

# データリポジトリの信頼性と要求要件に 関する調査検討状況報告

## 研究データ利活用協議会リポジトリ小委員会

**絹谷 弘子** (東京大学地球観測データ統融合連携研究機構)

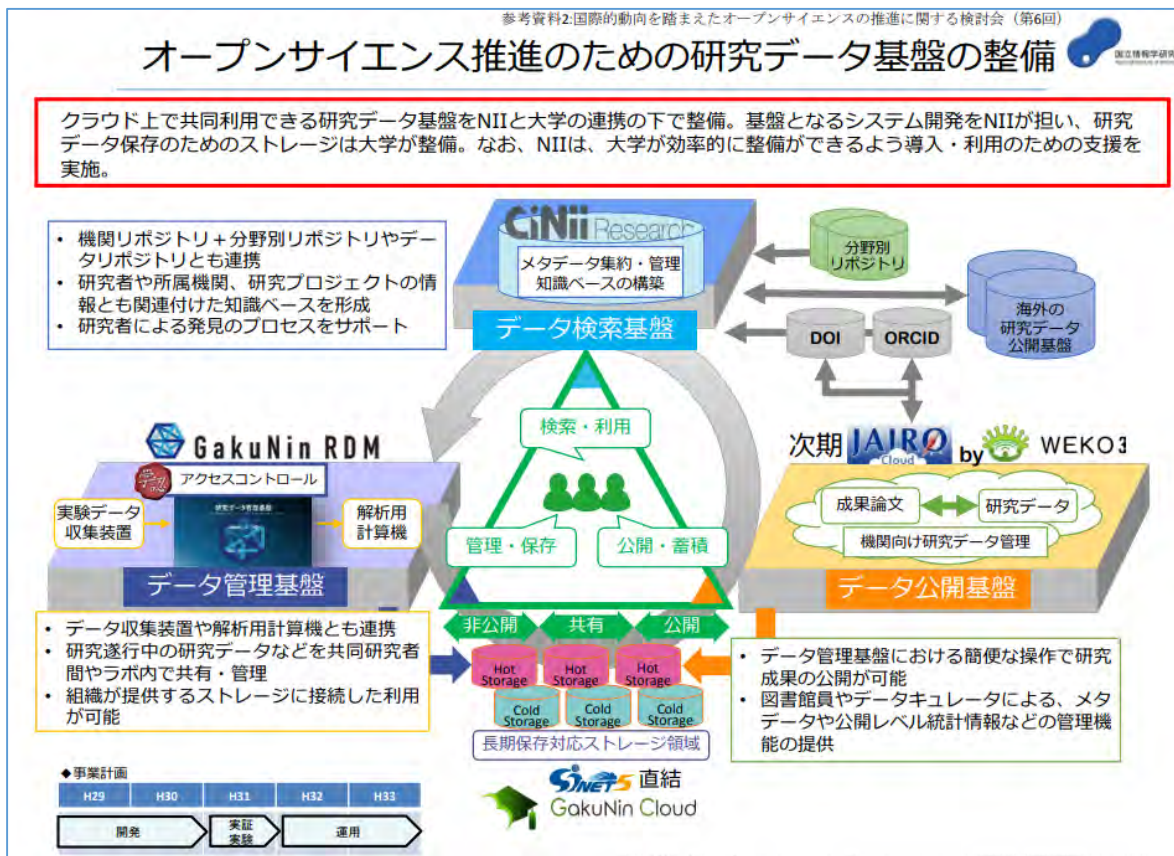
**北本 朝展** (情報・システム研究機構 データサイエンス共同利用基盤施設  
人文学オープンデータ共同利用センター / 国立情報学研究所)

## 発表内容

- 背景
- リポジトリ小委員会(JDARN)とは？
- データリポジトリの信頼性
- データリポジトリの要求要件に関する考察、検討
- 委員会活動のふりかえりと今後

