



北海道大学

2021年 2月18日

第3回 SPARC Japanセミナー2020

研究データ公開実践のための課題を探る

北海道大学での実例を通じて

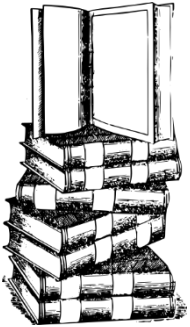


北海道大学附属図書館 研究支援企画担当

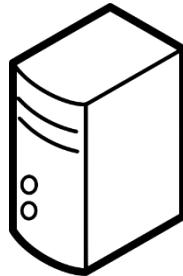
三上 絢子

自己紹介

所属：北海道大学附属図書館
研究支援課
研究支援企画担当

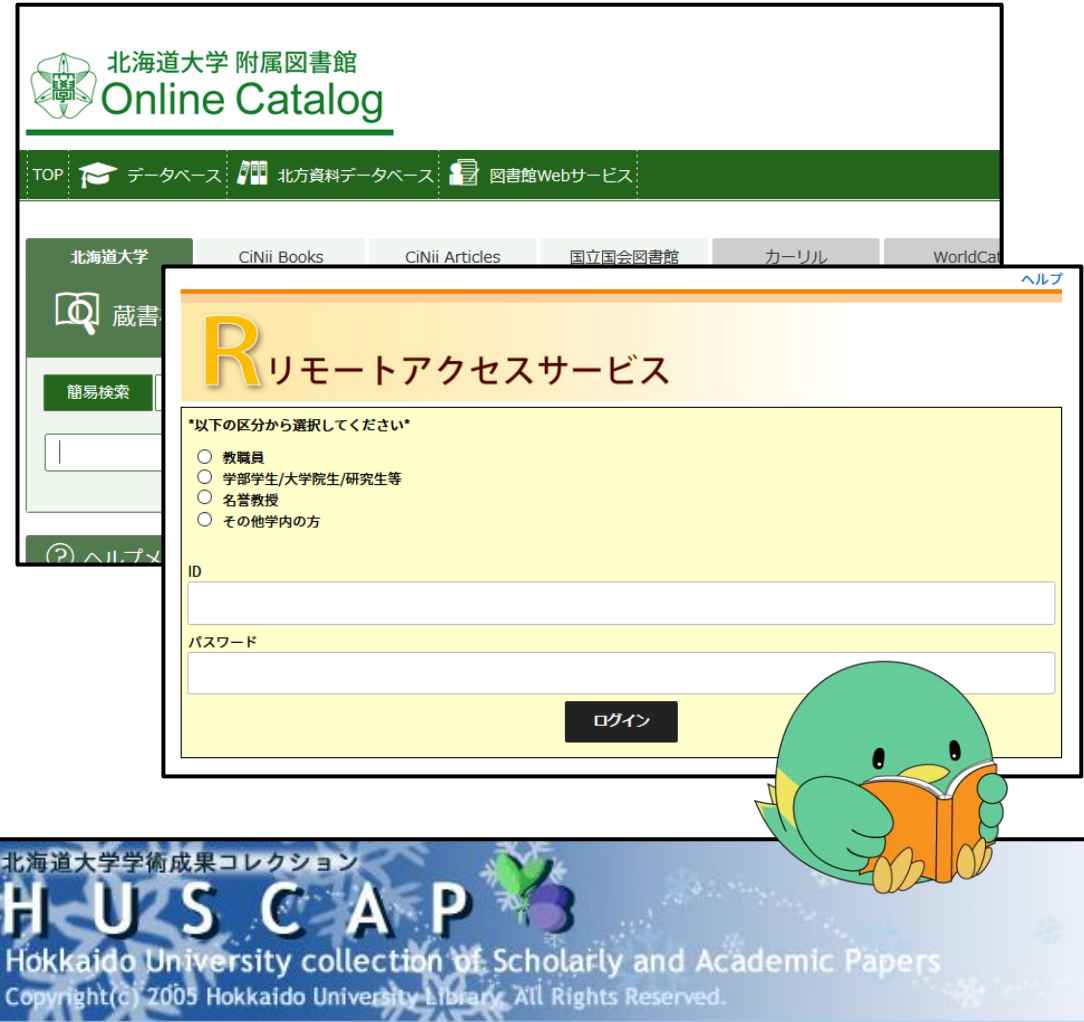


本よりもサーバの方を
よく触る担当です



主な業務：各種オンラインサービスの管理

- ・図書館システム（OPACなど）
- ・電子ジャーナル等のリモートアクセス（学外から）
- ・機関リポジトリ（HUSCAP）



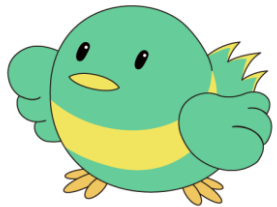
北海道大学学術成果コレクション(HUSCAP)の概要

66,910件の学術資料を公開する機関リポジトリ(2021/02/04現在)

オープンソースソフトウェア「DSpace」をベースに構築している(運用は2005年から)

主な公開資料:

- ・北大所属の研究者から公開依頼のあった資料(論文、教育資料、学会発表資料など)
- ・北大発行の紀要
- ・研究データ(少数)





