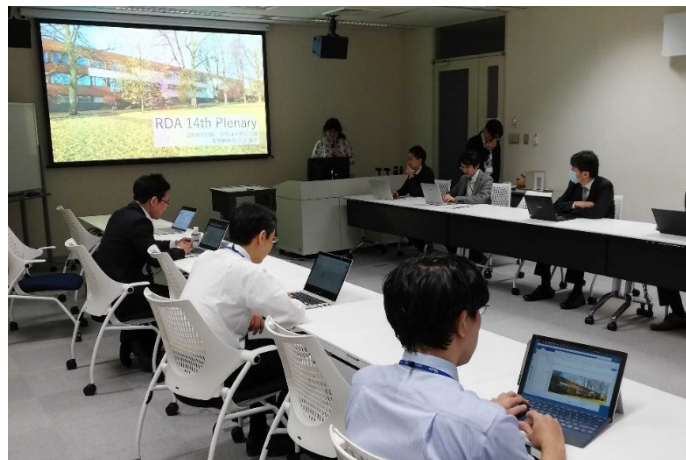


各地の鳥
(フィンランド、横浜、福岡)

研修風景



各地の会議で情報収集



まとめた内容を発表や打合せに
意見を元に更に練っていきます



イベントのお手伝いもしました

テーマ決定まで

10月前半はテーマ決定まで色々試した

- WEKO2、WEKO3やGakuNin RDMを触る
 - 北大の機関リポジトリはJAIRO Cloudではないため操作は初
 - Linux環境で構築してみたり
(Windows8上でVagrant+VirtualBoxを使って
Ubuntu18.04でGRDMのサーバ立ち上げまでは成功)
- WEKO3打合せに参加する

公開基盤 & 管理基盤の情報収集がメイン

研修テーマについて

研究データ管理から
研究データ公開を行うための
研究者向けの
ワークフローを検討・作成

研究データとは

研究の過程、あるいは研究の結果として
収集・生成される情報

例：

観測データ

実験データ

シミュレーションを行った結果得られたデータ など

論文の根拠になるデータ（エビデンス）も含まれる

研究データ公開の背景

- 出版社や助成機関の要請
 - 不正防止 透明性の確保
- ICT技術の向上
 - 論文の公開はここ10年程盛ん（オープンアクセス）
→研究データもその流れ
 - データ中心科学 機械学習や統計的手法など
- データを公開することによるメリット
 - データの存在が知られること自体に利点（主に人文科学系）
 - 他の研究者とのコラボレート

※全ての研究データを公開するものではない
（オープン&クローズ戦略）

研究データの活用サイクル

検索

研究に必要なデータを主にWeb上から取得する。
公開時に付与されたメタデータを元にキーワードで検索

助成機関によっては管理/公開手順を事前に明示することが必要
(Data Management Planの義務化)

検索

公開

研究成果としてのデータを主にWeb上で公開する。
形態は公開リポジトリ、データベース、個人サイトなど。

公開

管理

研究活動中のデータ管理。
逐次、生成や解析を行いデータは増えたり変更されたりする。

管理

活用サイクルとNII情報基盤との関連

検索基盤: CiNii Research



研究データの検索

検索

適切なメタデータにより検索
(JPCOARスキーマなど)

メタデータにより検索
利用条件に基づいて再利用
(CCライセンスなど)

公開基盤:

JAIRO Cloud



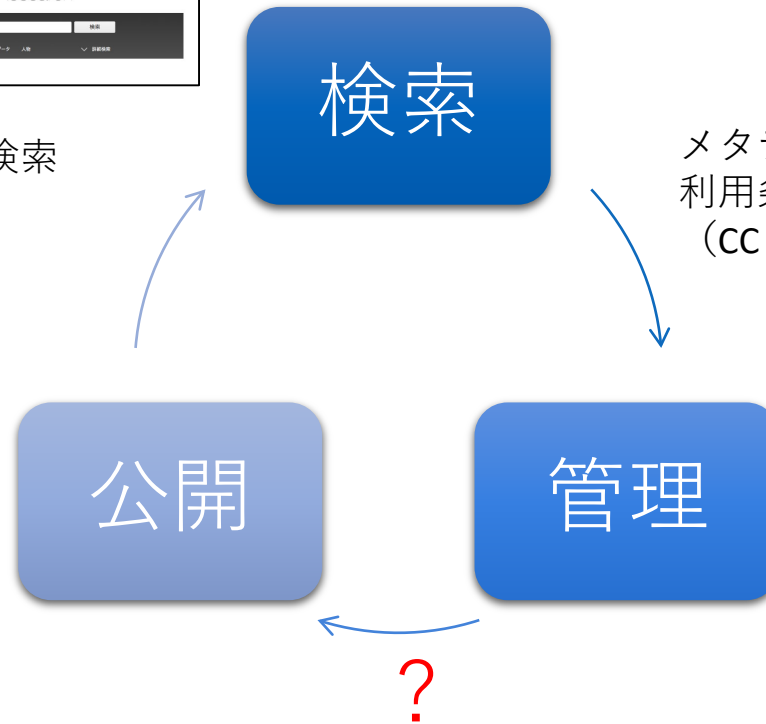
研究データの公開

管理基盤:

