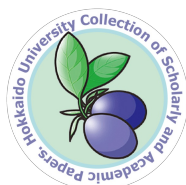




HOKKAIDO UNIVERSITY

Title	令和元年度国立情報学研究所実務研修
Author(s)	三上, 絢子
Description	令和元年度研修出張報告会（特別編）：オープンサイエンス・研究データ関係. 2020年2月4日. 北海道大学附属図書館本館 4階大会議室, 札幌市
Issue Date	2020-02-04
Doc URL	https://hdl.handle.net/2115/76791
Type	lecture
File Information	02_Mikami_1.pdf, 本文



令和元年度 国立情報学研究所 実務研修

北海道大学附属図書館 三上 絢子
2020/02/04 令和元年度第2回出張報告会

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（東京 皇居のすぐ横）
- 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

研修テーマ

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

研究データ活用サイクルの課題

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:

JAIR Cloud



研究データの公開

公開

管理基盤:

GakuNin RDM



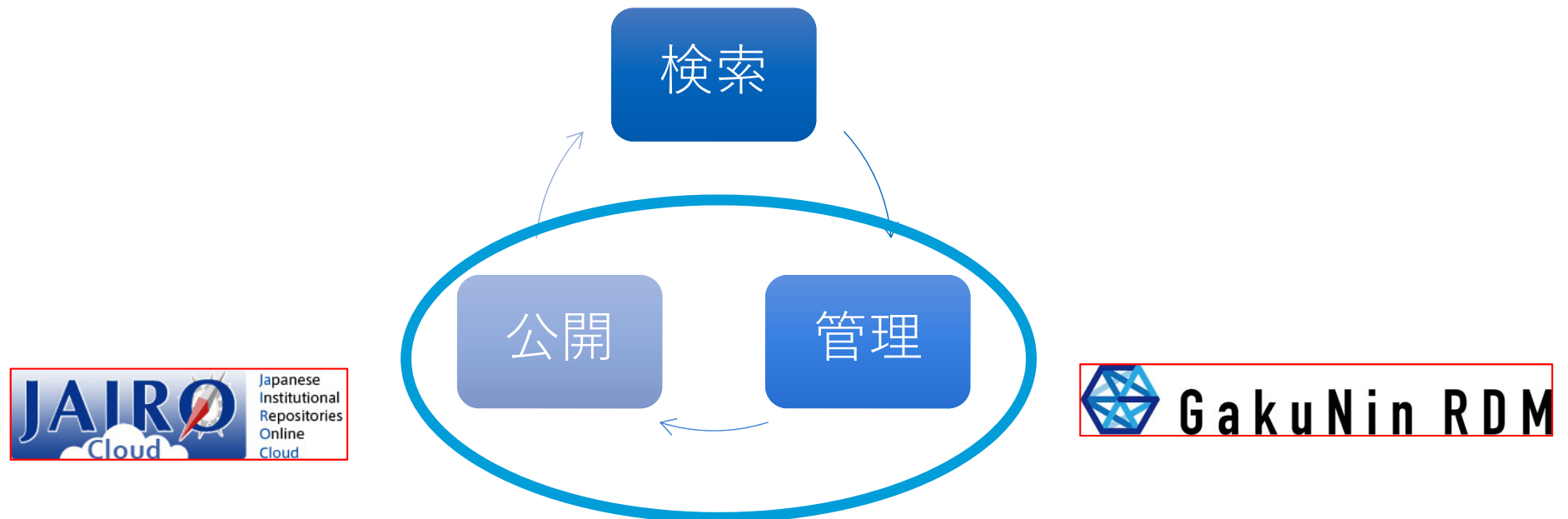
研究データの
生成・解析・保存

管理

?