北海道大学

研究データ公開事例の紹介

北海道大学学術成果コレクション(HUSCAP) 2017～2020

研究データ公開の件数

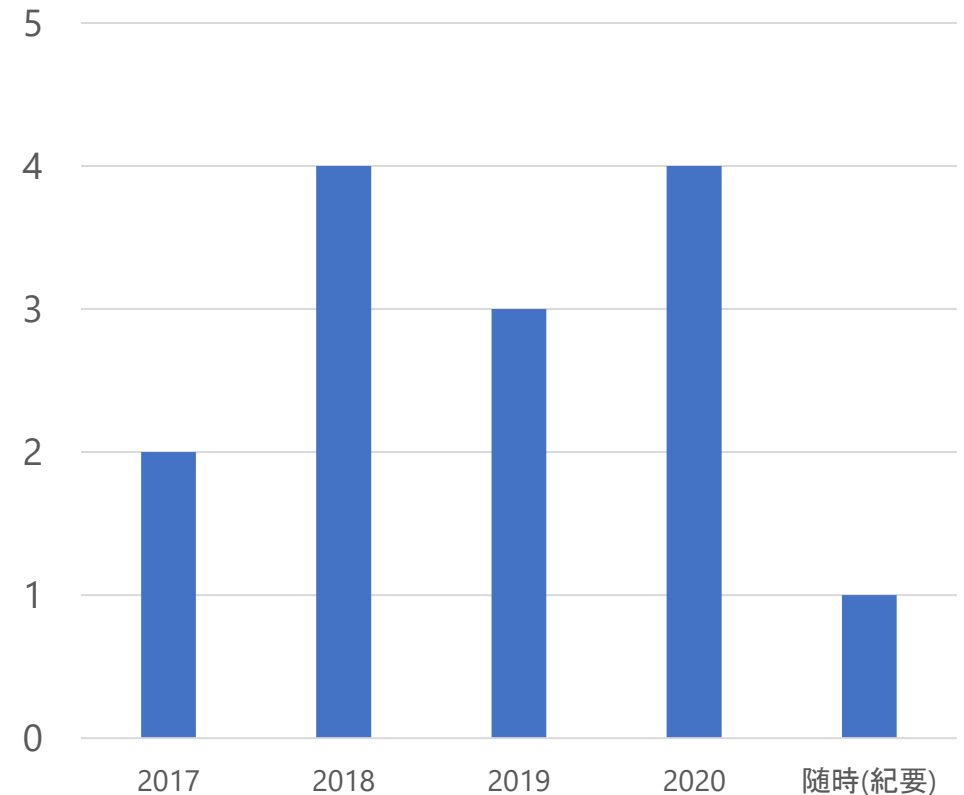
本事例紹介で扱う「研究データ」について

北海道大学の機関リポジトリ(HUSCAP)上で
公開された資料のうち、
資料種別が「**dataset**」のもの

2017年から2020年で14件

- うち、1-2年間隔で発行の
水産学部紀要が1件
- 残りは雑誌投稿論文に関連するデータ

2016年以前に公開したデータはない



2017年から2020年にかけての研究データ公開件数



依頼元の研究者が所属する部局

部局	件数
低温科学研究所	7
医学研究院	3
その他医系部局	2
理学院	1
水産学部紀要	1

投稿雑誌

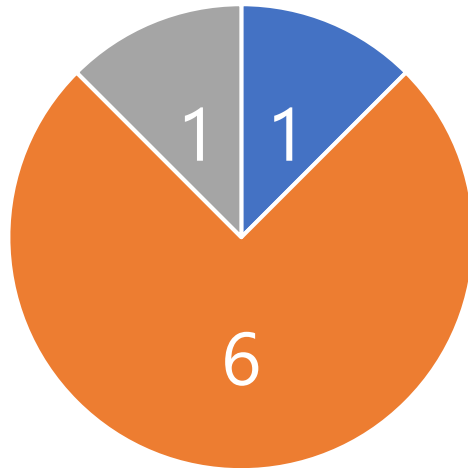
低温科学 研究所	Journal of Geophysical Research: Atmospheres
	Mammal study
	Earth and Planetary Science Letters
	PNAS
	Polar Science
医学研究院	Behavioural brain research
	Scientific data
	Molecular Cell
その他医系部局	The Journal of Clinical Endocrinology & Metabolism
	Tectonics
	海洋調査漁業試験要報
理学院	
水産学部紀要	