GakuNin RDM



研究データの
生成・解析・保存



現状の問題

研究データの管理→公開の部分は検討が十分ではない

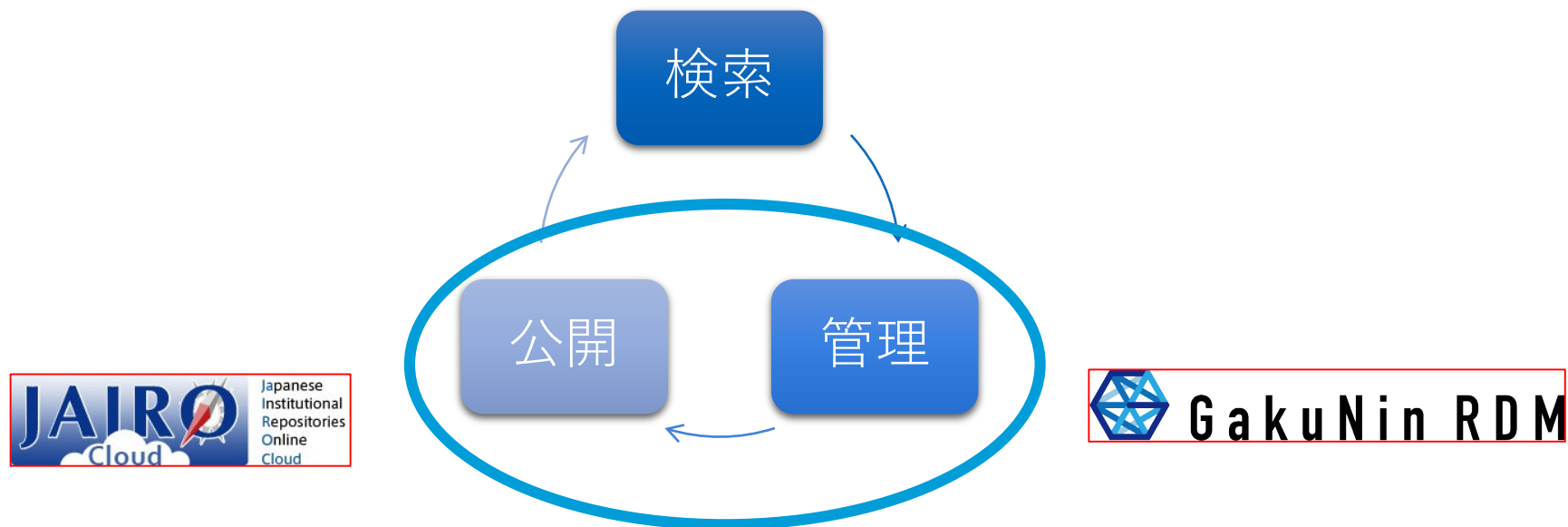
- 管理基盤（GakuNin RDM）と公開基盤（WEKO3）は開発中
- 管理基盤→公開基盤の流れ 明確なワークフローのない部分