# 背景

# 国内動向



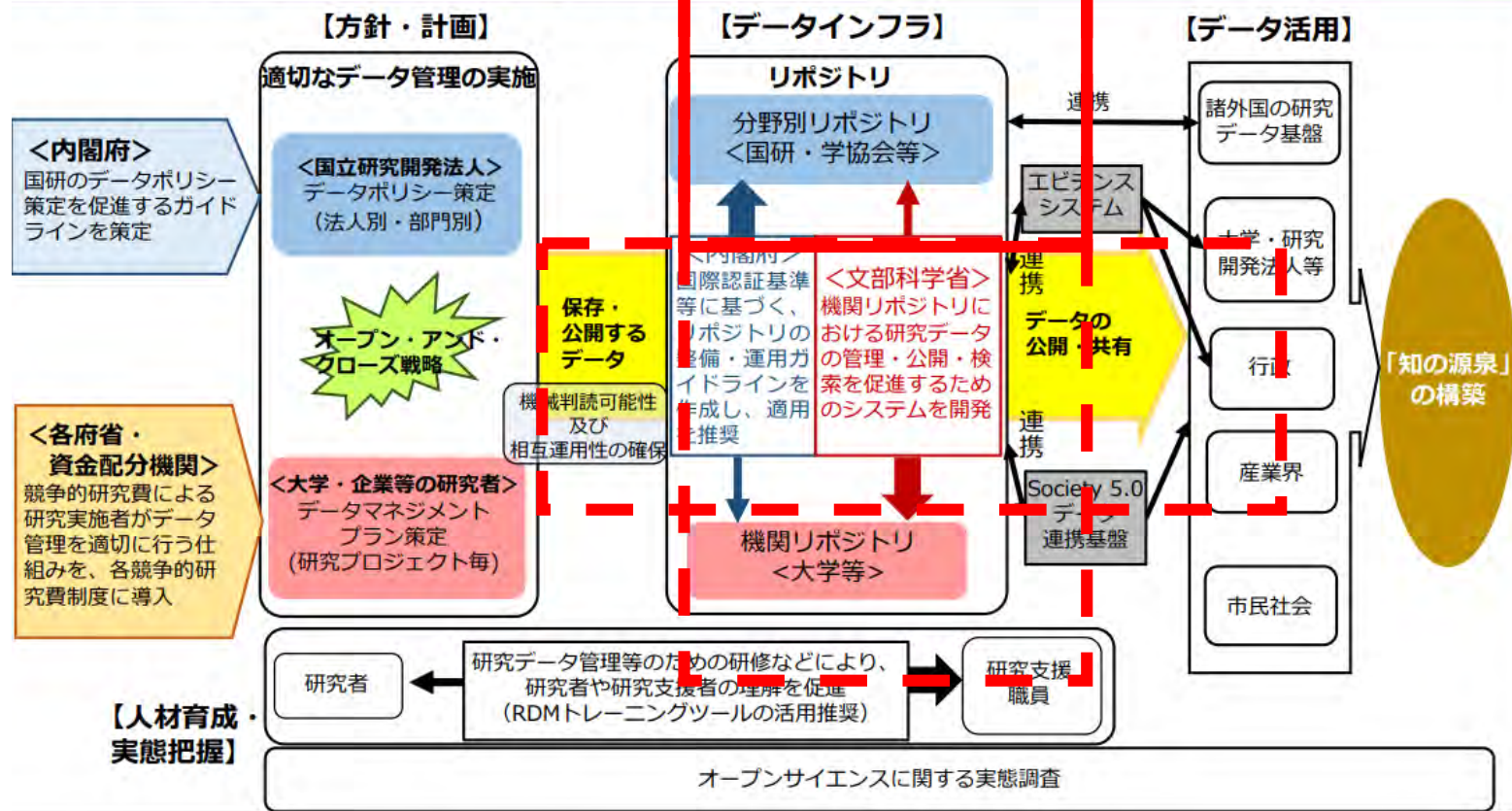
分野ごとにリポジトリの運営状況・国際連携は異なる

# 統合イノベーション戦略

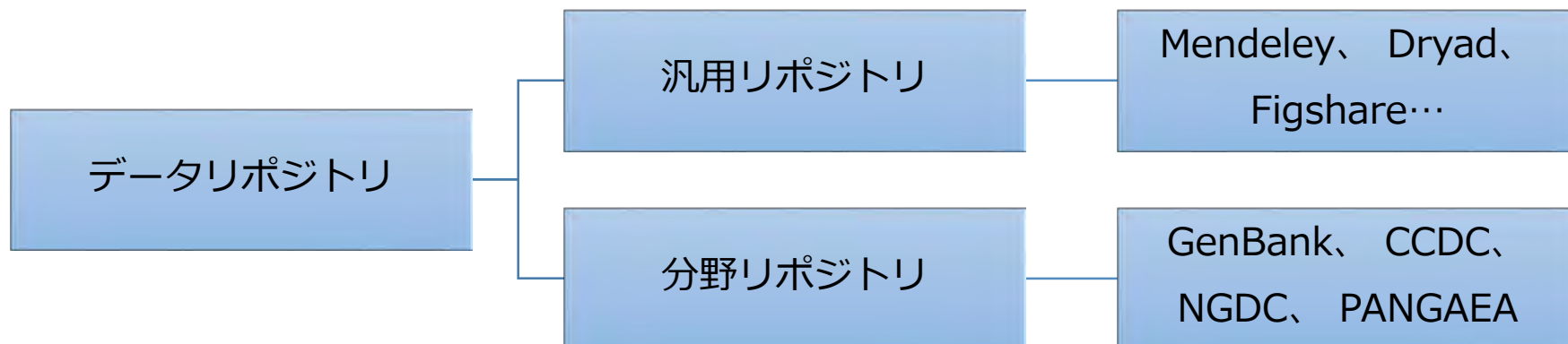
## オープンサイエンスのためのデータ基盤の整備

資料1: 国際的動向を踏まえたオープンサイエンスの推進に関する検討会(第6回)

- 国益や研究分野の特性等を踏まえて、オープン・アンド・クローズ戦略を考慮し、サイバー空間上での研究データの保存・管理に取り組み、諸外国の研究データ基盤とも連携して巨大な「知の源泉」を構築し、あらゆる者が研究成果を幅広く活用
- その結果、所属機関、専門分野、国境を越えた新たな協働による知の創出が加速



## 国際動向



その他の例：Research Data Allianceでのデータリポジトリ検討部会設置状況：

- Research Data Repository Interoperability WG
- Repository Platforms for Research Data IG
- RDA/WDS Certification of Digital Repositories IG
- Domain Repositories Interest Group

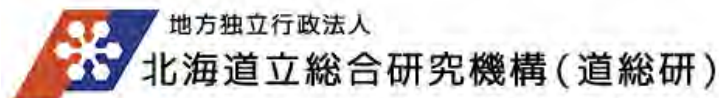


# リポジトリ小委員会とは？

## Japan Data Repository Network (JDARN)

## 委員の参加機関

### 参加者：18名





# 研究データ利活用に関する活動



## 国際

研究データ同盟：  