データ公開の依頼時期

大きく分けて2パターン: 関連する論文の出版前/出版後

出版前のデータ登録依頼

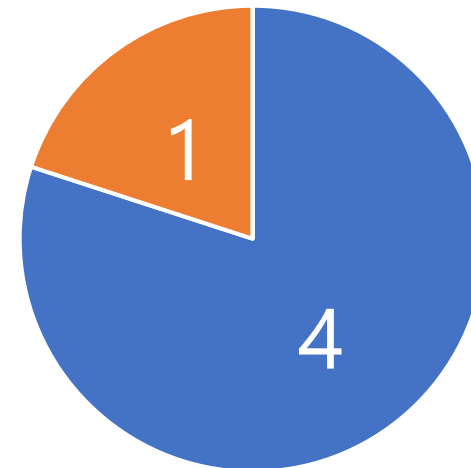


■ 医学系 ■ 低温研 ■ 理学院

出版前公開の理由

- ・アクセプト条件として出版前にデータを公開する必要があった
- ・アクセプトのために査読者とデータを共有する必要があった

出版後のデータ登録依頼



■ 医学系 ■ 低温研

出版後公開の理由

- ・Supplemental Dataが付属しているOA論文をアーカイブする際にそのデータも紐づけて登録した



事例の紹介

公開事例をパターン別に紹介

パターン1: 投稿前の論文の根拠データを公開

1-1: 校正時にURL確定

1-2: 査読時に査読者と共有

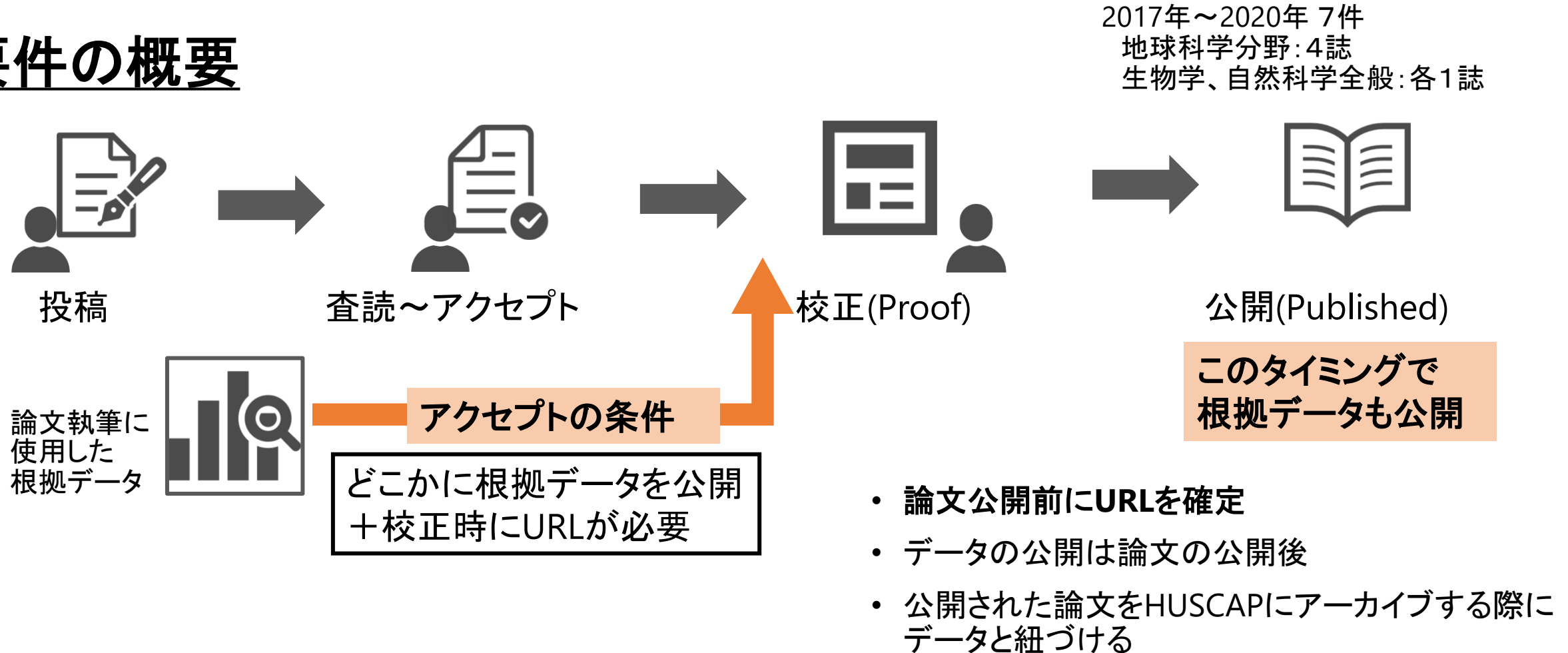