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



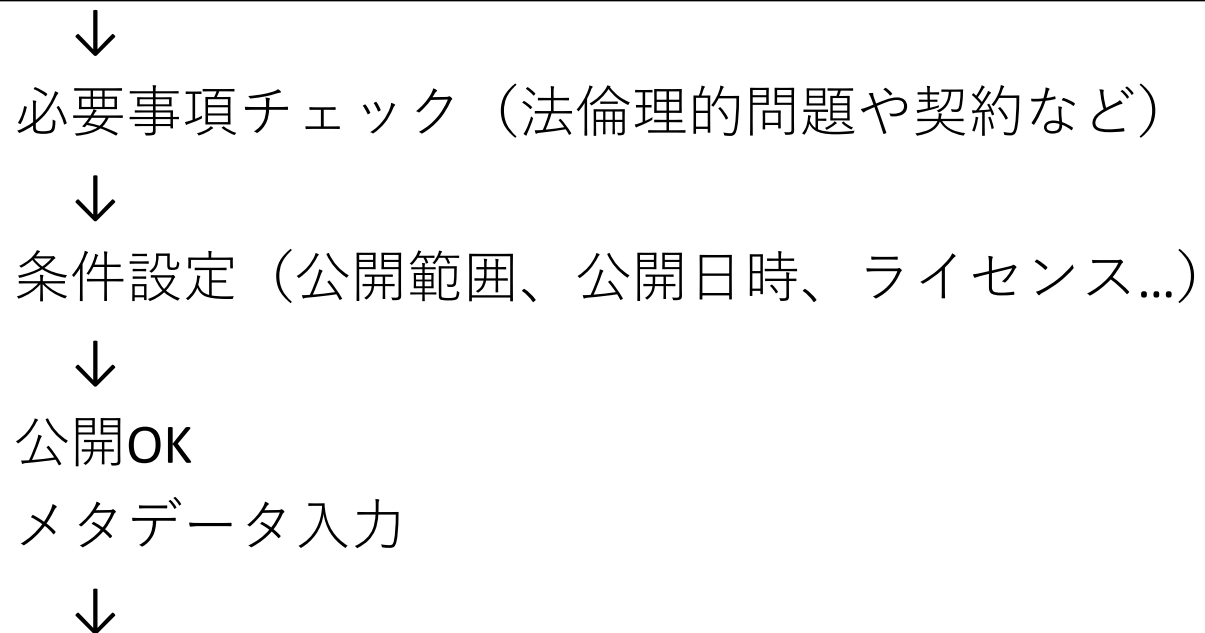
アプローチ方法：管理→公開のワークフロー作成

研修成果：ワークフローの概要

前提（仮定）：出版社のポリシーにより公開が義務化されている

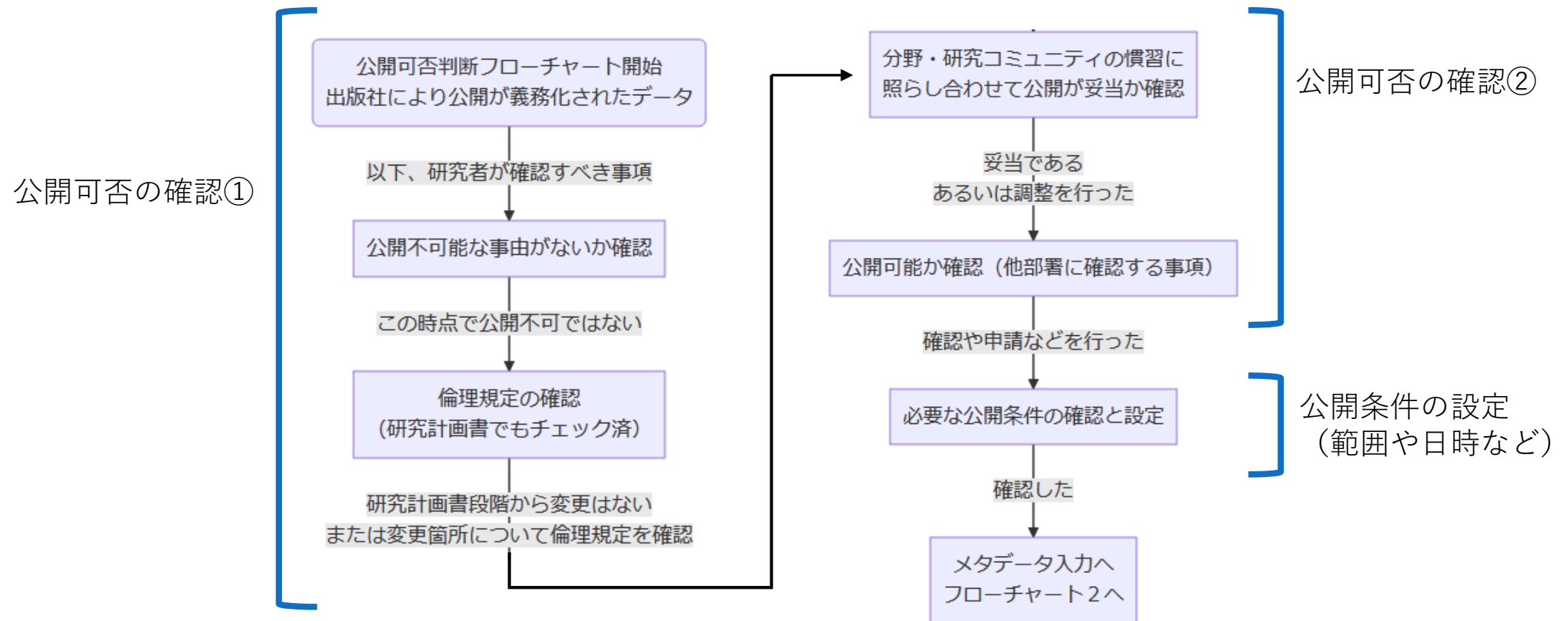
研究データ管理基盤

フローチャート作成
+
メタデータ要素整理
(JPCOARスキーマ
ベースに)



