

データの時代に生きる



国立情報学研究所所長、東京大学教授

喜連川 優

「コンピュータの役割は情報処理と情報管理である」と小生が大学院学生の1980年頃に教えられた。当時は情報処理が花形であり、スパコンに代表される高速計算が注目されていた。膨大な情報の管理はデータベースの役割であるが当時はまだまだで、どちらかと言うと地味であった。弾道的高速計算から始まったコンピュータとは言え、必ずデータ管理が中心になる時代が来るに違いないと学生時代に真剣に考えたことが思い出される。

その後もコツコツと地道な研究を続け、2004年、たまたま、情報系の大型科学研究費プロジェクトの公募に申請する機会に恵まれ、「情報爆発」をタイトルとしてみた。幸い採択され、500名以上の多くの先生方と研究を進めることが出来たが、当該プロジェクトがまさに終了した年、つまり提案から8年後の2012年に、皮肉にも米国からビッグデータインシアティブが発表された。日米の共同研究は叶わず残念であったが、目指すところはほぼ同じであり、日本は当該研究領域に於いて遅れているところは一切ない。近年は、米国で盛り上がった研究テーマを日本が追いつく形でプロジェクトが立案・実施されることが多い中で、情報爆発という妙なタイトルのテーマをビッグデータがブームになる前によくぞご採択頂けたものだと思ながら感謝の気持ちで一杯である。

研究者人生で常に考えてきたことは、今まで誰もいじったことがない、誰もいじることが出来なかったでっかいデータを使って誰も見たことの無い世界を見よう、ということだったかもしれない。その為の多様な技術を長年研究する中でどんなに大きなデータでも怖くなくなった。日本で30人限定という内閣府最先端研究開発支援プログラム（FIRST）の中心研究者を務めさせて頂いた際に、医療経済研究機構と一緒して年間400億レコードに上る我が国全レセプトデータプラットフォームの構築に挑戦することとした。400億件はでっかい！これまで約1か月程度かかっていた処理が目の前ですぐ終わるようにした。FIRSTで開発した超高速データベースエンジン（OoODE）がその実力を発揮する。システムはまだまだ改良の余地が大きい、取り憑かれたように東大駒場に来られる研究者もおられる。一方でデータが汚いというご批判もある。ただ、いろいろとデータと付き合っていると、「きれいなデータなどどこにも無い」ということは言える。とりわけ膨大なデータになればなるほどノイズは多い。しかし、だからと言って何もしないというのはあまりに勿体ない。ノイズを許容し如何にして新たな知見を絞り出すかが知恵の見せ所であろう。発病と治療の地域差や経時的変化、医療費の地域差が透き通るように見える。国家の支出の中で社会保障費に占める割合は圧倒的に大きい。介護を含め医療のクオリティを維持しつつ如何に効率化を図るかに関し、データがヒントを与えると確信する。医療に限ったことではない。ありとあらゆる分野で極度に複雑化したシステムが対象となり、エビデンスとしてのデータに頼らざるを得ず、多様な方法論が模索されると共にデータプラットフォームが注目されている。

長い年月を経て、ようやくデータの時代になってきた。しかし、データを沢山持っている人のところに行き解析をして面白い結果を導き出すという時代はもはや終焉に近い。即ち最初からデータは揃っていない。如何にデータを作るかがポイントとなる時代になったと言える。筆者は最近「Observable By Design」という言葉を捻りだしてデータデザインの説明に利用している。何らかの施策を打とうとする際、その効果を示すデータを可観測になるように設計することが肝要という主張である。データ解析の時代からデータデザインの時代へのシフトである。

以前に財務省から講演のご依頼を頂いた際に、こんなことを申し上げたことが思い出される。「我々研究者は論文を書くためだけに研究をしているのではない。財政政策に役立つデータはまだまだある。その活用に携わり、国家に資することが出来れば大きな喜びである。」

[財務省広報誌「ファイナンス」はこちらからご覧いただけます。](#)