1-3: 査読時に一般公開

パターン2: 投稿後の論文に付随するデータを公開



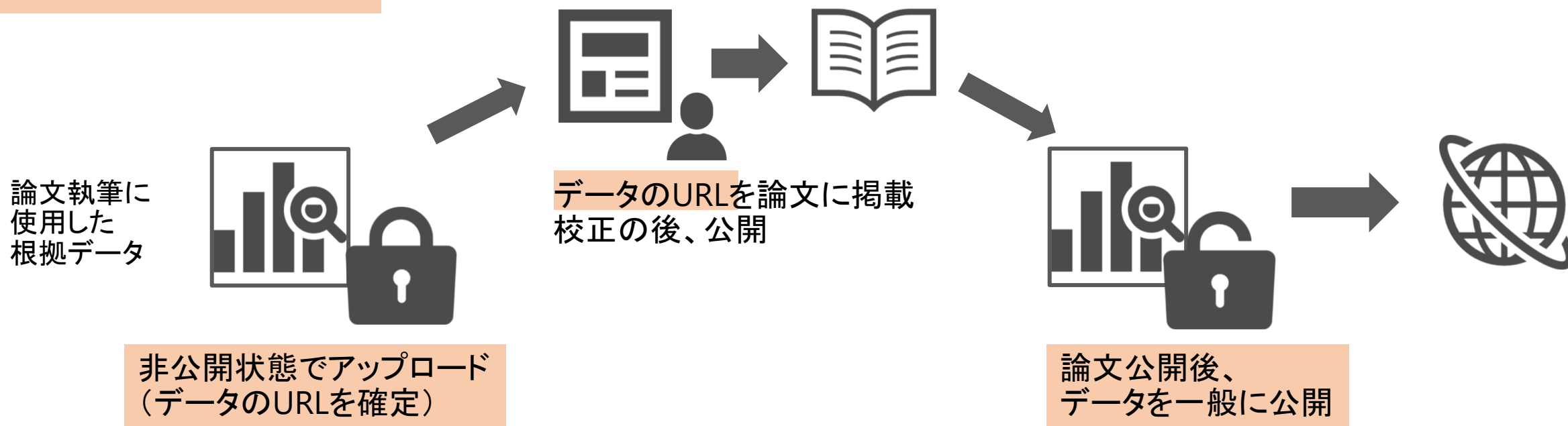
パターン1-1: 校正時にURL確定

要件の概要



パターン1-1: 校正時にURL確定 対応例

機関リポジトリで行った対応

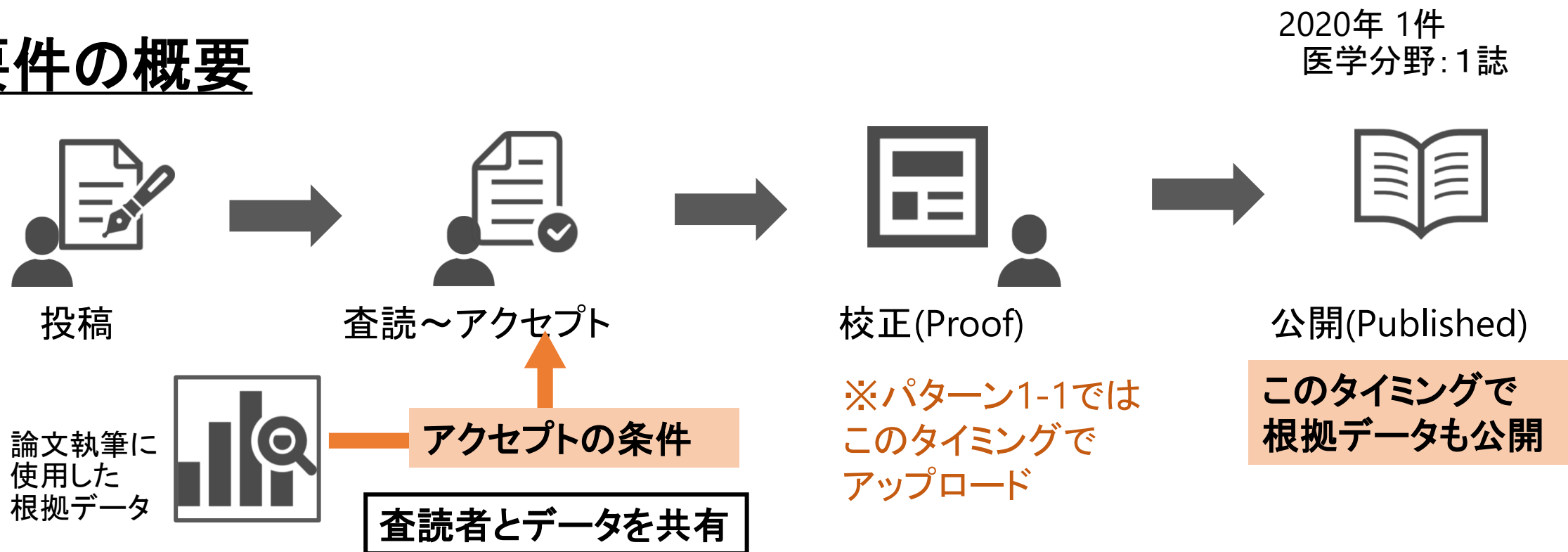


- 論文公開前にURLを確定 (HUSCAP上の非公開の場所にアップロード)
- データの公開後、依頼された先生から連絡を貰う
 - データをWeb上で一般公開
 - メタデータ+リンク作成



パターン1-2: 査読者と共有

要件の概要



2020年 1件
医学分野: 1誌

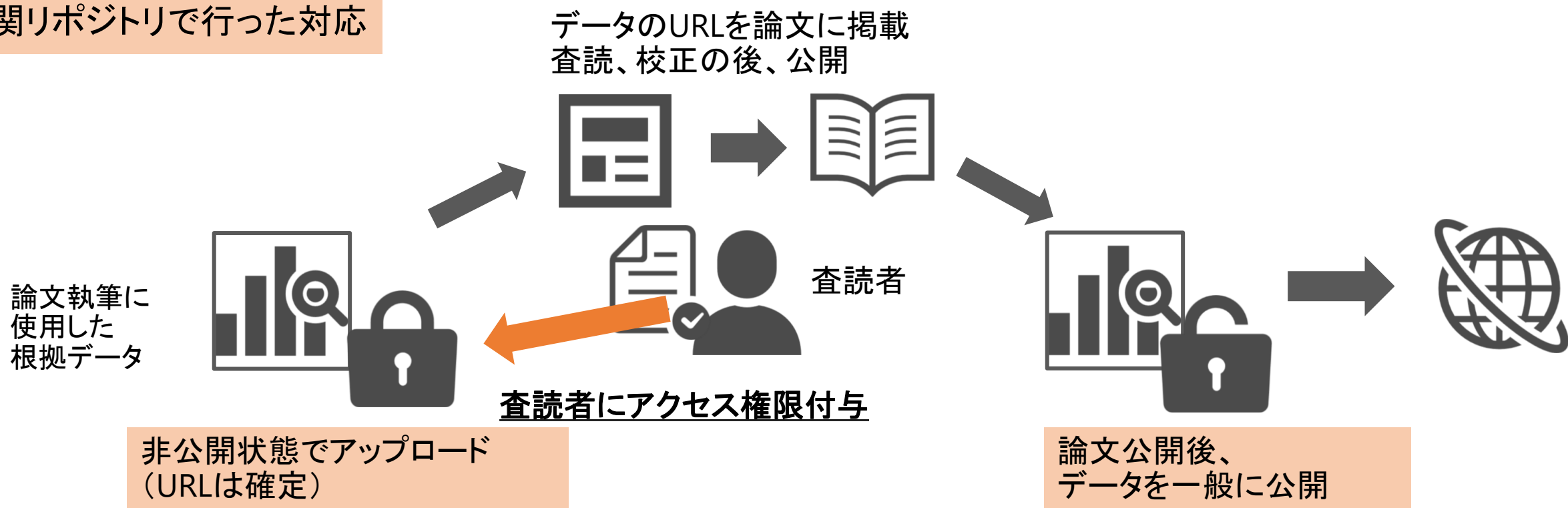
パターン1-1との違い:

一般公開ではなく、特定の相手にのみ公開したい



パターン1-2: 査読者と共有 対応例

機関リポジトリで行った対応



論文公開前にURLを確定(HUSCAP上の非公開の場所にアップロード)

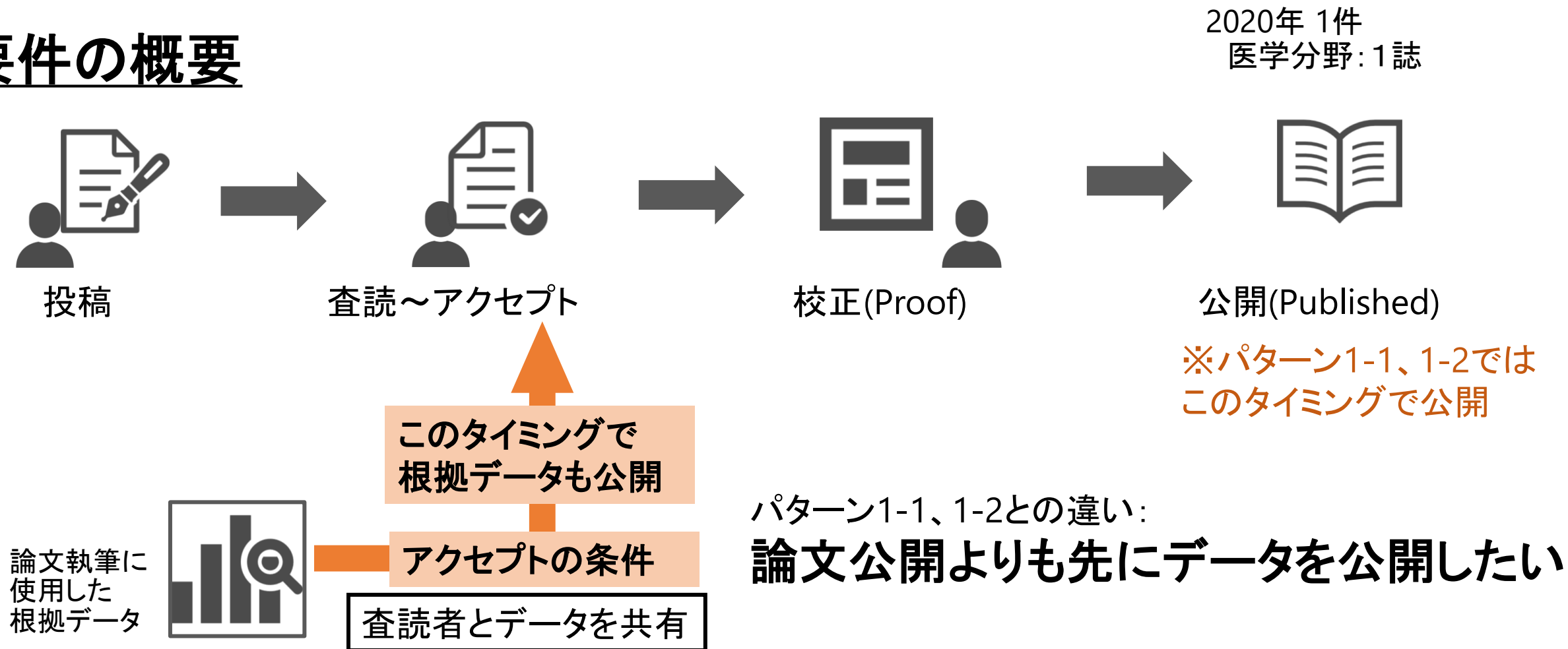
管理者アカウントは非公開の場所を読み書きできる(権限は個別に設定可能)

→非公開場所のうち、指定したアイテムだけ読み取り可能な管理者アカウントを作成→査読者に伝える



パターン1-3: 査読時に一般公開

要件の概要



パターン1-3: 査読時に一般公開 対応

ここで、現在のHUSCAPの公開細則にひっかかる
(明文化された運用ポリシーにおいて公開が想定されていない)

HUSCAPのポリシーは原則として、「公表済資料」の公開を前提とする

これまでは公表済論文のアーカイブを想定(GreenOA)

→未公表論文はあまり想定する必要がなかった

しかし、今回のデータ元の論文は公表済ではない

→ 付随するデータも未公表資料

なので現状では都度、決裁して対応
(「特に館長の認めた資料」という扱い)

※ポリシー改訂予定

(4) 資料の種別として以下のいずれかに該当するもの。

(ア) 公表済資料

① 商業出版社もしくは学協会が発行する、冊子体または電子的な学術雑誌

(中略)

(エ) その他、特に館長の認めた資料

[北海道大学学術成果コレクション資料提供・公開細則](#)より



パターン2: 論文アーカイブに伴った付随データの公開

要件の概要

データと合わせて公開されていた論文をHUSCAPに登録する際に
付随するデータも一緒に登録するパターン



既に公開済の論文

論文の登録希望



論文に
付随する
データ

または



論文の登録依頼
& 承諾

元々の公開場所

- ・出版社ページ
- ・Figshare
- ・Mendeley Data



パターン2: 論文アーカイブに伴った付随データの公開 対応

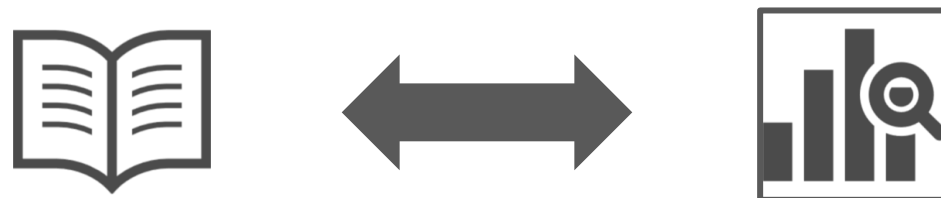
既に公開されている論文＋データのアーカイブの場合は
従来の論文アーカイブと同様に登録、公開
(一番、判断がスムーズなパターン)