研究データ公開基盤へ

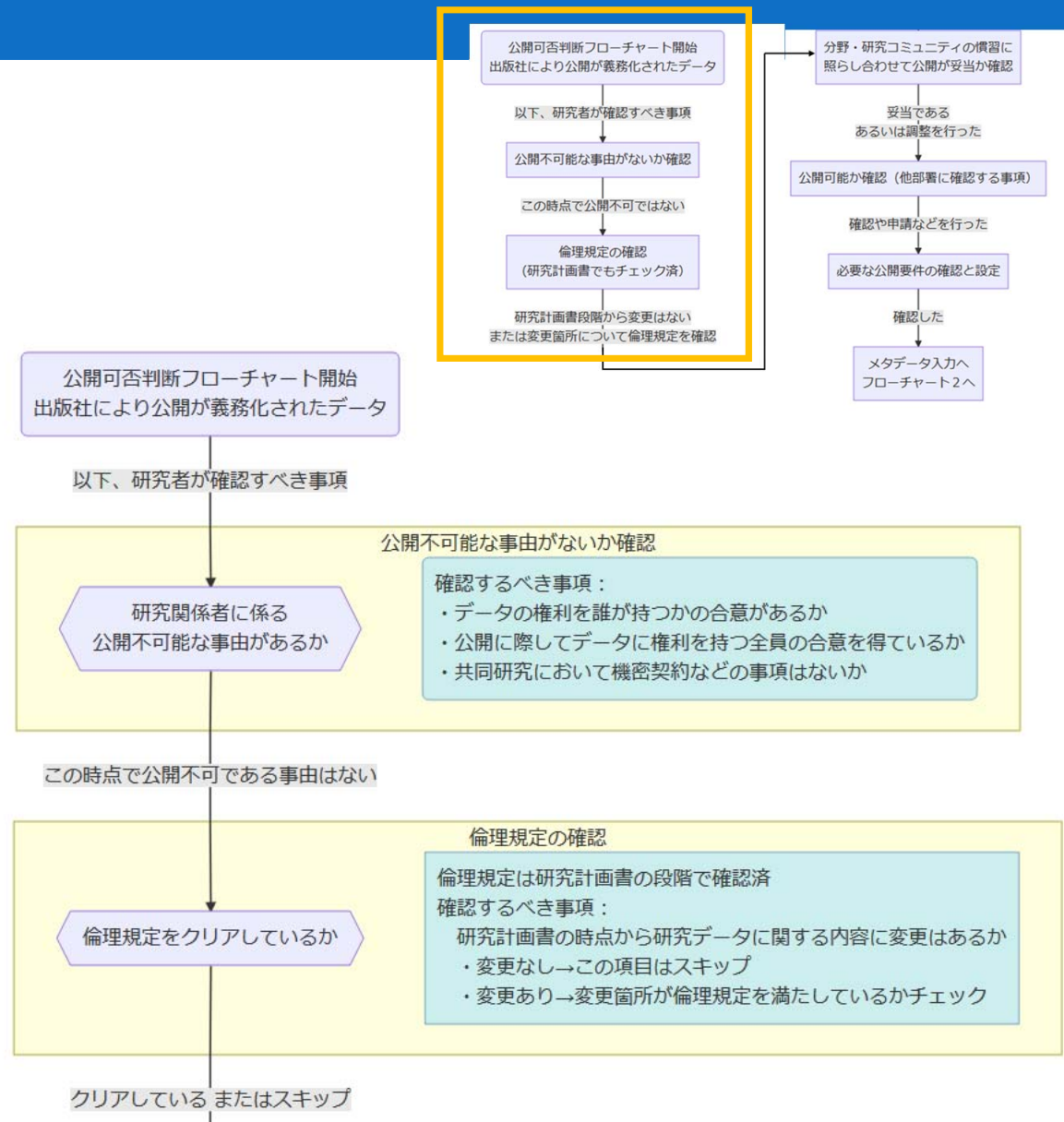
フローチャート図（前半）の概観



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

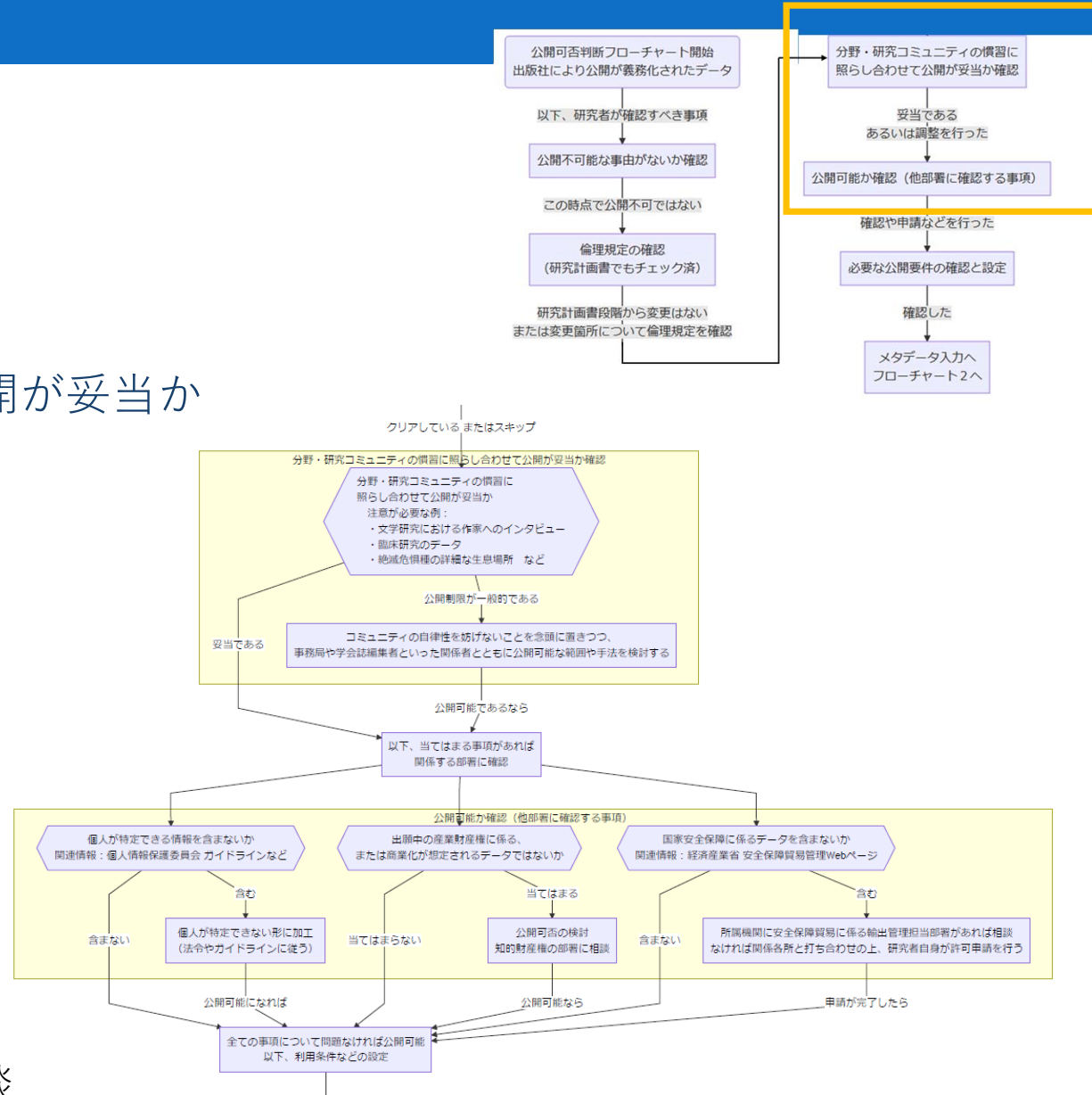
