仮説なき研究の時代

小林信一 こばやし　しんいち
科学技術イノベーション政策アナリスト，広島大学高等教育研究開発センター長・特任教授。

今回は，今後の科学技術研究の重要分野の一つであり，科学技術のみならず産業活動や社会生活にも広範な影響を及ぼす可能性のあるビッグデータ分析など，いわゆるデータサイエンスに関わる研究活動について考えてみたい。最近の政策議論では，中身に関する検討がないか，検討内容が公表されずに，ポンチ絵として施策が提示されることが多い。第 6 期科学技術基本計画の策定が近づく中で，個人の力では限りはあるが，政策立案に際して，こんなことを考えたいという願望を例示したい。政策は本来，ボトムアップでこうした議論を活発に行うことで立案に至るのが望ましいが，そう言って何もしないでいるのも無責任なので，稚拙ながら論点をいろいろと考えてみた。素人談義なので間違いもあると思うが，その訂正も含めて，各所で活発に議論が行われること，それにより科学技術基本計画がよりよいものになることを期待したい。

1．医療ビッグデータ時代の幕あけ

「次世代医療基盤法」（医療分野の研究開発に資するための匿名加工医療情報に関する法律）（2017 年 5 月公布）は，2018 年 5 月に施行され，ちょうど 1 年が経過したところだ。次世代医療基盤法を医療ビッグデータ法と呼ぶこともあるように，この法律は医療情報を全国レベルで大規模に収集・集積し，研究開発にビッグデータ，AI などを活用して，新しい健康・医療・介護システムを構築しようという壮大なゴールを目指している。2020 年度にはデータの利活用基盤を本格稼働させるべく，今後は準備が急速に進むと思われる。

医療分野ではすでに，がんに関して医療情報の全国レベルの収集を開始している（全国がん登録）。従来収集されてきたがん患者の情報は必ずしも網羅的ではないなどの限界があった。2013 年に議員立法で成立した「がん登録推進法」（がん登録等の推進に関する法律）（2013 年 12 月公布，2016 年 1 月施行）は，対象医療機関を拡大するなどして，がんの罹患，診療，結果などに関する情報を都道府県経由で，国立がん研究センターの全国がん登録データベースへ集約し，越境受診などを調整して，がん患者に関する正確な情報を把握することを目指した。

このような医療情報は，個人情報の中でもとくに慎重に取り扱うべき情報である。その扱いに関しては，厳しい制約が課されるかと思えば，医療研究の促進のために条件付きで緩和されるといったふうに揺らいできた。全国がん登録と前後して，「個人情報の保護に関する法律及び行政手続における特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律」が 2015 年に公布，施行された。これに対応して厚生労働省は，病歴や遺伝子情報などを要配慮個人情報として位置づけた。その結果，病歴や遺伝子情報などの要配慮個人情報を，本人の明示的な同意を得なくても第三者に情報を適用できるオプトアウト方式の適用対象から除外し，本人の同意を要するオプトイン方式の適用対象とした。そうなると，医療情報の全国レベルでの正確な実態把握が困難になる可能性が生じる。

そこで登場したのが次世代医療基盤法であり，これは個人情報保護法に特例を設け，一定の制約の下で，とくに研究開発のための利用を容易にするために，医療情報の収集にオプトアウト方式を

適用した。これにより，ビッグデータ，AIなどの活用による新しいタイプの医療研究が可能になる。医療ビッグデータ時代の幕あけである。

ライフサイエンス分野では従来から，バイオリソースの収集，保存が行われていたが，これに遺伝子情報も加わり，バイオバンク，さらにはゲノムデータバンクなどの整備が進められている。とくに2010年代以降，多人数のゲノム情報，生活歴，病歴などを長期にわたり追跡調査をするゲノムコホート研究が活発化し，大規模なデータ集積が目指されるようになってきた。有名な計画の一つが，オバマ大統領(当時)が2015年に打ち出したPrecision Medicine Initiative(精密医療イニシアティブ)である。従来の医療は基本的には「平均的な患者」のためにデザインされている。しかし，同じ病気だと思われていたものも，遺伝子レベルで見ると原因となる遺伝子異常には複数のタイプがあったり，同じ遺伝子異常があっても発症に違いが出たりと単純ではないことが明らかになってきた。抗がん剤の効き方に個人差があるという話はよく聞くことだ。このような個人の遺伝子・生活環境・生活習慣を考慮した医療への転換を目指すのが，精密医療イニシアティブである。その基盤となるのが「100万人ゲノムコホート」，すなわち100万人規模のゲノム情報，生活歴，病歴などを長期にわたり追跡調査しようとする壮大なデータバンク計画である。日本でも，100万人には及ばないものの，複数のゲノムコホートデータバンク計画が進行中である。日本の次世代医療基盤法は，個人の病歴などの情報を大量に集積し，ビッグデータ解析などにより，精密な医療の実現を目指すという点では，ライフサイエンス研究の近年のトレンドを反映したものと言える。