ちなみに登録の経緯は大きく分けて以下の2つ

- 研究者から依頼があったもの
- 図書館側から登録依頼→研究者から承諾があったもの
(Web of Scienceのデータを元に主要な著者が北大所属の論文について定期的に抽出 & 依頼)



パターン1、パターン2に共通する対応：論文とデータの関連付け



公開後、論文とデータを関連付ける

メタデータで論文とデータを関連付け（「Relation」要素の属性「isRequiredBy」）

(Relation)isRequiredBy: [Original reference: Ryoto Furukawa, Ryu Uemura, Koji Fujita, Jesper St. Yoshimura, Sumito Matoba, and Yoshinori Iizuka, Seasonal scale dating core from Greenland using oxygen isotope matching between data and Journal of Gephysical Research Atmospheres, accepted Aug. 2017](#)

元論文のDOIを入力
→出版社ページへリンク

HUSCAP内に論文がある場合は別途、内部リンクを貼る

HUSCAP内関連資料

- Related data set
- [\[Dataset\]SE-Dome ion concentration aerosols\(Iizuka2017\)](#)
- Article
- [Seasonal-Scale Dating of a Shallow Ice Core From Greenland Using Oxygen Isotope Matchi](#)

→HUSCAP内のページへリンク



その他: datasetとして公開されている紀要

水産学部紀要「海洋調査漁業試験要報」
(1957-2016: 冊子体で発行 2017～: 電子版のみ)

内容: 練習船の航行記録、センサーデータ、捕獲した魚類の種類/数など
PDF形式(冊子体時と同様)



その他: Dataset Searchへの対応(全パターン共通の対応)

GoogleのDataset Searchに対応するように個々の資料の
メタデータページ内にJSONを仕込む

```
33 <script type="application/ld+json">[  
34   "@context": "http://schema.org", "@type": "Dataset", "author": { "@type": "Organization", "name": "北海道大学大学院水産科学研究  
35 ]</script>
```

所定の書式をセット
→Dataset Searchに
無事収録された

The screenshot shows a search result for a dataset. At the top, there are tabs for '最終更新日', 'ダウンロード形式', 'ライセンス', 'トピック', and '無料'. Below these, a message states '100 件以上のデータセットが見つかりました'. The main result is for 'PART I. OSHORO MARU CRUISES FROM MARCH TO DECEMBER 2016' by 'eprints.lib.hokudai.ac.jp', with a 'pdf' icon and a '更新日: 2020/03/31'. To the right, there is a detailed view of the dataset, including a search button '探す: eprints.lib.hokudai.ac.jp', a note 'このデータセットを引用している学術記事: 3 件 (Google Scholar で表示)', a 'pdf' icon, and additional information: 'データセット更新日 2020/03/31', 'データセットの作成・提供組織: 北海道大学大学院水産科学研究院・水産学部 練習船運営委員会', and a '説明' link.