- コミュニティの慣習に照らし合わせ公開が妥当か

- 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

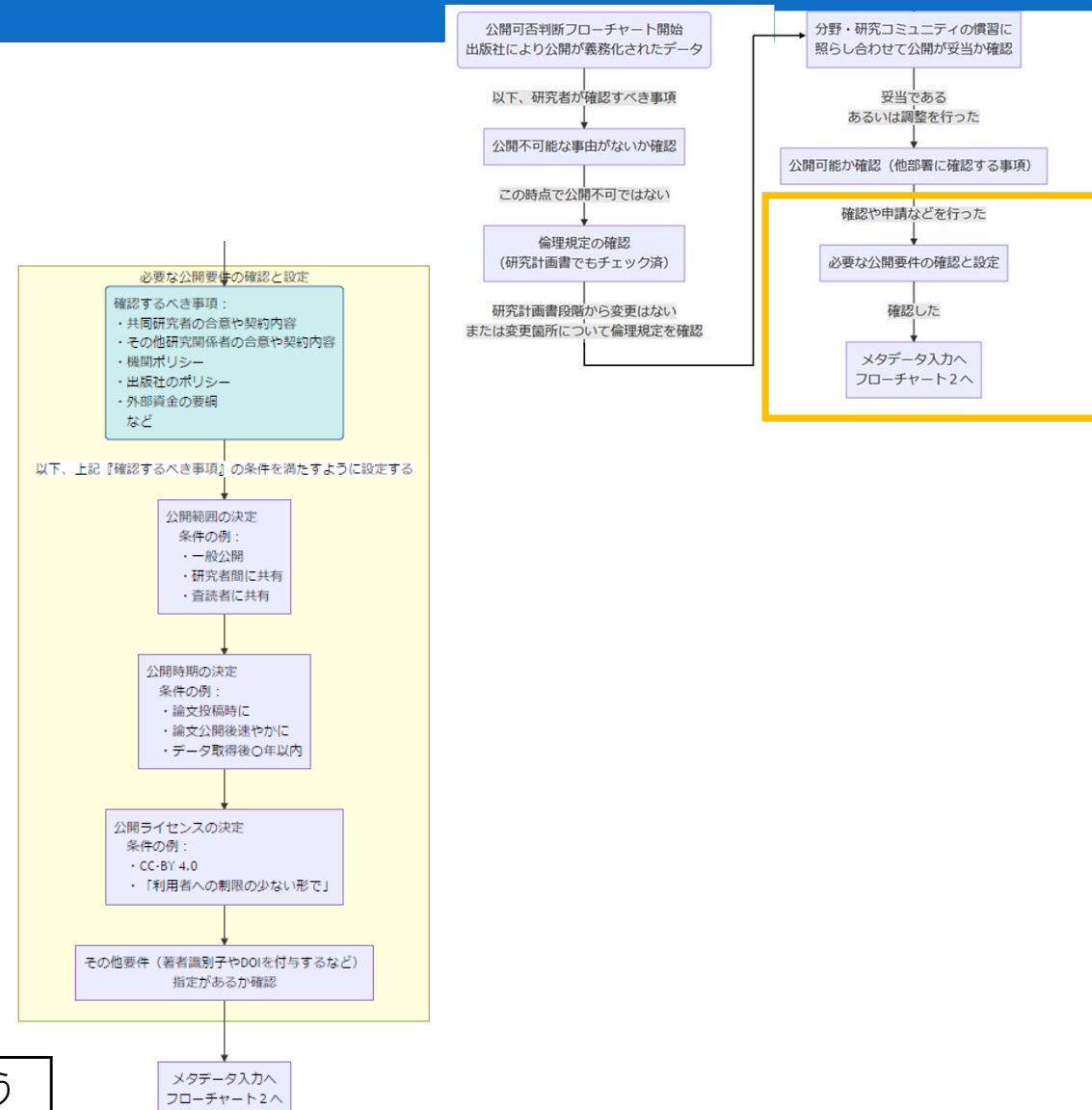


公開フローチャート： 公開条件の設定

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:creatorjpcoar:creatorName xml:lang="ja">安達, 淳</jpcoar:creatorName>
  <jpcoar:affiliationjpcoar: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

