

第 3 回 SPARC Japan セミナー2020

「初めての研究データ」

「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」と 「国内機関における研究データ管理の取り組み状況調査」

安原 通代

(国立情報学研究所)

講演要旨



2020 年 10 月にオープンアクセスリポジトリ連合 (COAR) が "COAR Community Framework for Good Practices in Repositories" を公開しました。JPCOAR では、このフレームワークの日本語訳を作成し、2020 年 12 月に JPCOAR サイトで公開しました。本報告ではこのフレームワークの概要説明及び、2020 年に JPCOAR 研究データ作業部会 RDM 事例形成プロジェクトと AXIES-RDM 部会が協力し行った「国内大学・研究機関における RDM 取り組み状況調査」について紹介します。



安原 通代

2020年度SPARC Japanセミナー企画ワーキングメンバー。京都大学図書系職員。2020年4月から国立情報学研究所学術基盤推進部図書館連携・協力室オープンアクセスリポジトリ推進協会 (JPCOAR) 事務局担当。JPCOAR研究データ作業部会会員。

本日の内容

本日は、前半では昨年 10 月にオープンアクセスリポジトリ連合 (COAR) が公開した「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」についてお話しします。セミナーのテーマは「初めての研究データ」ですが、この COAR コミュニティフレームワークは、データに関する項目もありますが、研究データリポジトリのためだけのフレームワークではないので、その点をご了承ください。後半では、2020 年 11 月から 12 月にかけて AXIES-RDM 部会と JPCOAR 研究データ作業部会が連携して実施した「国内機関における研究データ管理の取り組み状況調査」について説明します。

JPCOAR とは

オープンアクセスリポジトリ推進協会 (JPCOAR)

は、2016 年 7 月に、リポジトリを通じた知の発信システムの構築を推進し、リポジトリコミュニティの強化と日本のオープンアクセスならびにオープンサイエンスに資することを目的として設立されました (図 1)。国公立大学図書館協力委員会と国立情報学研究所 (NII) との間の協定書に基づき設置された「大学

JPCOAR とは

- ・ オープンアクセスリポジトリ推進協会
 - JPCOAR : Japan Consortium for Open Access Repository
 - <https://jpcoar.repo.nii.ac.jp/>
- ・ 2016年7月 リポジトリを通じた知の発信システムの構築を推進し、リポジトリコミュニティの強化と、我が国のオープンアクセス並びにオープンサイエンスに資することを目的として設立
- ・ 国公立大学図書館協力委員会と国立情報学研究所 (NII) との間の協定書に基づき設置された「大学図書館と国立情報学研究所との連携・協力推進会議」と連携し、活動を行っている
- ・ JAIRO Cloud (共用リポジトリサービス) を NII と共同運営
- ・ 会員機関数 : 662機関 (2021年1月8日現在)

(図 1)

図書館と国立情報学研究所との連携・協力推進会議」と連携して活動しています。共用リポジトリサービスである JAIRO Cloud を NII と共同運営しており、2021 年 1 月 8 日現在、JPCOAR の会員機関数は 662 機関です。

JPCOAR の運営組織体制は図 2 のようになっています。参加機関の意思を直接反映させる場が総会です。運営委員会は、運営に関する基本方針を策定する役割を担っています。その運営委員会の下にあり、研修、広報など各種事業の具体的な活動の中心となる組織が作業部会です。作業部会は、JPCOAR の参加機関に所属する職員で構成されています。「国内機関における研究データ管理の取り組み状況調査」は、このうちの一つである研究データ作業部会が行ったプロジェクトの一つです。

事務局は、JPCOAR の活動に必要な業務を遂行する部署です。事務局の専任職員は参加機関からの出向職員となっています。私は京都大学からの出向職員としてここに所属しています。

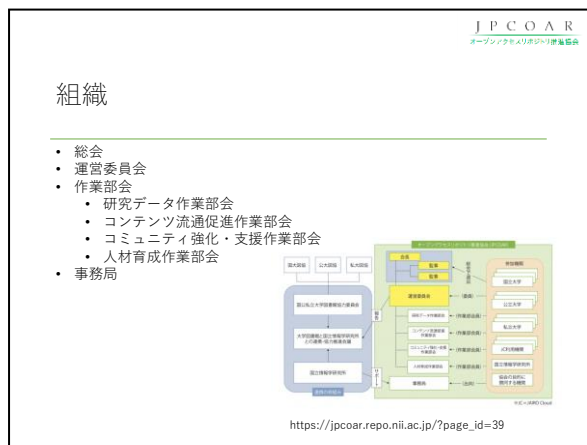
COAR とは

次に、オープンアクセスリポジトリ連合 (COAR) という組織について簡単に説明します。COAR は、図書館、大学、研究機関など世界中の 157 のメンバーとパートナーから成る国際協会で、JPCOAR もメンバーとして加入しています。国際的なリポジトリの協力体制による学術研究成果の可視化と利活用の促進を目的