RESEARCH DATA  
ALLIANCE (RDA)(2013-)

- ◆ 7,400名登録
- ◆ 94ワーキンググループ
- ◆ 実質的な国際標準・国際相互結合体制の形成を目指す

## 国内

研究データ利活用協議会  
(RDUF)委員会 (2017-)

### 小委員会

- ◆ **リポジトリ**
- ◆ ライセンス
- ◆ 研究データ管理

リポジトリ小委員会は月1～2  
回開催、現在まで13回実施

## JDARNの目的

- データリポジトリコミュニティを作ること
- よりよい研究データ管理と運営がわかる専門家集団を作ること
- データリポジトリへの要求の多様化に対し、共通の課題を議論すること
- 国内のデータリポジトリの信頼性を、国際的に期待される水準に高めること

# データリポジトリの信頼性

# リポジトリの信頼性

信頼性

コンテンツ

データ品質

コンテナ

組織・人材・情報システムなど

リポジトリの**コンテナ**としての信頼性に焦点を当てる

安定した運営

持続可能性

堅牢なデータ管理

高い専門性

高度なキュレーション（収集・分析・再構成）

明確なライセンス・ポリシー

データ利用の実績

# 信頼性の段階

- 1. Self-assessment:** 自分で自分を評価。  
「査読なし」の状態。
- 2. Certification:** 第三者が公開情報を基に評価。「査読を通った」状態。
- 3. Registry:** re3data等、情報集約を担うプレプリントサーバ的なサービスも登場。
- 査読の有無に関わらず、**信頼性に関わる情報の公開**自体は必要なこと。