公開に際しての手続き？
データ移行時のメタデータ付与？

研修の目的

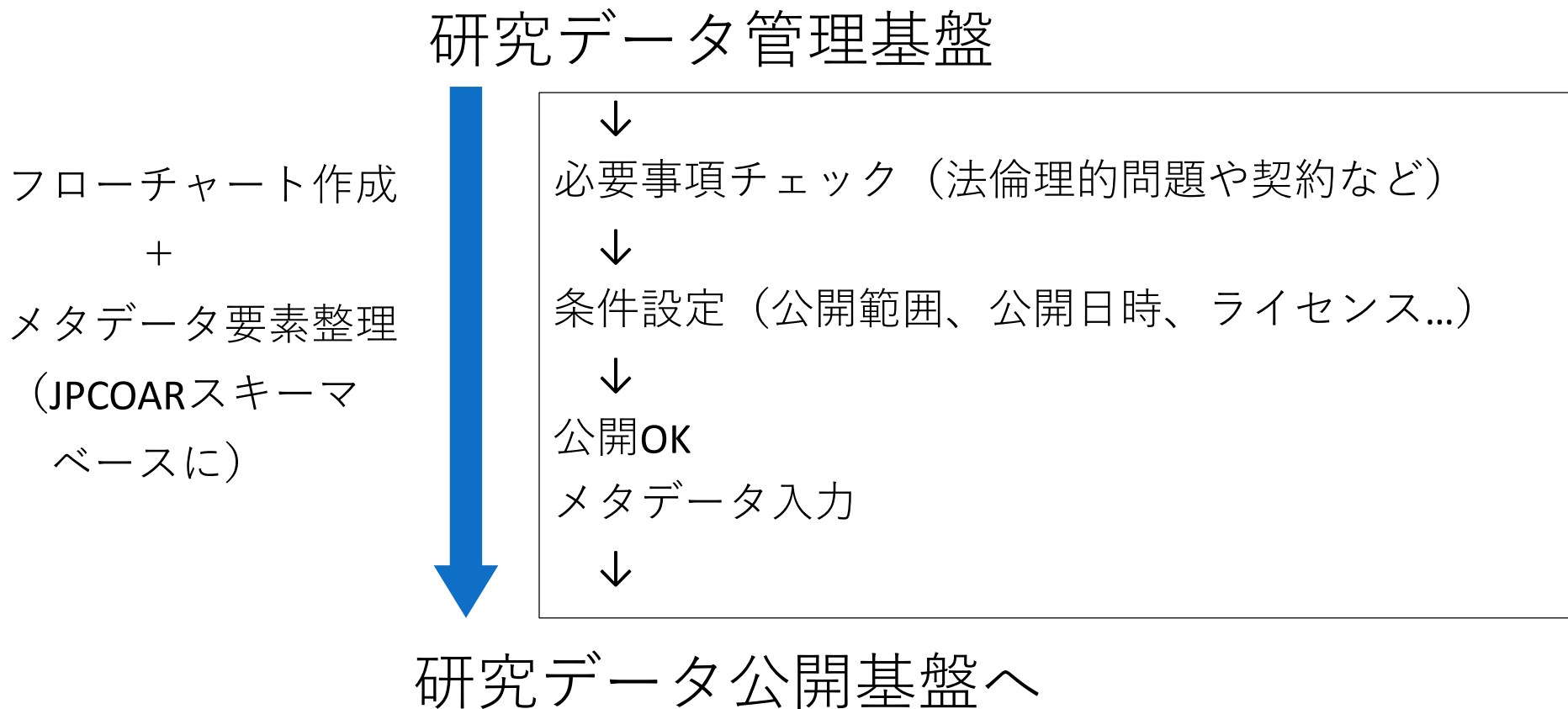
研究データを公開したい研究者向けに、公開方法を示したい



アプローチ方法：管理→公開のワークフロー作成

研修成果と今後の展望

研修成果：ワークフローの概要



前提条件（想定するシナリオ）

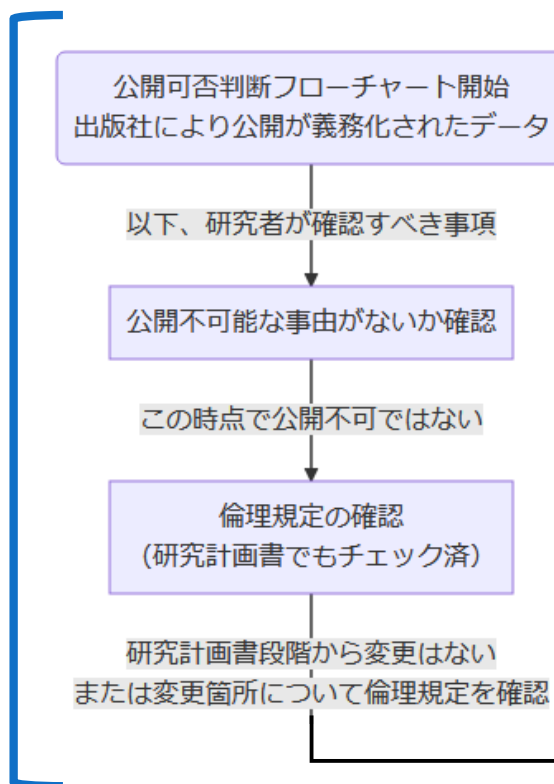
- 公開したいデータが手元にある
データ生成・解析が既に済み、公開後に第三者が利用できる状態
 - 出版社のポリシーにより公開が義務化されている
例えばSpringerのデータポリシーではジャーナルによって以下の4種類がある
 - データ公開を推奨
 - エビデンス（全ての関連する生データ）の公開を強く推奨
 - エビデンスの公開を強く推奨し、一般公開しない場合でも利用条件の明記が必須
 - 原則利用条件の制約なしでデータを公開し、データの査読を受けることが必須
- 今回は一番下の条件を想定
- 最も条件が限定的
 - 出版社の情勢を鑑みるとこれからも研究者からの要請が増えると見込まれる

参考資料

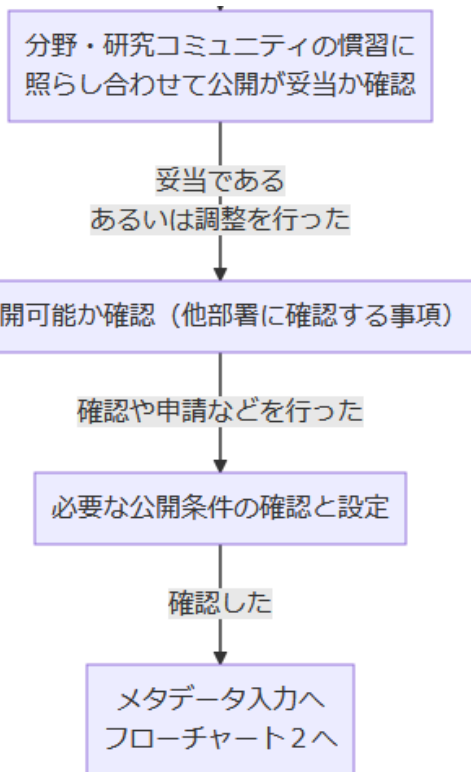
- 研究データの公開・利用条件指定ガイドライン
研究データ利活用協議会（RDUF）研究データライセンス小委員会 作成
(https://doi.org/10.11502/rduf_license_guideline)
- JPCOARスキーマガイドライン
オープンアクセスリポジトリ推進協会 作成
(<https://schema.irdb.nii.ac.jp/ja>)
- RDMトレーニングツール
オープンアクセスリポジトリ推進協会 作成
(<http://id.nii.ac.jp/1458/00000023/>)

