

● 会議報告

KKJ セミナー「研究・実験データの保管・共有の推進方策」を受講して

中京大学 工学部 情報工学科 教授
長谷川 明生

8月21日(火)に紀尾井町の明治薬科大学剛堂会館で開かれた地域科学研究会 高等教育情報センター主催のセミナー「オープンサイエンスの情報基盤 - 研究・実験データの保管・共有の推進方策」に参加した。受講者は10数名であったが、名大や阪大教授、国公立大学や私学の研究支援課の役職者等の出席があり、今回の企画が時宜にかなったものであったことが感じられた。

セミナーのオーガナイザは国立情報学研究所オープンサイエンス基盤研究センターの船守美穂氏で、講演者は他に同じオープンサイエンス基盤研究センターの込山悠介氏、および京都大学情報環境機構IT企画室の青木学聡氏であった。

最初に船村氏によって、基調講演が行われた。

講演では、オープンサイエンスやオープンデータという概念が生まれてきた背景やオープンサイエンスの基本的な事項について提示された。オープンサイエンスの始まりは世界的な学術雑誌の寡占状態とそれによる購読料の高騰ということであった。1986年から2011年にかけて4倍に高騰した。ただし、この間、日本では円高であって、それでも倍額となっている。ドイツ大学学長会と出版社間で価格問題について交渉がなされていたが、交渉が決裂し、ドイツの大学では学術雑誌にアクセスできない状態が続いているということである。スウェーデン、ペルー、台湾もドイツと同様の状態とのことであった。

これらの動きに対してネットワークを介しての研究論文の共有の動きについて1994年に論文のインターネット上でのオープンアクセスの考え方が提示された。国家的な動きとして、「学術研究の費用は主に税金で賄われているのに、その成果を見るのに更にお金を払わなければいけないのは、納得がいかない！」という治療法を探している重病医療患者からの声があり、また研究助成機関による公的資金を得た研究の公開義務化の動きもある。ただし、オープンアクセスに対して、どのように質を担保するかAPC(論文掲載料)の処理について課題が残されている。

このような流れを受けて、論文レポジトリが用意されつつある状況である。その一方で、公表された論文の再現性に関する深刻な議論についても紹介された。自著、他著にかかわらず、50%近い論文について追試および再現ができないというレポートが示された。そのような問題点の解決策として、既存の論文レポジトリに、論文の根拠データ(論文中のグラフや表の元データ等)を追加することが行われつつある。この動きを進めて論文の根拠データのみではなく、研究のために統一的な視点で収集された基礎データ全体をアーカイブし積極的に公開していこうという動きがオープンデータである。データの公開について、徐々にではあるが拡大を続けており、科研費についてもDMP(Data Management Plan、データ管理計画)への言及が求められるようになった。

このような流れを受けて、国立情報学研究所で開発中のプロジェクト管理システムの概要について国立情報学研究所の込山氏より報告された。このシステムでは、プロジェクト進行管理のみならずプロジェクトで共有するデータを一元的に管理し、必要に応じて取得・保管したデータをオープンデータとして公開する機能を持っている。このシステムは2020年公開を目指して開発が進行している。

京都大学では、すでに情報メディアセンターの機種更新に合わせて自前で研究データ・アーカイブシステムを構築し運用中であり、そのシステム概要、および運用について、京大の青木氏より報告がなされた。その中で、膨大になりがちな研究データに対応するための、システムの容量設計、および保管期間の設定は根本解決の困難な問題であることが示された。

再度、船守氏により、まとめが行われ、パネル討論が実施された。

本セミナーは、「研究とその関連データ」をどのようにアーカイブし公開していくかをテーマとしたものであった。昨今、大学や研究機関においては論文をアーカイブおよび公開するための論文リポジブリの整備が行われている。このような論文リポジブに対して、発表された論文の根拠データの追加が要請されつつある。今後は、論文データ公開に加えて、研究のために系統的に収集したデータ全体のアーカイブと再利用可能な形式での公表が要請されそうである。システム的には論文データのアーカイブは、さほど困難ではないと想定されるが、基礎データのアーカイブと公表には記憶装置の容量が問題になりそうである。オープンデータの実装のためにはアーカイブ期間の設定等、ポリシーの検討と設定が各機関に求められている。