研究データ公開例のまとめ

2017年～2020年に登録した14例(論文の根拠データ13例、紀要1例)
論文の根拠データの公開手順はいくつかのパターンに分けられた

公開手順を判断する要素

いつ、登録/公開するか？

- 論文アクセプト前(査読時)
- 論文アクセプト後(校正時)
- 論文公開後

誰と共有するか？

- 非公開状態
- 査読者とのみ共有
- 一般公開

公開手順に関係する事柄

制限公開の機能

- URLのみ確定させ非公開の状態
- 特定の相手とのみ共有

研究データ公開を想定した運用要綱

データと論文の関連付け

検索エンジン(Dataset Search等)への登録



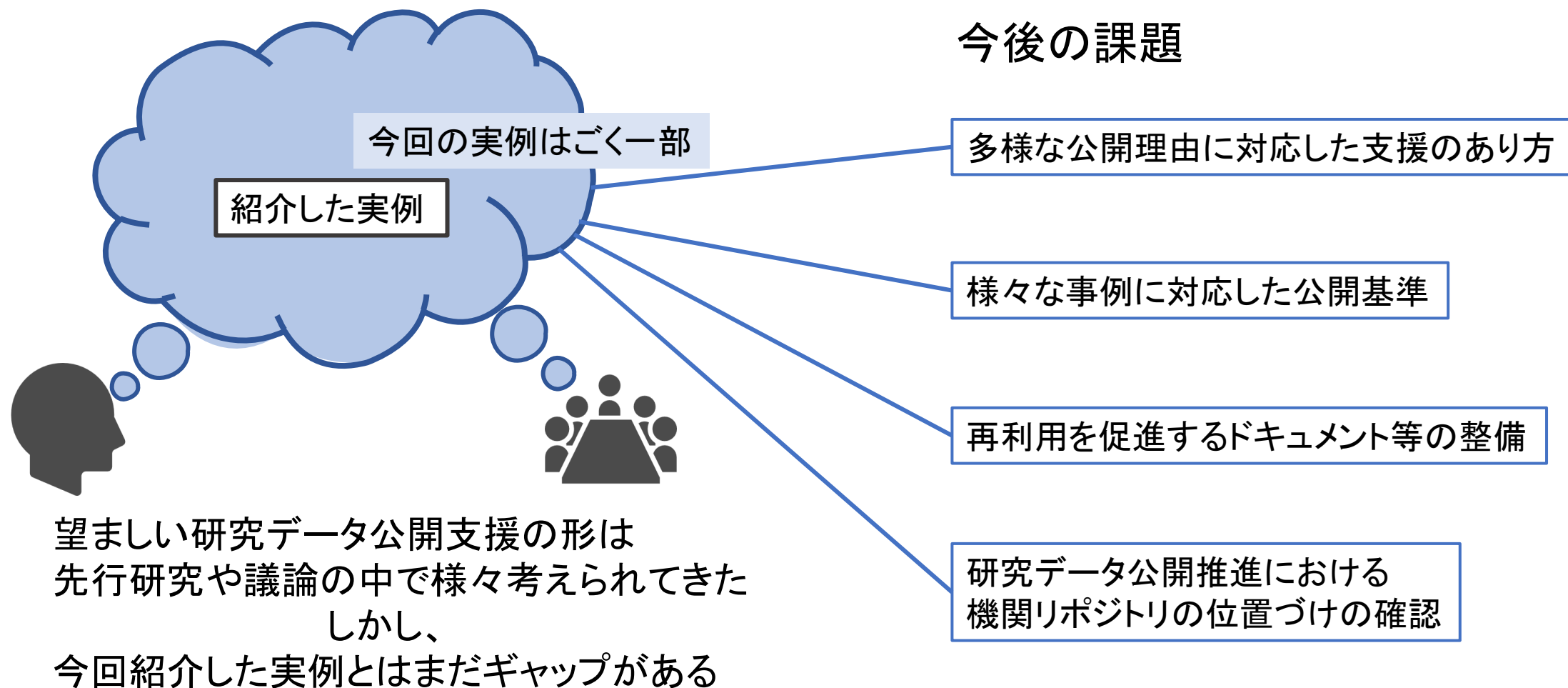


北海道大学

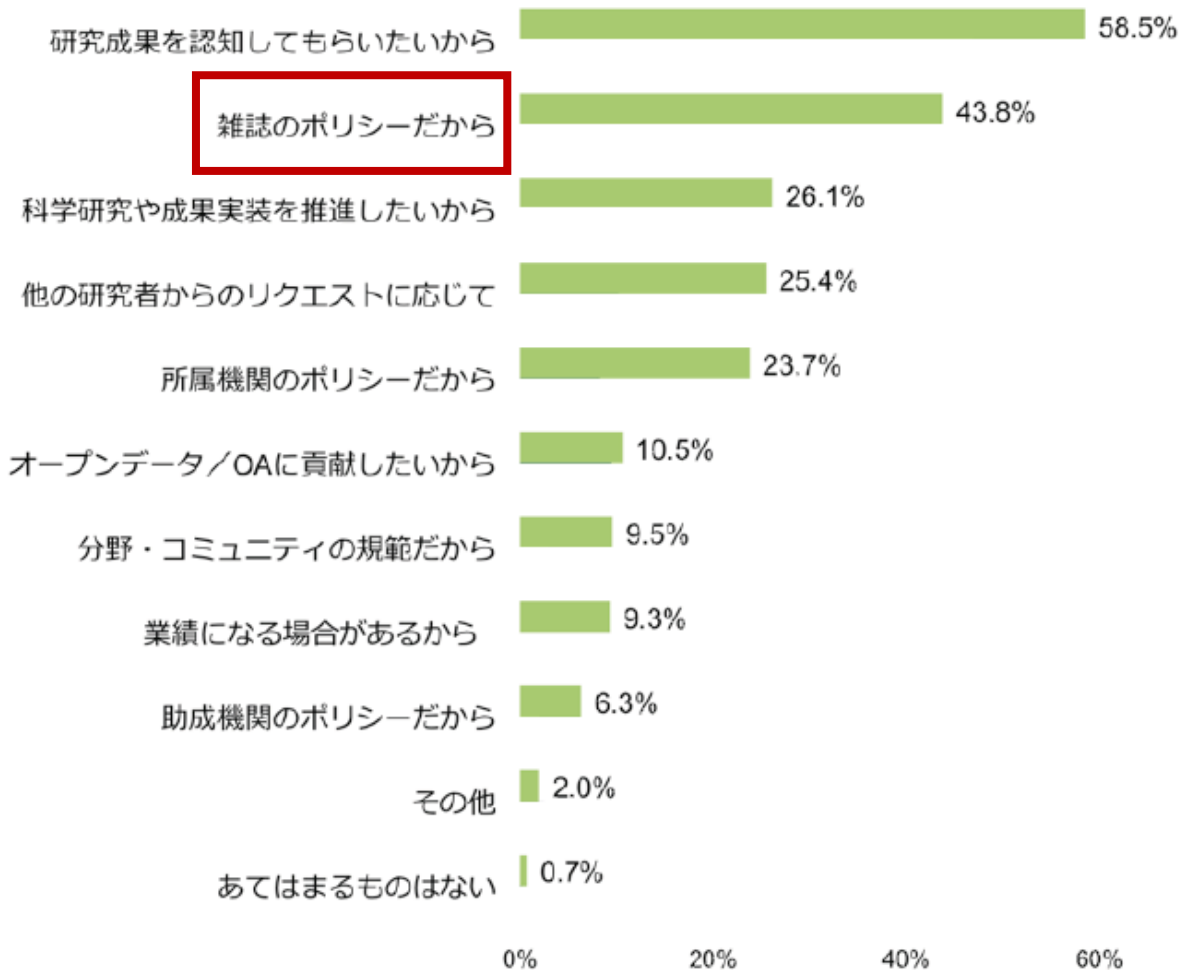
研究データ公開事例から考える今後の課題



今までの事例とこれまでの議論とのギャップから見てくる今後の課題



多様な公開理由に対応した支援のあり方



今回の事例のほぼ全て(紀要除く)のデータ公開理由は「雑誌のポリシーだから」であった

これまでデータ公開を行った経験のある研究者の中にはその他の理由によるものもいることが(池内ら,2017)の調査で明らかになっている(左図)