フローチャート図（前半）の概観

公開可否の確認①



公開可否の確認②

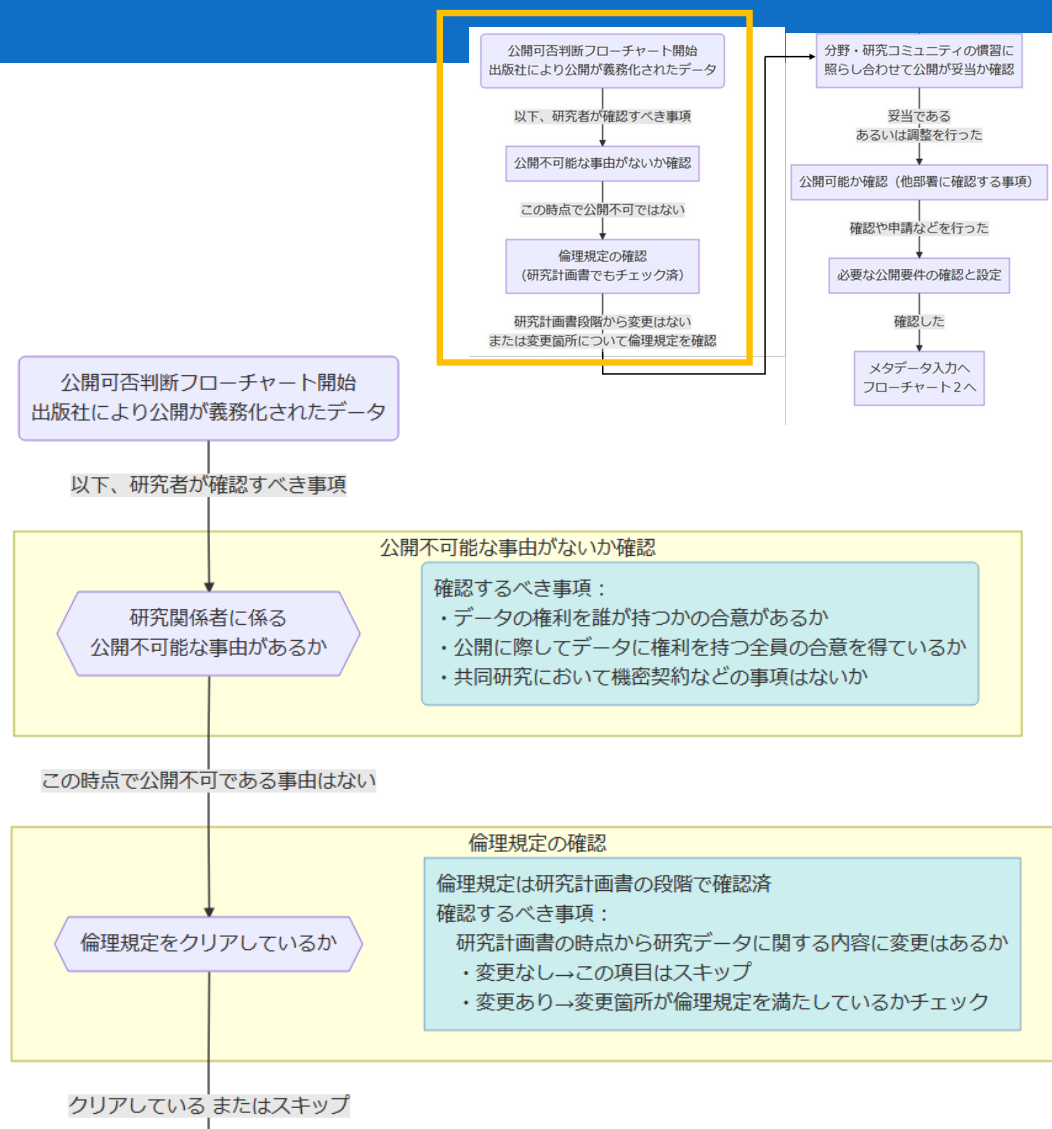


公開条件の設定
(範囲や日時など)