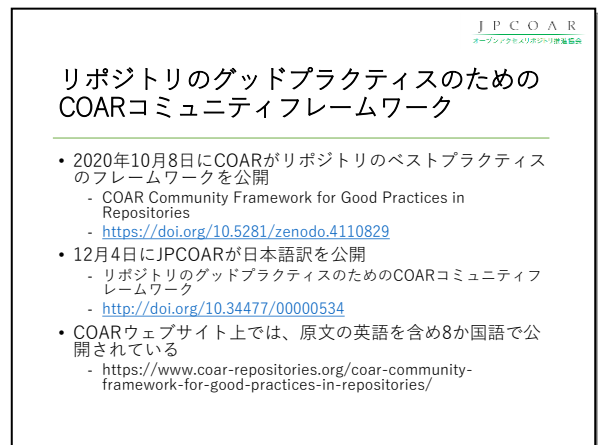
としていて、世界中のリポジトリコミュニティの情報共有を中心に、コミュニティ間をつなぐ役割を果たす国際組織です。COAR がリポジトリの運用の評価、改善を支援するために策定したのが「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」です (図 3)。これは 2020 年 10 月 8 日に COAR の Web サイトで公開されました。日本語訳は JPCOAR が作成し、同年 12 月 4 日に JPCOAR の Web サイトで公開しています。この日本語訳は JPCOAR の機関リポジトリからダウンロードでき、原文と同様に CC BY で利用可能になっています。COAR の Web サイト上では現在、原文の英語を含めて 8 カ国語で公開されています。

このフレームワークは、リポジトリの運用の評価、改善を支援することを目的として策定されています。他にも多くのリポジトリの評価フレームワークが存在していますが、それらの評価基準は異なる組織に散在しており、特定の地域や特定のタイプのリポジトリしか利用できないことが多いです。COAR フレームの目的は、そういった既存のフレームワークの基準をまとめて、さまざまな地域、さまざまなタイプのリポジトリで利用できるようにすることだそうです。

COAR の現事務局長である Kathleen Shearer 氏が回答しているインタビューの記事に、フレームワークの目的や今後の作業についてももう少し詳細に書かれていたので紹介させていただきます (図 4)。既存のフレームワークの多くは、求める基準、リポジトリ要件を



(図 2)



(図 3)

非常に高く設定していて、世界中の多くのリポジトリ、特にあまりリソースのないリポジトリはその基準に到達できないので、そういったフレームワークは無関係なものになってしまっている。そのため、COAR のフレームワークは、多くのリポジトリが採用できるように、基準を高く設定し過ぎないようにしたとのこと。また、COAR のフレームワークは、正式な認証を実施するために作成されたものではなく、リポジトリの改善を支援する自己評価のためのツールとされています。一方で、このフレームワーク作成の次のステップとして、より詳細なガイダンスをリポジトリの管理者に提供できるための作業が既に開始されているとのこと。

この記事は 2020 年 10 月 30 日に公開されたもので、インタビューの時点では 11 月から COAR のワーキンググループがそのような活動を開始すると事務局長が答えているのですが、現時点で COAR から次のステップに関する情報はまだ出ていません。

作成プロセス

次に、COAR のフレームワークがどのようなプロセスに基づいて作成されたかを紹介します (図 5)。まず、既に存在する多くのリポジトリの評価フレームワークを COAR のワーキンググループがレビューしてギャップを特定し、その重要性、妥当性、実施可能性のレベルを評価して、各特性を「必須」または「望ましい」に分類して草案を作成しました。2020 年 6～8

月にこの草案が COAR メンバーに配布され、実施が困難なものはないか、重要な基準が欠けていないか、解釈が難しいものはないか、不適切なものはないかという観点からさらにフィードバックを受け、それを反映して公開に至ったということです。

今後は、コミュニティがこれらのベストプラクティスを採用する際に役立つ事例やガイド、インストラクションへのリンクや引用を COAR のホームページで提供することを予定しているそうです。また、このフレームワークは完成形ではなく、毎年 1 回、7 月から 8 月にかけてフレームワークを見直し、最新のベストプラクティスを反映して更新することを予定しているそうです。