Japan Open Science Summit 2018: 研究データ管理を考える～データリポジトリのサービスとCoreTrustSeal認証～（研究データ利活用協議会リポジトリ小委員会）  
doi:10.11502/joss2018\_c5\_01より引用

# リポジトリの信頼性評価

## 1. 保存重視 : CoreTrustSeal(CTS)

組織や予算の持続性など、データ保存の側面を重視。

## 2. 利用重視 : GeoLabel、 Enabling FAIR Data Project

データ利用の基盤整備など、データ利用の側面を重視。

## 3. 出版重視 : 出版社の要求水準 Fair Data Principles

より軽い傾向だが、査読対応などの特定要求あり。

## 4. JDARNが目指すガイドライン

自分たちのリポジトリのための評価・レベルを見極める。

# データリポジトリの要求 要件に関する考察、検討

# データリポジトリの要求要件の例（その1）

- 組織的な基盤（Organizational Infrastructure）
  - R1. ミッション／視野（Mission/Scope）  
研究領域におけるリポジトリのミッションを明らかにするもの
  - R2. ライセンス（Licenses）  
データ提供者との合意事項、データ利用者に対するライセンス等
  - R3. アクセスの継続性（Continuity of access）  
保有するデータへのアクセスとデータの保存を保証する計画等
  - R4. 機密保持／倫理（Confidentiality/Ethics）  
システムのセキュリティ（特に機密保持）、スタッフへの倫理教育等
  - R5. 組織的な基盤（Organizational infrastructure）  
ファンディングとガバナンス
  - R6. 専門家によるガイダンス（Expert guidance）  
専門家の指導とフィードバック

参考：<https://www.coretrustseal.org/>



# データリポジトリの要求要件の例（その2）

## ■ デジタルオブジェクト管理（Digital Object Management）

- R7. データの完全性・信頼性（Data integrity and authenticity）  
データの変更管理
- R8. 評価（Appraisal）  
データの受け入れ管理
- R9. 保管手続きの文書化（Documented storage procedures）  
データの保存ポリシーやバックアップ
- R10. 保存計画（Preservation plan）
- R11. データの品質（Data quality）
- R12. ワークフロー（Workflows）  
データの受け入れから流通までのワークフロー
- R13. データの発見と識別（Data discovery and identification）  
データの検索方法と永続的な識別子
- R14. データの再利用（Data reuse）  
メタデータ

参考：<https://www.coretrustseal.org/>

# データリポジトリの要求要件の例（その3）

## ■技術（Technology）

- R15.技術的な基盤（Technical infrastructure）  
リポジトリのインフラ（ハードウェア・ソフトウェア）
- R16.セキュリティ（Security）  
特にシステム障害への耐性

参考：<https://www.coretrustseal.org/>

## データリポジトリの要求要件の 課題 = 理解しづらさ

1. 抽象的な表現が多いため、具体的なアクションがイメージしづらい。
2. 長期的な方針のもと人員配置し運営することが前提で、日本の現状と異なる。



要求要件の内容を分解し、アクション可能な項目ごとに整理する。

Japan Open Science Summit 2018:信頼性を高める為に必要なアイテム  
八塚茂 (バイオサイエンスデータベースセンター)  
doi:10.11502/joss2018\_c5\_03より引用

# 抽出したアイテム

(八塚委員試案をもとに委員会で議論)

## 組織

- 安定した運営組織
- 運営組織内でのフォロー委員会または外部諮問委員会

## 運営に必要な役割 (事例として業務分担者として表現した)

- 責任者・運営委員会
- 企画担当者
- 調達担当者
- ポリシー作成者
- 外部アドバイザー (分野専門家他)
- 外部アドバイザー (法律専門家他)
- 契約担当者
- データ開示管理者
- 個人情報管理者
- データ受け入れ担当者
- データ品質管理者
- メタデータ管理者
- データキュレーター
- データ提供担当者
- リポジトリ利用解説者
- 設備管理者
- システム担当者
- ハードウェア管理者
- ネットワーク管理者
- セキュリティ担当者
- スタッフ教育担当者
- 広報担当者