公開フローチャート： 公開可否の確認①

フローチャートに従い研究者が確認

- 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



公開フローチャート： 公開可否の確認②

フローチャートに従い研究者が確認

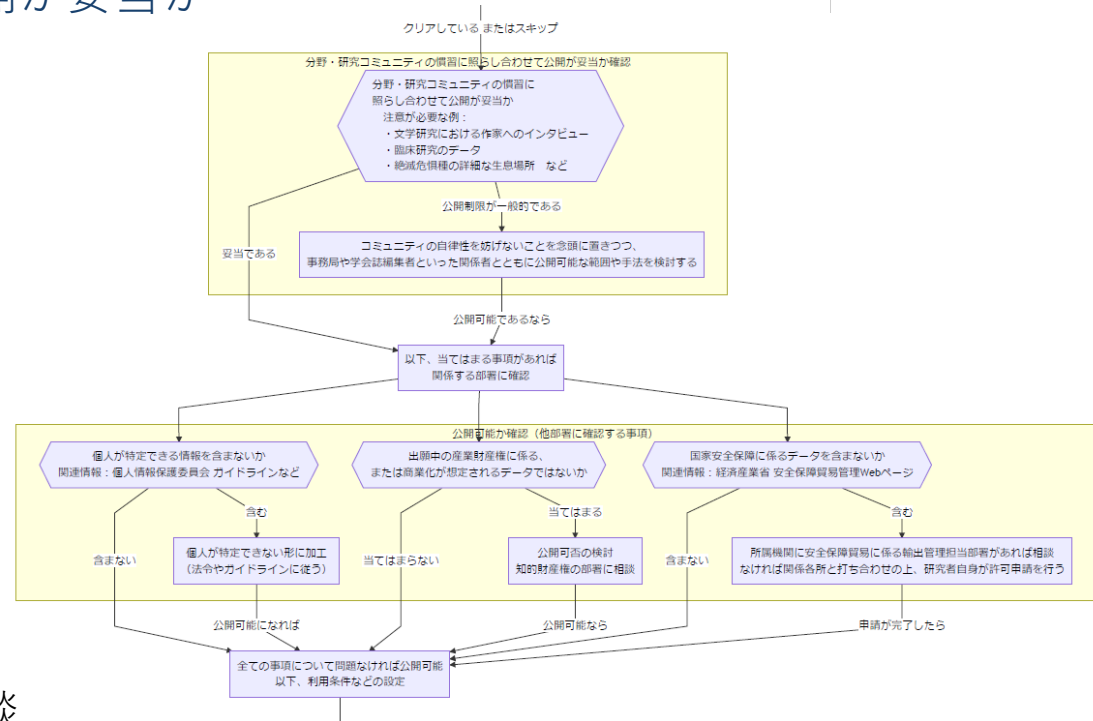
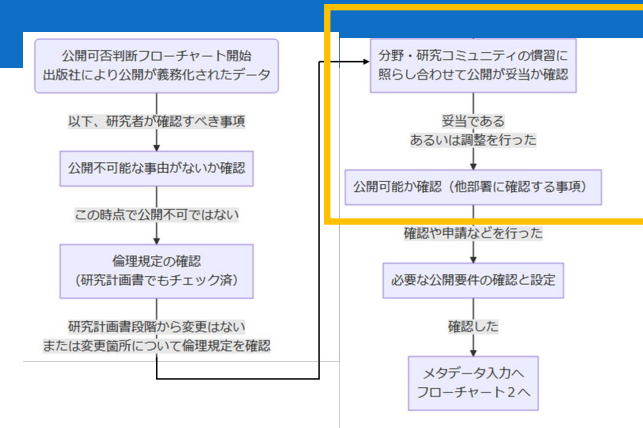
- コミュニティの慣習に照らし合わせ公開が妥当か