2. ビッグデータ研究の気持ち悪さ？

ところで，バイオバンク，ゲノムデータバンク，ゲノムコホートなどは，実は仮説のない研究である。もちろん，仮説を適当に並べ立てることはできる。しかし，仮説の検証のために最適な研究計画をデザインし，それに沿ってデータを収集しているわけではない。データ量が多ければ多いほど，意味のある結果を得られる可能性が高まるという合理的根拠はないのである。そこにあるのは，大量にデータを集めて，それを分析すれば，従来の研究では見落とされてきた事実や関係が発見できるのではないか，という期待または予想である。いわば「期待(予想)にもとづく研究」である。それにもかかわらず，ゲノムコホートデータバンク計画は，大量の資金とマンパワーを投資して進められている。実はビッグデータを利用した研究のほとんどは，「期待(予想)にもとづく研究」である。この種の研究様式は比較的新しく，伝統的な研究様式とは大きく異なる。そのため，なんとなく気持ち悪い，落ち着かないなど，違和感を覚える人も少なくないようだ(とくに年長者)。

我々は学生時代に，研究というものは，まず仮説を立案し，それを検証するための方法を考え，実際に検証し，仮説の適否を判断するものだと教えられてきた。これは，自然科学だけでなく，社会科学の多くの分野にもあてはまる。もちろん仮説は，単なる思いつきではダメで，意味のあるものでなければならない。過去の理論や知識の漸次的な改変のような「安全な」仮説は，検証に成功するかもしれないが，その結果に革命的な意味はない。一方，過去の理論や知識を覆すような大胆な仮説は，検証も難しく，失敗する可能性も高いが，検証されれば，その結果は学問的に大きいインパクトをもつことになる。

それでは仮説はどのように導かれるのか。一つの方法は過去の研究の延長線上で，解くべき問題を選択し，それを仮説の形に変換することである。これは，パラダイム論における通常科学の標準的な営為であり，その仮説の検証はパズル解きに喩えられる。一方で，対象とする事象を観察することで，過去の研究では説明できないことを見出したり，過去には説明されていない事象を発見したりすることなどを通じて，仮説を設定することもある。この場合，地道な観察を必要とすることがあり，そうした観察とその記録自体が研究とされ

る場合もある。通常の研究が仮説に駆動される研究であるとすれば，これは仮説を探索する研究とも言える。

学生時代には通常，仮説駆動型(仮説検証型)研究は研究として認めてもらえるが，仮説探索型の研究は研究ではないと指導される。要するに「仮説がないものは研究とは言えない」と教えられるのである。その考え方は多くの研究者の思考や行動に染みついている。研究費の獲得のための申請書の作成に際しては，仮説検証型のストーリーで研究計画を作成する。どんな結果が出るかわからないが，とにかく調べてみる，というタイプの研究は，先行研究や既存の理論などとの関係が明確でなく，研究手法の妥当性や研究結果が得られる可能性も判断しにくいため，研究計画が熟れていないと評価される可能性がある。対象とする事象によほどの意味がない限りは，仮説探索型の研究は評価されにくい。これが年長者の感覚であろう。

ところが，ゲノムデータバンクのみならず，データマイニング，テキストマイニング，ビッグデータ分析などを用いた研究のほとんどは，仮説のない研究，仮説非依存型の研究である。前述のように，研究として意味のある結果が得られるという合理的根拠はない。また，仮説検証よりも仮説の発見に重点を置く仮説探索型の研究だとも言える。伝統的にはあまり評価されなかった研究様式だが，これが今日，急速に進展しているのである。後述するようにデータサイエンス，データ駆動型研究と言われたり，科学の第4のパラダイムと言われたりして，新しい科学研究の様式として確立しつつある。次世代医療基盤法は，医療ビッグデータの収集とその分析により，仮説非依存型の新しい様式の研究活動を促進することを支援する法的枠組みだともいえる。

