

人文社会科学データの共有に向けての取り組みと課題

国立情報学研究所 朝岡誠

1. 問題の所在

現在、我が国の人文社会科学（SSH）データはデータを共有するインセンティブがないことから研究データが死蔵され、散逸することが危惧されている。（佐藤 2012 橋本 2015）一方、国外ではデータ共有のガイドラインをもとにデータの共有や利活用が急速に進展しており、このままではこの世界的な流れから日本が取り残され、国際的な共同研究を行うことができなくなる恐れがある。そこで日本学術振興会は研究者が共にデータを共有・利用しあう文化を醸成し、データ分析に基づく SSH 研究の飛躍的発展を目的とした人文学・社会科学データインフラストラクチャー事業を新設し、SSH データを共有する基盤の構築を進めている。事業はデータ共有機能を担う拠点機関（一橋大学経済研究所、東京大学社会科学研究所、慶應義塾大学パネル設計・解析センター、大阪商業大学 JGSS 研究センター）およびデータ利活用促進機能を担う中核機関（日本学術振興会、国立情報学研究所）から構成される。各拠点機関は所有するデータやデータ共有のためのインフラを整備し、中核機関は各拠点機関のデータを横断的に検索可能にし、研究者による利活用を促進することが求められている。中核機関は現在、分野横断的なデータカタログを設計しているが、データカタログを運用するためには、カタログに登録するデータの特性を記述するデータ（メタデータ）が必要になる。そこで各拠点機関のデータの整備状況を調査し、本データカタログで用いるメタデータの形式を検討する。そして、将来的に様々な研究機関が本データカタログに自分たちの所有データを登録して SSH データの共有を行う際、どのような課題があるのかを検討するために、各拠点機関がどのようにデータ提供を行っているのかを調査した。

2. メタデータの検討

拠点機関のデータの整備状況を調査したところ、各拠点でデータの性質が異なるため、データ公開の際に提示する情報に差があり、本カタログ用に共通のメタデータを整備する必要があることが判明した。現在、海外の研究機関との連携のためにも社会科学分野の国際標準メタデータ規格である DDI を利用することを念頭においているが、検索の利便性のために一部の項目について語彙統制を行うことを検討している。

3. 各拠点のデータ提供状況とデータ共有の課題

拠点機関のデータ提供状況を調査したところ、調査対象者への配慮から利用を学術目的に制限し、厳密なデータ管理を行っている拠点機関が複数あった。これらの拠点機関は利用者の所属や研究目的をもとにデータを提供するかどうかを判断し、利用者に対してデータの利用期間を設け、利用報告を義務として課しており、そのために利用者管理に根ざしたデータ管理システムを配備していることが判明した。SSH データを共有するためには、公開したデータに対して学術利用かどうかを簡易に判別し、提供するシステムと利用者からのデータ利用報告や研究成果報告を促す仕組みが必要であると考えられるが、現行のデータレポジトリの機能との兼ね合いから利用者管理がどこまで必要かを検討するのが今後の課題となる。

文献

佐藤博樹, 2012, 『実証研究におけるデータアーカイブの役割と課題－SSJ データアーカイブの活動実績を踏まえて』フォーラム現代社会学.

橋本雄太, 2015, 『人文学資料オープンデータの可能性と現状』情報の科学と技術.