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

- 他部署に確認する必要がある問題

- 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

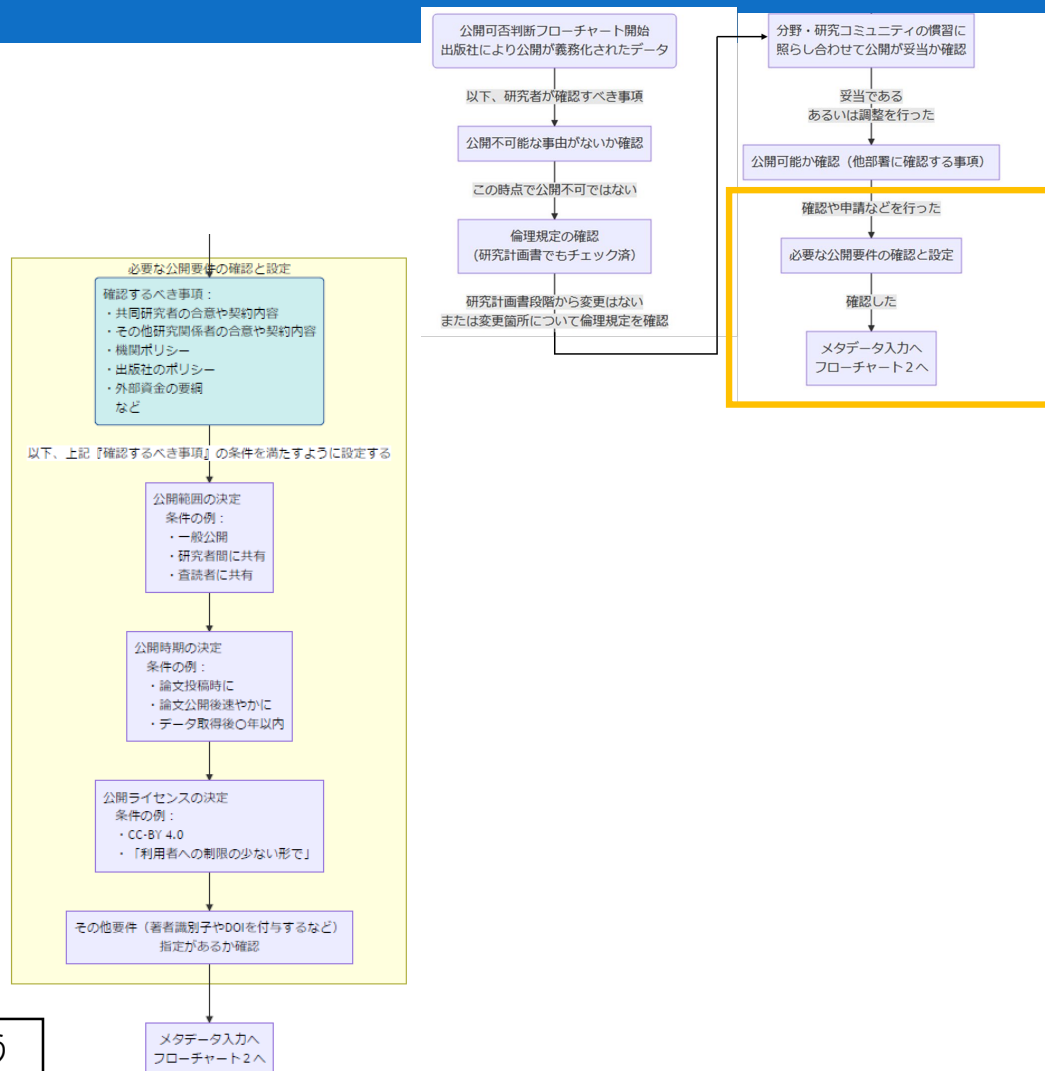


公開フローチャート： 公開条件の設定

フローチャートに従い研究者が設定

- 公開範囲
 - 一般公開
 - 研究者間に共有
 - ...
- 日時
 - 即時
 - 論文公開後速やかに
 - ...
- ライセンス
 - CC-BY
 - ...
- 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



メタデータについて

メタデータ：データ自身に関する情報

メタデータの持つ情報の例：

「このデータはいつ誰に作られた、出典〇〇の論文である」

記述要素は決まっており、機械可読なのが望ましい

メタデータの例：JPCOARスキーマ（**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>

研究データのメタデータ

論文のメタデータの場合：

- 本文中に明確に含まれる要素も多い（著者名、雑誌名...）
→図書館員などが入力
- 権利情報は明記されていない場合も多い（論文は引用の形で利用可能）

研究データのメタデータ場合：

- 通常、データセットに権利者名は含まれない
→おそらく研究者のみが知っている
- 再利用のライセンスを明確にする必要がある
(引用だけでなく内容の加工を伴うことも多い)

論文とは性質が異なる→入力業務の方法もまた異なる？

要素をどのように入力するか？

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(ORCID IDなど)	入力	-
	姓名	自動補完	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 運用によってあらかじめ設定 (ISNIや科研費の機関番号など) 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	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

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	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	
		本文URL	jpcoar:URI	メタデータ入力者	可
ファイル情報	jpcoar:file	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

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

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

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

フローチャート図（後半）の概観

メタデータ入力分担

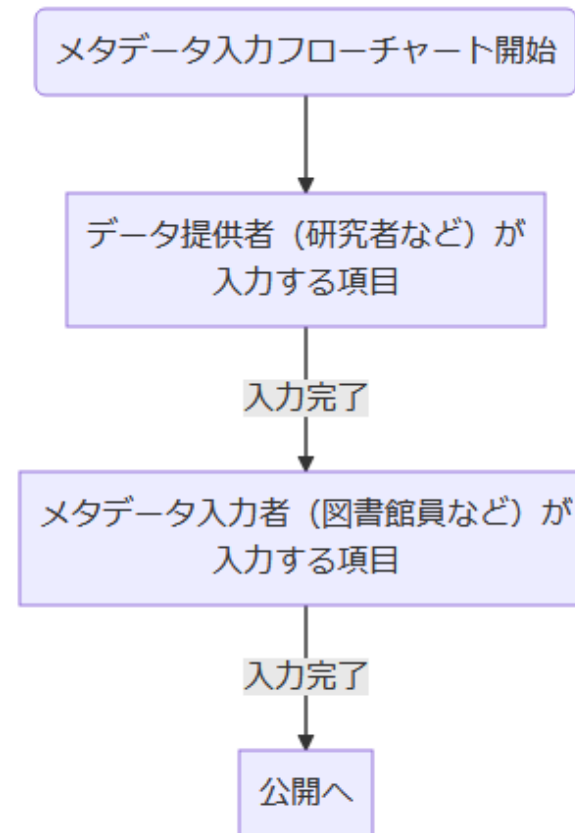
JPCOARスキーマをベースに検討
(Datasetの場合)

