

ひとはく 研究員 だより

近ごろビックデータとかオープンデータという言葉をよく耳にします。ウェブ環境やパソコンの性能が著しく進化したおかげで、一昔前には考えられなかった膨大なデータを使った解析や調査ができるようになりました。

自然史系博物館の昆虫や植物などの標本には、いつ、どこで、誰が採集したかというラベルデータがついています。これらがオンラインで使えれば、研究だけでなく保護区域の設定など、環境政策に

主任研究員 高野温子さん



大いに役立ちます。文字情報だけでなく標本画像も使えば、さらに有益です。

例えばヒマラヤ産シャクナゲ属の標本画像1万点余りを解析し、地球温暖化により最近125年間にシャクナゲの開花日が2・2日早まっていることを示した研究があります。世界各国の主要な植物標本庫は画像付きの標本データベース(DB)をウェブで公開しています。世界中の標本庫を訪ねないでできなかった研究が、研究室にいながらできる時代になったのです。

残念ながら日本は、画像の

公開が標本点数の1%足らずにとどまり、他国に比べて大きく遅れています。これは、標本貼りからデータの入力まで手間暇がかかる標本の整理作業に、十分な人手と予算がつかないためです。画像撮影やDBのウェブ公開にまで手が回らないのです。

これを自動・省力化するため、一定の画質の標本画像を早く大量に撮影し、その画像からラベルデータを自動で読み取れないか、数年前からアイデアを温めていました。欧米とは異なり日本語には漢字、かな、カタカナがあり、植物標本ラベルに和文と英文が入り交じるため、数年前までは技術的に不可能でした。

しかし、この1、2年に日



画像化を進めている頌栄短期大学の植物標本

標本を画像データ化

取り可能なアプリが開発され実現の可能性が見えてきました。当館の岩槻邦男名誉館長に寄付を頂けることになり、標本の画像化に踏み切りました。

今は約25万点ある頌栄短期大学植物コレクションの標本画像化に取り組んでいます。少ない予算で効率的に作業するため、市販のカメラにさまざまな工夫を施し、誰が撮影しても一定の画質を確保でき、1日で最大千枚の標本画像が撮影できる装置を共同で開発しました。2017年11月に撮影を開始し、これまでに15万点以上の標本撮影を済ませていきます。撮影と並行して、標本画像からのラベルデータ抽出の自動化も進めています。