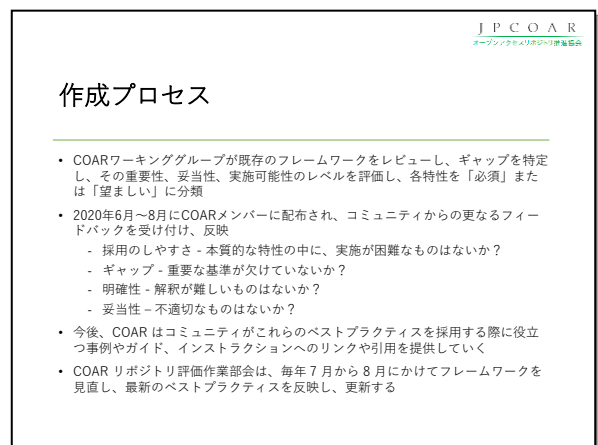
フレームワークで取り上げる特性の項目

フレームワークの内容について、ごく簡単に説明します。図 6 にあるような九つの項目に分類され、各項目に必須の特性と望ましい特性が挙げられています。実際に読んでいただくと、明確なチェック項目を定めているわけではないことがお分かりいただけると思います。

一つ目の項目では、リポジトリのリソースの発見可能性を高めるために必要とされる特性が取り上げられています。リポジトリ内のリソースが利用できなくなった場合でもメタデータが利用可能か、検索機能を提供しているかなど、必須とされる特性と、望ましいとされる特性に分けられています。



(図 4)



(図 5)

二つ目は、リソースへのアクセスを保証するための項目です。必須とされる特性として、ユーザーがリポジトリ内のリソースにアクセスするためのコストは発生しないこと。リポジトリは、公示された期間におけるリソースへの継続的なアクセスを保証すること。リポジトリは、障害者のために、文書やメタデータへのアクセスをサポートすること。デバイスの中立性として、ユーザーがリポジトリにアクセスするために特定のデバイスを必要としないこと。こういった四つの特性が挙げられています。具体的な期間や、サポートのための具体的な整備内容については規定されていません。期間が定められて公示されているか、何らかのサポートがされているかといった確認内容になっています。アクセスに関する望ましい特性としては、大容量ファイルの取り扱いや、アクセス制限のあるリソースへの間接的アクセスのサポートがあるかなどが挙げら

れています。

三つ目は、リソースの再利用についての条件や、フォーマットに関する項目です。必須の特性が四つと、望ましい特性が二つ挙げられています（図7、8）。

四つ目は完全性と真正性です。私は最初、この項目名をただけでは内容が分かりませんでした。原文では integrity and authenticity となっています。リポジトリのリソースが不正に変更されていないことを保証するための文書やポリシー、記録の保持に関する内容となっています。バージョンの管理については必須事項とされています。

五つ目は品質保証に関する項目です。メタデータやリソースの品質を保証する項目が図9のように挙げられています。

六つ目はセンシティブデータのプライバシーに関する項目です。特に研究データに関わる項目になるかと

J P C O A R
オープンアクセスリポジトリ推進協会

フレームワークで取り上げる特性の項目

- ・ 発見可能性
- ・ アクセス
- ・ 再利用
- ・ 完全性と真正性
- ・ 品質保証
- ・ センシティブデータのプライバシー（例：被験者情報など）
- ・ 保存
- ・ 持続可能性とガバナンス
- ・ その他

(図6)

J P C O A R
オープンアクセスリポジトリ推進協会

再利用（望ましい特性）

3.5 リポジトリは、フルテキストハーベスティングやテキストマイニング・データマイニングをサポートするためのオープンなAPIを持つ

3.6 リソースは、機械可読なコミュニティ標準フォーマットで保存される

(図8)

J P C O A R
オープンアクセスリポジトリ推進協会

再利用（必須の特性）

3.1 リポジトリは、再利用条件を規定するライセンス情報をメタデータレコードに含む

3.2 リポジトリは、引用可能なPIDを提供する（1.4参照）

3.3 リポジトリは、推奨される、独自仕様でないフォーマットのリストを提供する

3.4 ランディングページには、引用に必要な情報を含むアイテムに関するメタデータが機械及び人間が読める形式で掲載される

(図7)