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

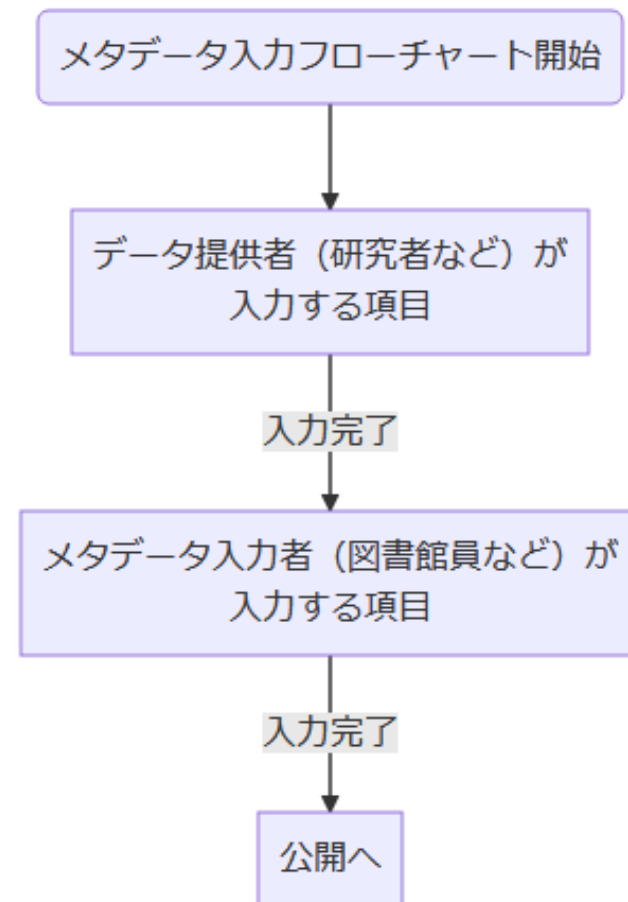
JPCOARスキーマをベースに検討
（Datasetの場合）

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：

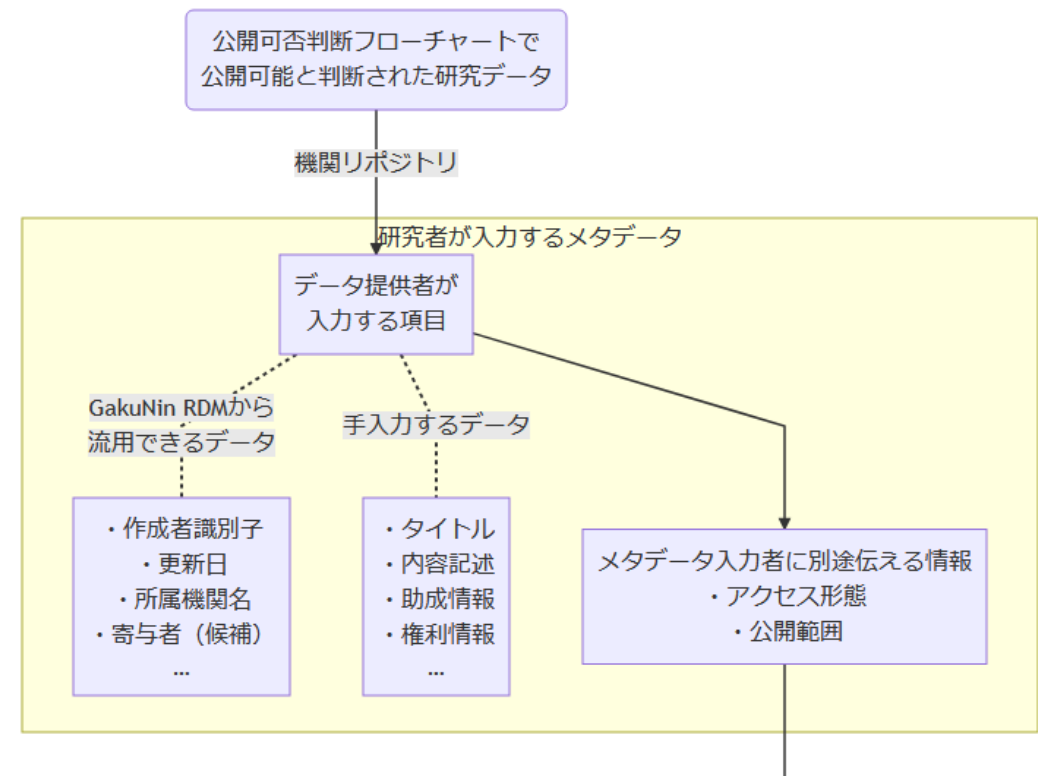
- データ提供者本人しかわからない項目は本人が入力
- その他の項目はメタデータ入力者が入力