# 抽出したアイテム

(八塚委員試案をもとに委員会で議論)

## 文書・ポリシー・マニュアル類 (例)

- ① リポジトリのミッションや組織での承認レベルを記載した文書またはWebサイト
- ② リポジトリの中期・長期運営計画
- ③ リポジトリの終了計画
- ④ リポジトリシステム仕様 (システムマニュアル)
- ⑤ データ保存ポリシー
- ⑥ データ開示ポリシー
- ⑦ 個人情報取り扱いポリシー
- ⑧ データコレクションポリシー (品質、メタデータ、ファイルフォーマット等)
- ⑨ 調達契約
- ⑩ データ提供者との契約文書
- ⑪ データ利用者との契約文書
- ⑫ アドバイザリー報告書
- ⑬ データ管理マニュアル
- ⑭ データキュレーションマニュアル
- ⑮ 障害復旧マニュアル (セキュリティ、災害対応、データ復旧)

# 役割と 文書・ポリシー・マニュアル類

文書等番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
責任者・運営委員会	■	■	■	■	■	■	■	■							
企画担当者	■	■	■												
調達担当者				■					■						
ポリシー作成者					■	■	■	■					■		
外部アドバイザー（分野 専門家他）												■			
外部アドバイザー（法律 専門家他）															
契約担当者									■	■	■				
データ開示管理者						■									
個人情報管理者							■								
データ受入担当者													■	■	
データ品質管理者								■						■	
メタデータ管理者								■							

# 役割と 文書・ポリシー・マニュアル類

文書等番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	
データキュレーター																
データ提供担当者																
リポジトリ利用解説者																
設備管理者																
システム担当者																
ハードウェア管理者																
ネットワーク管理者																
セキュリティ担当者																
スタッフ教育担当者																
広報担当者																

# 抽出したアイテム

(八塚委員試案をもとに委員会で今後議論する予定)

## 情報インフラ（ハードウェアを含む）

- リポジトリのインフラ
- データ管理システム
- データバックアップシステム
- データ検索システム

## 教育

- スタッフへの教育（ポリシー、データ保護）
- スタッフへの教育（ITスキル、データのキュレーション）

## その他

- 永続的な識別子(PID)



# アイテム整理による知見

- 「アイテム」ごとに整理することで、リポジトリ運営の改善に向けたヒントが得られる
- 「アイテム」の整備状況を公開することが、データリポジトリへの信頼感につながる
  - データを預けるとき判断材料となる
  - リポジトリ運営担当者の仕事見える
- 「アイテム」の再構成が、データリポジトリの新たなガイドラインに寄与する可能性がある
  - 小委員会でのアドバイスや事例紹介
  - データリポジトリ設立・運営ノウハウ提供

# 委員会活動のふりかえりと 今後

# アイテムからガイドラインへ

(小委員会で今後想定される議論)

- 役割と文書 → 運営しているリポジトリに不足している文書類がわかり、今後関係者と共に運用内容を文書化するきっかけとなる
- 人材 → 役割ごとに必要なスキルを整理し、現状のメンバーでの兼任可能性、転任可能性が見える
- アイテムごとの必須度 → リポジトリに不足しているアイテムについて優先順位がつけられる
- アイテムが整備・更新される → 国際的な認証要件を満たすことになる

# 委員会活動を通して データリポジトリコミュニティ構築

- リポジトリの運営にたずさわる共通課題を共有
  - リポジトリの持続可能性（運営資金、設立母体）
  - ひとりで多数の役割を兼務
  - リポジトリメンバーのコミュニケーション
- 国内のデータリポジトリの認知・利用拡大
  - 研究者はデポジットしやすいリポジトリを利用する
  - 学会運営データリポジトリ、分野別リポジトリ、機関リポジトリに拡大
  - 国内で失われつつあるデータを救出する
  - 世界のデータが日本のリポジトリにあることをめざす
- 国際的なデータリポジトリ動向の共有
  - 論文誌が推奨するデータリポジトリ
  - 認証による宣伝効果
  - 国内リポジトリからの情報発信

**ご清聴ありがとうございました**