J P C O A R
オープンアクセスリポジトリ推進協会

品質保証

- ・ 必須の特性

5.1 リポジトリはメタデータ（及び該当する場合はリソース）の基本的なキュレーションを行う

5.2 リポジトリは、どのようなキュレーションプロセスがリソースとメタデータに適用されているかを概説した文書またはポリシーを提供する
- ・ 望ましい特性

5.3 リポジトリは、リソースとメタデータの外部注釈、コメント、レビューをサポートする

(図9)

と思いますが、リポジトリがセンシティブデータを収集している場合には、データ所有者が許可されたユーザーのみにアクセスを制限できる仕組みを提供していることが必須の特性とされています。望ましい特性になると、さらにデータのセキュリティ要件のレベルの違いに基づいて、段階的なアクセスを提供できることが加わります。

七つ目は保存に関する項目です。

八つ目に挙げられているのが、リポジトリの管理責任を持つ組織の明確化や、管理組織における運用体制の明確化についての項目です。この項目のみ、望ましいとされる特性はなく、全て必須とされる特性になっています。

最後の項目はその他です。支援窓口やヘルプデスクがあるかや、利用統計についてなど、他の八つの項目には含まれていない特性が挙げられています。

フレームワークで取り上げられている項目は以上です。必須となっている特性でも、ポリシー作成が必要だったり、機能面ではシステム構築から必要になったりして、すぐには対応が難しい内容も多く含まれています。最初にも申し上げましたが、COAR のフレームワークは、正式な認証を実施するために作成されたものではなく、リポジトリの改善を支援する自己評価のためのツールとされています。グッドプラクティスに基づいて自機関のリポジトリを自己評価して改善点を見つけ、さらに将来に向けて機能やポリシー、運用について検討するために活用するものと考えています。

JPCOAR の関係者から、このフレームワークを基にして日本国内に向けたチェックリストを作ってはどうかという案が出されており、現在、ミーティングや意見交換を行っています。また、COAR のフレームワークも毎年更新が予定されているので、その進展についても今後報告していきたいと思っています。

「国内機関における研究データ管理の取り組み状況調査」

次に、2020 年末に実施した「国内機関における研

究データ管理の取り組み状況調査」について紹介します（図 10）。この調査は、JPCOAR の作業部会の一つである研究データ作業部会が AXIES-RDM 部会と連携して行っている RDM 事例形成プロジェクトの一環になります。RDM 事例形成プロジェクトは、参加機関による RDM の取り組みを支援し、データポリシー策定の足掛かりを作成することや、ディスカッションやワークショップを通じて事例集を作成・公開することなどを目的としています。プロジェクト全体については、2020 年の NII オープンフォーラムで、NII オープンサイエンス基盤研究センター（RCOS）の南山泰之先生が報告されていますので、ご参照ください。

今回行った「国内機関における研究データ管理の取り組み状況調査」は、国内の大学・研究機関を対象に、機関内での研究データ管理の取り組み状況を把握することを目的として、JPCOAR 会員機関、AXIES 会員機関、その他の国内機関にご協力いただいたアンケート調査です（図 11）。質問は全部で 46 問あり、七つのセクションに分かれています。それぞれのセクションについて、内容に応じて適切と想定される回答部署を記述していますが、想定回答先は各大学・研究機関の体制によって異なると思いますので、あくまでも想定です（図 12）。

この調査を行うに当たっての AXIES-RDM 部会と連携体制は図 13 のような形になっています。こちらの図も 2020 年末の AXIES 年次大会で RCOS の南山先生が講演で使われたものを使用させていただいています。

J P C O A R
オープンサイエンス基盤研究センター

**RDM事例形成プロジェクト
(AXIES-RDM部会との連携事業)**

- 参加機関によるRDMの取り組みを支援し、データポリシー策定の足掛かりを作成する
- ディスカッションやワークショップを通じて事例集を作成・公開する
- AXIES-RDM部会とJPCOAR研究データ作業部会が連携して実施する

2020 NIIオープンフォーラム
AXIES研究データマネジメント部会合同トラック「学術機関における研究データ管理フレームワーク」
https://www.nii.ac.jp/openforum/2020/day2_cs4.html
研究データ管理のステークホルダーと事例収集 南山泰之（国立情報学研究所）
<https://www.nii.ac.jp/openforum/upload/cb0bfc6d76ac96dde6c1340e2f86e6e68b4c155.pdf>

(図 10)

調査開始時（11 月）の JPCOAR の会員機関数は 657 機関、AXIES の会員機関数は 133 機関、両方に重複して加盟している機関が 112 機関でした。また、どちらにも加盟していない国内機関にもご連絡して、最終的には合計 658 機関にご協力をお願いしました。