疑問

その他の理由・モチベーションに対応したリポジトリの運用方針や公開手順などはまた別に検討が必要だろうか？

その他の理由で公開を希望する研究者が、研究データ公開に求める要件には何があるだろうか？

様々な事例に対応した公開基準

今回の実例の場合：

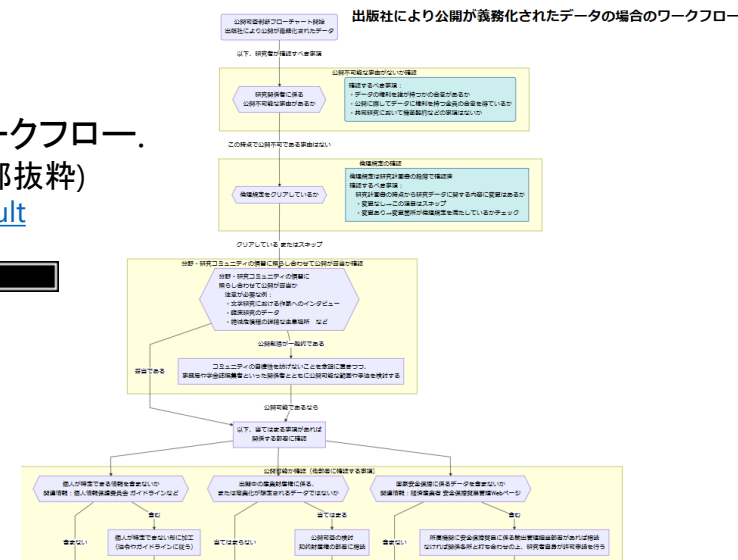
特段、データの公開可否の判断をしていない
(明確な判断基準も現時点では存在しない)

データによっては公開の際に 確認すべき問題があることが予想される

- 権利表記の合意(誰を作成者とするか？等)
- 共同研究者との契約上、守秘義務が生じる場合
- ゲノム情報など分野ごとのガイドラインが存在する場合
- 個人情報を含む場合
- 知財関連(特許など)が絡む場合

判断に専門知識を必要とする問題もある

公開可否判断ワークフロー例：
三上 絢子, 研究データ公開ワークフロー.
(2019年度NII実務研修成果物、一部抜粋)
<https://hrd.nii.ac.jp/jitsumu/2019/result>



疑問

誰が問題を判断+解消すべきか？

支援部署だけでの対応は難しい？

研究者へ: あらかじめ注意すべき点を周知？

専門部署(知財等)へ: 事前に対応可能性を協議？

複数の利害関係者間で問題意識を共有し、協働できるか？



再利用を促進するドキュメント等の整備

今回の公開事例では特別な公開手順は定めていない(論文同様、提供ファイルをリポジトリで公開)

- ファイルの命名規則は特にない(論文公開も同様)
- データの内容を説明するドキュメントは特にない
- 論文のAcceptに必要→出版社の定めた要件を満たしていればよい
- データの中身については研究者に任せている
- 再利用については考慮していない

研究データ公開で考慮すべき点 (JPCOAR研究データタスクフォース作成「RDMトレーニングツール」(2017)より)

- ファイルのフォーマット
- **ファイルの組織化(命名規則)**
- **再利用のためのドキュメントの整備**
- メタデータの付与
 - 発見性やアクセス性の向上
 - **再利用時のライセンスの明示**

※現時点で考慮できていないものを朱色で表記

疑問

従来の公開手順では考慮できない点に対し、
どのように対応していくか？

出版社等の要件以上のこと、例えば「再利用の
利便性を高めるため、ドキュメント整備を行う」などの
場合の研究者/支援者のモチベーションは何か？

こういった手順でそのモチベーションに対応できるか？



研究データ公開推進における機関リポジトリの位置づけの確認

HUSCAPでのデータ公開を選んだ事情

第一候補だった分野別リポジトリはデータフォーマットが厳密に決められていたため、制約のないHUSCAPで公開することを希望した

既に他の場所で公開されているデータのアーカイブ

公開場所例: FigShare、MendeleyData、出版社のリポジトリ

図書館で研究データを公開したいという問い合わせを受け、FigShare等の公開リポジトリと併せてHUSCAPでも公開できる旨を伝えた

初めからHUSCAPで公開希望との連絡があった

その他のリポジトリでの公開状況は
現時点では把握していない

疑問

論文公開における機関リポジトリの位置づけ
→GreenOAに対応したアーカイブ公開場所

データ公開における機関リポジトリの位置づけとは何か？
(同じくデータ公開を担う分野別リポジトリ、公開リポジトリとの違いは何か)



模索は始まったばかり

今後の課題

多様な公開理由に対応した支援のあり方

様々な事例に対応した公開基準

再利用を促進するドキュメント等の整備

研究データ公開推進における
機関リポジトリの位置づけの確認

とそれぞれについて現時点の疑問

事例紹介



新しい形を目指した雪だるまを作るはすかつぷちゃん
(HUSCAPのマスコットキャラクター)

