



開かれた学術情報の知識基盤整備

込山 悠介 KOMIYAMA Yusuke (コンテンツ科学研究系 助教)

専門分野：セマンティック Web、バイオインフォマティクス

■研究の成果を社会で広く共有する

研究論文だけでなく、その根拠となる研究データを広く公開し、共有しようという「オープンサイエンス」の動きが国際的な広がりを見せています。その推進のため、国立情報学研究所 (NII) では、全ての学術分野の研究データをまとめて管理し、検索できる情報基盤づくりを進めており、私はそのシステムの構築に取り組んでいます。

欧米では、税金で行われた研究の成果は、国民が自由にアクセスできるのが当然だと考えられています。論文の根拠となったデータがあれば実験や調査の再現が可能になり、研究も加速され、不正な論文も減らせます。産業界や一般の市民もだれもが利用できれば、新しい発想の研究が生まれるかもしれません。その成果が国民に還元されることになるのです。日本でもこうしたオープンサイエンスのしくみを早急に整える必要があるでしょう。

私は生命情報学が専門で、ゲノム科学用のデータベースや医療用研究データの共有のためのシステムについて研究してきました。それを人文科学や自然科学も含めた全ての学術分野に広げて、国民が自由に利用できる学術情報の知識基盤を構築しようと考えています。NII は全国の研究機関を結ぶ学術ネットワーク「SINET」、学術認証フェデレーション「GakuNin」、機関レポジトリのクラウドサービス「JAIRO Cloud」、国内の研究論文を検索できる「CiNii」など多数の事業を展開していますから、それらとうまく結びつけて展開していく必要があります。

■膨大なデータ結びつけ巨大データベースに

今は、複数のルートから登録された論文や研究者情報がデータベース上で統合されていない状況で、そこに研究データまで結び付けるのは現実問題として難しい課題です。加えて、大学以外でのプロジェクトではデータを集めにくく、アーカイブ化も難しいでしょう。データ整備に関わるような貢献者の名前も評価対象として上がってきません。貴重な研究データは各機関にバラバラに分散していて、その粒度や管理のポリシーも分野毎に異なります。管理が放棄されると失われてしまい

ますので、オープンサイエンスでは研究データを再利用しやすくすることが求められます。そこで役立つのが「リンクトオープンデータ (LOD)」とよばれる技術です。

インターネット (Web) 上の文書は人間には文字や画像として読めますが、そのままではコンピューターには理解できません。コンピューターが読み取りやすい形式でデータを公開し、それらをリンクさせることで、全く違うデータ同士が結びつく「データ版の Web」のようなネットワークを構築することができます。この LOD の技術を適用すれば、別々のデータベース上にあっても、インターネット全体を 1 つの巨大なデータベースとして検索できるため、探したいデータを一度で手に入れることが可能です。異なる分野の知識がつながると、例えば生物学のデータと化学のデータから薬の目的の化合物を絞り込むといった成果も期待できます。研究データをどう蓄え、共有するのか。どのタイミングでオープンにするか。全国展開や国際協調するにはどうすべきか。課題はたくさんありますが、最終的には、専門家だけでなく、高校生や小中学生であっても興味のある研究データにアクセスでき、問題を解くためのソフトウェアがクラウドサービスで即座に提供される、そんな社会を実現したいと思います。

(構成=ライター・財部恵子)