Google Form 経由での回答をメールで依頼して、Google Form にアクセスできない機関については、Word ファイルに回答を入力の上、メール添付での送信をお願いし、最終的に 352 の機関から回答を頂くことができました（図 14）。内訳としては、研究開発機関が 62%、国立大学が約 19%、公立大学が 12%、私立大学が 2%、大学共同利用機関等が 2%です。もう少し内容を報告できたらよかったのですが、現在、報告書は取りまとめ中になっています。

プロジェクト自体は 2019 年度から 3 年を予定しており、本年度が 2 年目です。今年度はアンケート調査

自体に対する報告書をまとめて、来年度は、調査結果やプロジェクトの成果報告などをまとめて公開します。

その他の研究データに関する取り組み

最後に、JPCOAR の作業部会による活動の中から、研究データに関するその他の取り組みを幾つか紹介します。

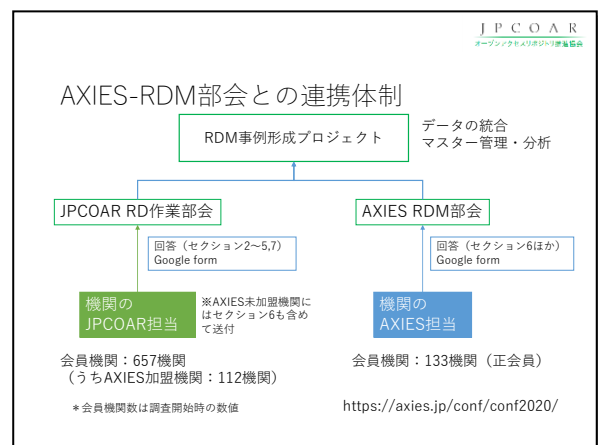
2020 年 10 月に教材「研究者のための研究データマネジメント」を公開しています。研究支援者の目線から、大学や研究機関等に所属する研究者の方に向けて作成した教材集です。研究データ管理の場面に応じた 12 のテーマに分かれていて、研究者自身がこの教材によって必要な知識を得る他に、研究を支援する側が各教育機関の研究環境やニーズに応じた形で、この教材を加工して研究データ管理サービスを提供することを想定しています。また、これは 2017 年公開の研究

J P C O A R
オープンアクセスリポジトリ推進協会

国内機関における研究データ管理の取り組み状況調査

- 目的：国内の大学・研究機関を対象に、機関内での研究データ管理の取り組み状況を把握する
- 調査期間：2020年11月27日～12月28日
- 質問数／所要時間：全46問／約50分
- 対象：JPCOAR会員機関、AXIES会員機関、その他国内機関（計685機関）

(図 11)



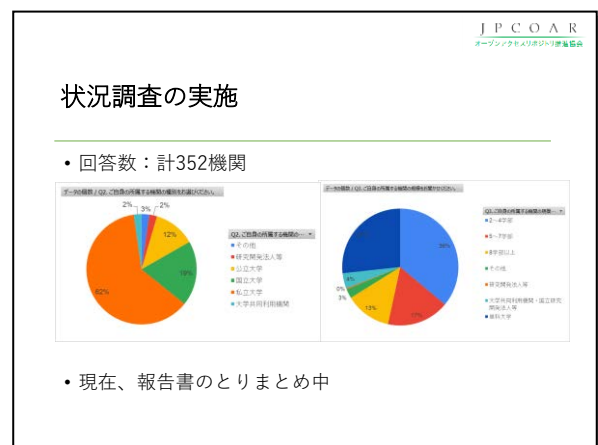
(図 13)

J P C O A R
オープンアクセスリポジトリ推進協会

セクション別回答想定先

- セクション1 (概要説明)：(全1問)
- セクション2 (基礎情報)：JPCOAR参加機関 窓口担当部署または機関リポジトリ担当部署 (全3問)
- セクション3 (ニーズの把握)：機関内の研究データ管理 (RDM) 担当者／対応部署 (全5問)
- セクション4 (データ管理体制の構築状況)：研究推進部署／URA等 (全6問)
- セクション5 (研究データ管理サービスの実施状況)：大学図書館等 (全11問)
- セクション6 (情報インフラの整備状況)：情報基盤センター等 (全18問)
- セクション7 (オープンアクセスリポジトリ推進協会 (JPCOAR) について)：JPCOAR参加機関 窓口担当部署または機関リポジトリ担当部署 (全2問)