データ提供労力削減が目的

- データ提供者の入力項目を最小に
- 必須項目を洗い出し

方針：

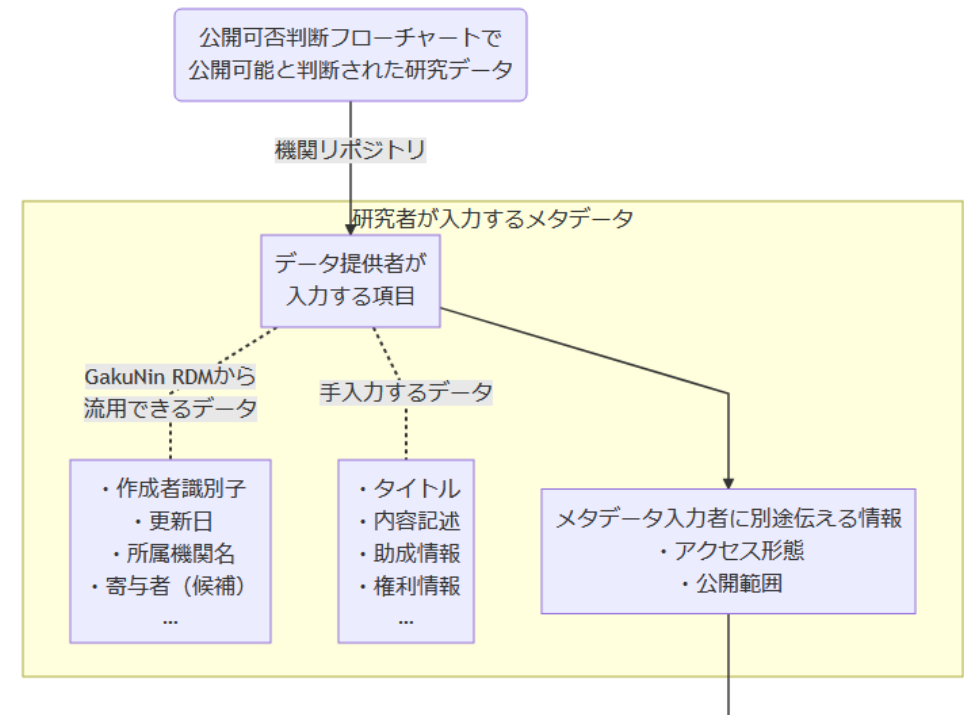
- データ提供者本人しかわからない項目は本人が入力
- その他の項目はメタデータ入力者が入力
- 自動で入力できる項目は自動で
- ORCID IDなどにより自動で入力できる項目も存在



メタデータ入力フローチャート①

データ提供者が入力する項目

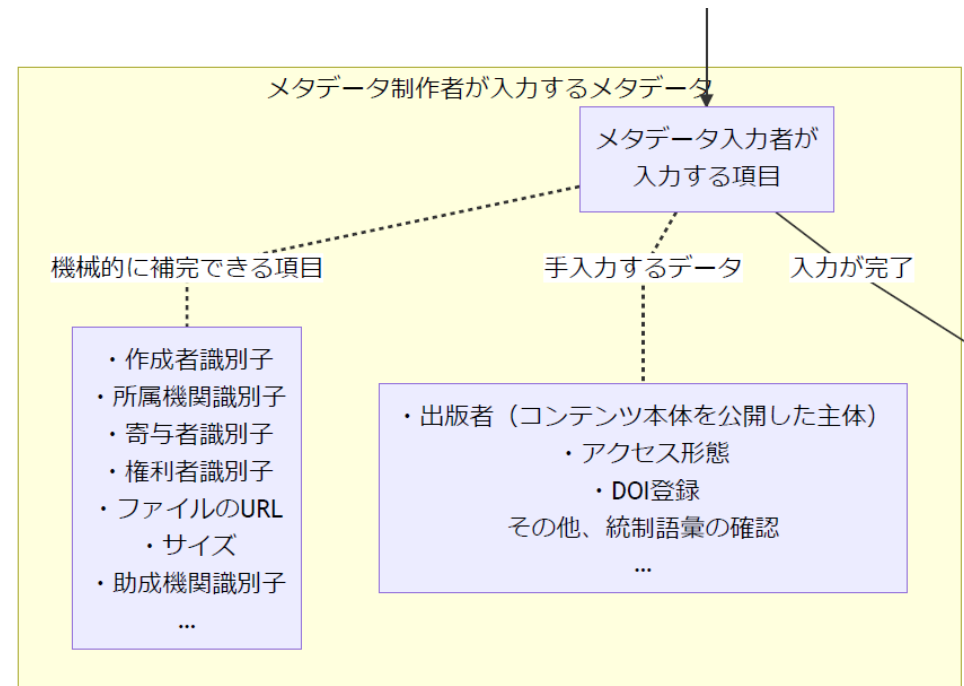
- GakuNin RDMから流用できるもの
 - 作成者の情報
 - 寄与者の情報
 - ファイルの更新日
- 手入力する必要があるもの
 - タイトル
 - 内容記述
 - 助成情報
 - 権利情報
- この後の工程で設定するもの
 - アクセス形態
 - 公開範囲