- 自動で入力できる項目は自動で
- ORCID IDなどにより自動で入力できる項目も存在



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

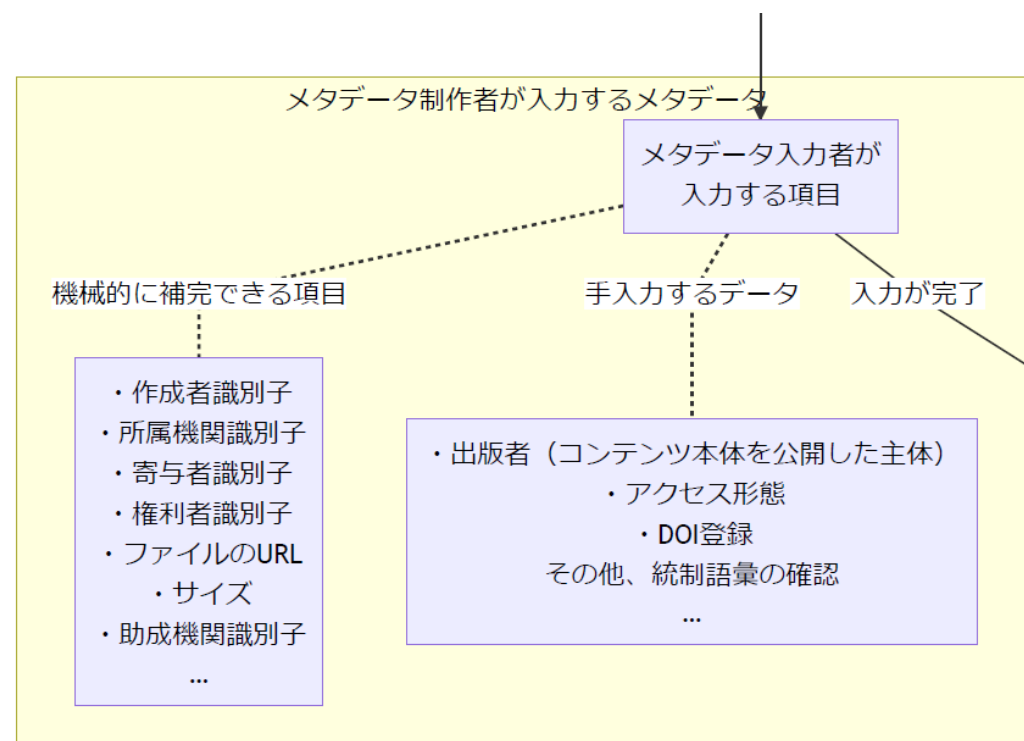
- GakuNin RDMから流用できるもの
 - 作成者の情報
 - 寄与者の情報
 - ファイルの更新日
- 手入力する必要があるもの
 - タイトル
 - 内容記述
 - 助成情報
 - 権利情報
- この後の工程で設定するもの
 - アクセス形態
 - 公開範囲