(図 12)



(図 14)

支援者向けの教材「RDM トレーニングツール」、2018 年公開の教材「研究データ管理サービスの設計と実践」に続く第 3 弾の教材となっています。2018 年公開の教材「研究データ管理サービスの設計と実践」については、2 月 10 日に第 2 版が公開されています。こちらは大学図書館員や URA、技術スタッフなど、研究支援職員が研究者の研究プロセスに沿って研究データ管理サービスの設計と実践の方法を身に付けることを目的としたものです。

そして、3 月にはデータベースレスキューマニュアルを公開する予定となっています。

●八塚 質問を頂いています。「COAR のフレームワークの JPCOAR 版という話がありましたが、研究データ利活用協議会（RDUF）が昨年策定したガイドラインとの違いや位置付けなどについて教えていただくと幸いです」。私が推測するには、これは RDUF の中のジャパン・データリポジトリ・ネットワーク（JDARN）で作成したリポジトリガイドラインのことだと思いますが、その辺の違いで、もしご存じのことがあれば教えていただけますか。

●安原 すみません、私の方で明確な違いは分かりません。

●八塚 違いについてはこれから検討しようとしているところだと思いますので、なかなか見えてこないと思いますが、私の感覚では、COAR のフレームワークは CoreTrustSeal などと違って、誰かに認証されるような類いではなく、あくまでもリポジトリ側で自己改善していくためのツールという位置付けだと思います。そういうことでよろしいでしょうか。

●安原 はい。そのとおりです。

●八塚 ありがとうございます。では、次の質問にしたいと思います。「調査状況については報告書を取りまとめ中ということでしたが、もし興味深い結果があれば、何か情報を頂けませんか」。

●安原 「研究データの管理サービスを展開するに当たって、自機関で、研究者から要望が高いと思われるサービスをお聞かせください」という質問があり、複数選択可として、こちらからデータ管理計画（DMP）の作成支援や研究データのストレージ提供など幾つかの選択肢を挙げたのですが、最も回答数が多かったのは「分からない」でした。その上で、「ここで選択した研究データ管理に関する支援を実施する際に必要と思われる知識やスキルはどういったものでしょうか」という質問をして、データキュレーションスキルや ICT スキルといったスキルを幾つか挙げて、こちらも複数回答可としたのですが、ほぼ全てのスキルを「はい」とした方が多かったです。ですので、どういった要望が高いかは分からないけれども、あらゆるスキルや知識が必要とされているという認識があるのだらうと思いました。

このアンケートを行った際に、セクションに分けて、取りまとめを依頼している部署から各関連部署に質問していただいたのですが、その際に非常に思ったのが、研究データ管理における関連部署の連携の重要性です。必要とされるものは分からないけれども、必要とされる能力は多岐にわたっているので、これは本当に一つの部署では対応できるものではありません。そうすると、学内・機関内のさまざまな部署で連携していくことが必要だと感じました。

●八塚 今の質問に関係するのですが、「結果については取りまとめ中ということですが、どういった設問だったのかをご紹介いただけないでしょうか」という要望が来ています。

●安原 図 12 のように、セクションに分かれており、

ニーズの把握やデータ管理体制の構築状況、研究データ管理サービスの実施状況、情報インフラの整備状況、最後は JPCOAR についての質問もさせていただきました。何らかの研究データ管理サービスを実施しているか、その中で図書館が実施しているものはあるかといった質問や、現在はしていないけれども図書館で実施を計画・検討しているものはあるか、また、過去に研究データ管理に関する研修やワークショップ、シンポジウム等のイベントを実施したことはあるかといった質問で、全部で 46 問あります。

●八塚 設問についても、まだ公開されていないというのでしょうか。

●安原 設問自体は機関の皆さまにお送りしただけで、まだ公開していません。

●八塚 私から一つ質問させてください。今回、COAR のフレームワークを翻訳されて、これが現状の日本の機関リポジトリや分野別リポジトリにどれだけマッチしていそうかという感触は何かありますか。

●安原 やはり日本のリポジトリに当てはめるには少し難しいものがあると感じまして、JPCOAR 内で、これを基にして日本に向けたチェックリストを作ろうということになっています。

●八塚 例えばどのあたりが食い違っていそうですか。

●安原 特定のことになってしまいますが、私は、障害者が文書やメタデータにアクセスするのをサポートするあたりで、少し具体的なリソース例やサポート対応を挙げていく必要があると思いました。