

データサイエンスと情報・システム研究機構の取り組み

北川 源四郎 (大学共同利用機関法人情報・システム研究機構 機構長)

情報通信技術の飛躍的発展によって、研究分野や社会において複雑・多様なデータが爆発的に増加し、その有効活用が今後の研究発展や産業におけるイノベーションの鍵となっている。産業革命が人間から多くの肉体労働を解放したと同様に、このビッグデータの登場は人間の知的労働を大きく変えるものと予見されており、いま人類社会は新たな歴史的転換点に直面している。やがて、データ中心の社会が出現し、すべての研究はデータサイエンス化するといっても過言ではない。ただし、それを実現するためには、研究基盤、科学的方法論および人材育成のすべてについて格段の変革が必要になっている。

実際、ビッグデータが内包する価値は膨大とはいえ、一般にその価値密度は低く、複雑・多様データから重要な情報や知識を取り出すことができなければビッグデータの活用は実現しない。この課題を解決するためには、科学的方法論の観点から言えば、従来の実験科学と理論科学、さらには近年急速に発達したシミュレーションを中心とする計算科学に加えて、大規模データからの知識獲得のための第4の科学（データサイエンス）の確立が不可欠となる。一方、研究の基盤の観点からは、データと知識の共有・統合・活用による知識発展を目指すオープンサイエンスを視野に入れたデータ共有基盤およびビッグデータからの知識獲得のためのデータ解析基盤の構築

が必要である。そして、データサイエンスを推進するためにはデータサイエンティストの育成も不可欠である。データサイエンティストは、データ解析、データ可視化、ビッグデータ工学などの要素技術と領域分野に習熟するだけでなく、問題の本質を見抜いて課題解決する力や研究倫理やセキュリティの能力も必要である。

このような認識に基づき、情報・システム研究機構ではこれまでの経験と実績を活かし、平成28年度から始まる法人第三期においては、データサイエンス共同利用基盤施設を設立し、機構が強みを持つ生命科学分野、地球環境分野、人間社会分野を中心にデータ共有支援事業を発展させるとともに、ゲノムデータ解析支援やデータ融合計算支援などのデータ解析支援事業を推進し、我が国全体のデータサイエンス化に貢献したいと考えている。

現在のデータ中心科学リサーチコモンズを発展させ組織化したこの施設は、他の研究所と同様に運営会議による開かれた運営体制を持ち、人事評価も人事自体もデータサイエンティストの仕事が正当に評価され、データサイエンティストがキャリアパスを形成できるような組織にする必要がある。データサイエンスの時代においては、データを中心に過去の仕事や他の領域と繋がり、そして知識が発展していくことが想定される。論文の形ひとつをとってもそれが可能になるような形に変わっていかなくてはならない。当機構はデータサイエンスの時代を先導する役割を目指していきたい。