メタデータ入力フローチャート②

メタデータ入力者が入力する項目

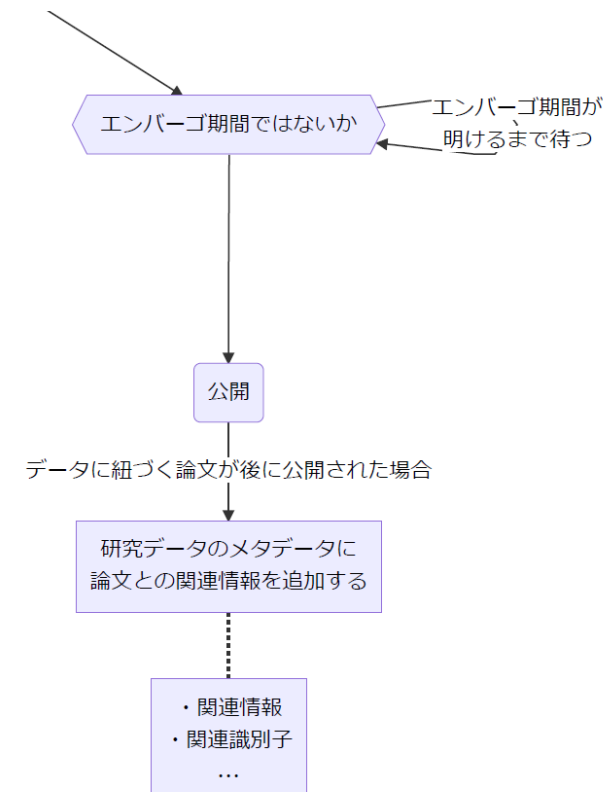
- 機械的に補完できるもの
 - 各種識別子
(対応するWebデータベースに登録されている場合)
 - スキーマに対応したURL
 - ファイルのURL
 - ファイルサイズ
- 手入力する必要があるもの
 - アクセス形態
 - DOI登録を行うための情報
 - その他、統制語彙の確認



メタデータ入力フローチャート③

メタデータ入力終了→公開

- エンバーゴ期間が明けるまで
待ってから公開
- 論文の根拠データかつ
論文の公開前に登録された場合は
後日、論文を関連情報として登録する



サマリー

研究データ公開に際して確認すべき事項

- 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がある問題
- 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
 - 再利用を考えてメタデータを作成

今後検討が必要な課題

メタデータについて

▶バージョンの付け方

- メジャー/マイナーバージョンの基準
- システム内部での更新回数=バージョンとすべきか？
あるいは研究者がコントロールできるようにすべきか？

▶コンタクトパーソンの連絡先

- データ利用時に確認したいことがある場合の連絡先は示されていない
データ作成者を追跡できる仕組みが欲しい（ORCID IDなど）

▶クレジット表記や権利者の範囲が難しい

- 推奨表記を自動生成できると省力化できる 各機関リポジトリの運用？
- 権利者の設定は研究者が決める
ポリシーやガイドライン的なものがあるとよい？

その他のシナリオ

今回の前提は以下の二つだが実際にはもっとある

- 公開したいデータが手元にある
- 出版社のポリシーにより公開が義務化されている

▶ 助成機関による公開義務化

- あらかじめData Management Planを提出している
はずなのでそれにも従う

▶ (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)

- 機関リポジトリの対象は通常、掲載済みの論文
→ 掲載済み論文の研究データも論文の場合と同様に掲載と判断？

▶ 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

- 基準は一意ではない ポリシーで明確にすることが望ましい

今後の展望①

図書館員向けの 研究データ公開支援ワークフローの作成

現状：

- 研究データ公開の手順は明確ではない
- 都度館長の決裁が必要

目標：

- 解消すべき問題を明確に → 業務として持続的に組み込めるように
- 手続きをスムーズに → 今後の体制作りの足掛かりに

研究データ公開の要望はこれからも増えていくはず
→ まずは支援体制を整える

今後の展望②

図書館以外の部署との連携



現状：

- 研究データ公開について北大では図書館以外の部署は消極的
- 公開に他部署の判断が必要になってくる可能性がある

目標：

- 問題別に担当する部署の整理 →業務の切り分けを考える
- 他部署との間での検討 →対応可能性を明確に

研修を振り返って

研修を通じて思ったこと

- とにかく色々な情報を収集しました！
 - 研究データ管理基盤や公開基盤の動向
 - JPCOARスキーマについて
 - 研究データ公開の試行例
 - 研究データポリシーの現状
- (北大に) 帰ってうまく業務に繋がられるかまだ不安
 - まだ全部整理が追い付いていないです...
 - まずは同じ方向に興味関心のある人達にシェアしたい
- 色々見て回るのは楽しい
 - 様々なバックボーンの人と関わる機会が多かった
 - 人によってニーズも様々→図書館もこれからも様々なニーズに対応