

Japan Data Repository Network (JDARN) : データリポジトリの信頼性向上を中心としたコミュニティ活動

Japan Data Repository Network (JDARN): Community-based activities for improving the trustworthiness of data repositories

*北本 朝展^{1,2,4}、絹谷 弘子³

*Asanobu Kitamoto^{1,2,4}, Hiroko Kinutani³

1. 国立情報学研究所、2. ROIS-DS人文学オープンデータ共同利用センター、3. 東京大学地球観測データ統融合連携研究機構、4. 総合研究大学院大学

1. National Institute of Informatics, 2. ROIS-DS Center for Open Data in the Humanities, 3. EDITORIA, The University of Tokyo, 4. SOKENDAI

1. JDARNの設立と経緯

Japan Data Repository Network (JDARN)とは、日本のデータリポジトリを対象として、世界の最新動向を共有しながら、信頼性を向上させるための取り組みを進める、コミュニティ活動である。その源流は、ジャパンリンクセンター (JaLC) が2014年10月から2015年9月にかけておこなった「研究データへのDOI登録実験プロジェクト」にある。このプロジェクトは、研究データに関する専門家が分野を越えて集まったという点では日本初とも言える画期的な場となった。この活動を引き継ぐものとして研究データ利活用協議会が2016年6月に設立され、その後いくつかの小委員会が提案されることとなった。その一つとして2017年10月に我々が立ち上げたのが「国内の分野リポジトリ関係者のネットワーク構築」小委員会である。2018年10月からは、より多くの分野と関係者を対象とするために「ジャパン・データリポジトリ・ネットワーク(Japan DAta Repository Network : JDARN)」と名称を変更した。

JDARNはデータリポジトリに関する動向を共有することが目的の一つであるが、中でも焦点となっているのがデータリポジトリの信頼性という問題である。研究データの生産者がデータを外部サービスに預ける際に、どこに預けるべきかを意思決定する基準として、データリポジトリの信頼性は重要な役割を果たす。そうした信頼性を示す基準の一つにCoreTrustSeal (CTS)がある。CTSはデータリポジトリに関する国際的な認証の一つであり、2019年2月現在で140あまりのデータリポジトリが認証を受けているが、日本ではまだ認証を受けているケースが少ない。CTSの認証が日本では少ない理由を探るため、2017年12月にセミナー「信頼できるデータリポジトリ～CoreTrustSeal認証に関する実践的情報共有の場～」を主催し、日本の有力なデータリポジトリがCTSの要求要件を用いて自己評価 (self-assessment) してみる試みを行った。その結果、CTSの背景となる考え方がわからないとCTSによる自己評価も難しいことが判明した。そこでまずCTSを理解するための資料の作成を開始し、これが小委員会の主要な活動となった。そしてさらに議論を重ねた結果、CTSありきでなく利活用の側面も考慮したデータリポジトリのガイドラインを作成する課題に活動がシフトしていった。

2. データリポジトリのガイドライン

現在作成中のデータリポジトリガイドラインは、基本的にCTSの要求要件 (16項目) を参考にしつつ、CTSを直訳するのではなくJDARNが独自に構成を提案するものである。このようにCTSを再考するきっかけとなったのが、バイオサイエンスデータベースセンター (NBDC) の八塚茂氏によるCTSのアイテム単位の整理である。CTSの審査過程では、様々なドキュメントを用意しそれを公開していることが透明性の一つの証拠となる。そこでCTSを実際に獲得したデータリポジトリの申請書を分析し、そこで言及されているドキュメントの種類を整理することで、CTSに必要なドキュメントを準備するという次のアクションがわかりやすくなると思った。CTSの抽象的な項目を具体的なヒト・モノなどに落とし込むことで、より理解しやすいガイドラ

インを作れる可能性が生まれたのである。

しかしドキュメントの整理に比べると、データリポジトリに関わる人に関する項目の整理はより困難である。データリポジトリではどんな職務が必要なのか、それを担うのは誰なのか。しかも職務については、その職務を専門家として何と呼ぶかという名称の問題もある。データ専門家として近年提唱される職名には、データライブラリアン、データキュレーター、データサイエンティスト、データエンジニアなどがあり、その意味も人によって異なる。これらの職務の概念を整理し、それらの長期的なキャリアパスを示すこと、それができればデータリポジトリを基盤としたオープンサイエンスの展開はおぼつかない。こうした問題についてはまだ確固たるモデルがあるとは言えず、我々は現在も議論を続けている。

3. 今後の展開

JDARNは設立以来、毎月1回ほどの会合を開きながら活発な議論を交わしてきた。そうした議論に参加するデータリポジトリの数が増えれば、日本のデータリポジトリの品質を高め、世界の中での存在感も高め、オープンサイエンスのための基盤としての価値も向上するであろう。そのためにはデータリポジトリが研究に不可欠な存在となる必要がある。データリポジトリというデータのコンテナとしての信頼性・持続性の向上がCTSの焦点であるが、それに加えてデータの統合、分析、可視化、社会実装などコンテンツの利活用に向けた多様な専門家も必要になる。これを単独で担える組織は限られるため、データリポジトリ間のコラボレーションも重要な課題であり、そこにデータリポジトリのネットワークが生きてくると考えている。

キーワード：ジャパン・データリポジトリ・ネットワーク、データリポジトリ、信頼性、CoreTrustSeal、ガイドライン、データ専門家

Keywords: Japan Data Repository Network, Data repository, Trustworthiness, CoreTrustSeal, Guideline, Data expert