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

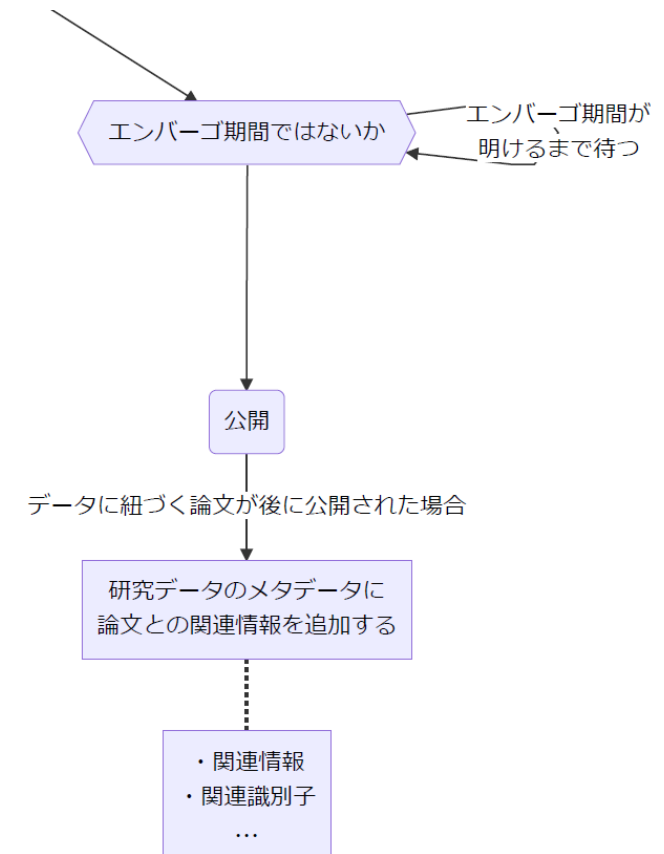
- 機械的に補完できるもの
 - 各種識別子
(対応するWebデータベースに登録されている場合)
 - スキーマに対応したURL
 - ファイルのURL
 - ファイルサイズ
- 手入力する必要があるもの
 - アクセス形態
 - DOI登録を行うための情報
 - その他、統制語彙の確認



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

参考資料

- 研究データの開示・ライセンス指定ガイドライン(草案)
(研究データ利活用協議会 研究データライセンス小委員作成
2020年公開予定 (NII南山さん提供の2019年11月現在のものを参考にしました))
- JPCOARスキーマガイドライン
(<https://schema.irdb.nii.ac.jp/ja>)

本スライドの研修成果は以下でも公開予定

NII実務研修過去の記録
(<https://www.nii.ac.jp/hrd/ja/jitsumu/result.html>)