3. バベルの図書館とビッグデータ

ビッグデータの話を聞いて，『バベルの図書館』を思い浮かべる人も少なくないだろう。『バベルの図書館』は，アルゼンチンの作家であり，短編小説を多数残したホルヘ・ルイス・ボルヘスの作品の一つである。表題にあるように図書館の話であるが，この図書館は空想上の図書館である。簡単に紹介すると，六角形の部屋が一つの単位となり，それが平面的に，また垂直方向に連なっている図書館で，一つの部屋には計20段の書棚があり，それぞれには32冊の本が所蔵されている。すべての本は，各ページ40行，1行に80文字の410ページの分量の情報が詰まっている。本は，アルファベットと空白，コンマ，ピリオドの組み合わせで埋め尽くされているが，書かれている文字の並びが同じ本は存在せず，すべての本で文字の並びは異なっているとされている。つまり，バベルの図書館には，あらゆる文字の並びの組み合わせが存在していることになる。その図書館の中には，意味のある文が出現するのを探す役割の人もいるという話である。そこにあるのは，無限の文字の組み合わせの中から物語を見つけ出せるという信念である。

英語の論文を例にとれば，特定の論文は，(図表や数字はとりあえず棚上げにすると)アルファベットと空白，コンマ，ピリオドの組み合わせであるから，その論文と同じものは，バベルの図書館の蔵書の中のどこかに存在しているはずである。つまり，バベルの図書館は文字によって表現しうる，過去と未来に出現するはずのすべての知識や物語を蔵書していることになる。たしかに言語というものは，少数の記号を組み合わせることで様々な表現を生み出すものであるから，文字のあらゆる組み合わせが存在する図書館であれば，すべての表現が，図書館の蔵書のどこかに記されているはずである。

このような図書館が存在するとして，蔵書の数は何冊になるだろうか。これは繰り返しのある順列なので簡単に計算できる。文字の種類をnとすれば，nの$(40\times80\times410)$乗の本が存在することになる。一部屋あたり20×32冊蔵書されているので，部屋の数は$n^{(40\times80\times410)}/(20\times32)$になる。この数字は，図書館の大きさが我々の存在している宇宙の大きさをはるかに凌ぐことを意味してい

るが，それでも本の数は無限大ではない。そこで，どこかに意味のある文章や論文が存在しているはずであり，それを探すという物語が成立するわけだ。ただし，その探索が現実的に意味のある時間内に終了する(アルゴリズムが存在する)保証があるかというと，ほぼ不可能である。いわゆる計算量の問題であり，コンピュータの性能が格段に向上しても，太刀打ちできないだろう。

データマイニング，テキストマイニング，ビッグデータ解析などのデータが駆動する研究(データ駆動型研究)では，大量のデータを処理するが，バベルの図書館の蔵書ほど大量ではない。測定や観測などデータ収集が可能な範囲の有限の分量のデータである。かつては不可能であると思われていた大量のデータもコンピュータの性能が大幅に向上した結果，実用的な時間内に一定の解を求めることは不可能ではなくなってきた。もちろん，問題の性質によっては実用的時間内で最適解を求められない場合もあるが，一定の範囲の近似解であればよいという条件や，何らかの制約条件を外部から与えると，比較的容易に暫定解が求まる場合もある。

つまり，大量データの中から，意味のある事実や関係を発見できる可能性が出てきたのである。もちろん，データの削除，追加などによりデータセットが変更されれば，解は変わる可能性があるので，解は暫定的なものである。ただし，現実的に有用な解であれば，時間をかけて最適解を求めるよりも，短時間で暫定解が決まるほうが望ましい場合も少なくない。とくに状況が目まぐるしく変わるような環境下では，状況適応的であるほうが望ましいとも言える。近年は技術開発のみならず，科学の分野でも「アジャイル」(敏捷な)という言葉が使われることがあるが，データ駆動型研究もアジャイルな性質をもっている。

4. 仮説駆動型研究の幻想

データ駆動型の研究は，ある意味では科学的な厳密性がなく，暫定的結果しかもたらさないのに対して，仮説駆動型の研究様式は，仮説検証というプロセスにより，理論的に望ましい結果をもたらすはずである。この考え方を支えているのは実証主義であり合理主義である。簡単に言えば，対象とする事象の背後には理論やモデルが存在し，それにもとづいて事象が作動しているので，それを直接または間接的に測定，観測，推定できると想定するのである。もちろん，そこで真理と想定しているものは反証可能性のある真理であり，絶対的な真理ではない。また，実証的分析も，普遍性のある結果をもたらすのではなく，あくまでも特定のデータ，事例，観測にもとづく推定値をもたらすという留保はついている。それでも，実証的分析により，分析結果と普遍的な理論やモデルとの関係について何らかのことが言えると考える。しかし，仮説駆動型の研究様式も，理論やモデルの存在を想定していることを除けば，実はデータに依存している。しかも，ビッグデータの場合と異なって，相対的には少数のデータから結論を導く傾向にある。

この部分は，素人の戯言だと思って読んでいただければいいのだが，伝統的な数学的，あるいは統計学的手法は，いかに計算量を節約するかを腐心した結果ではないかと思うことがある。いわば「計算量節約のドグマ」である。例えば統計に関して言えば，無限母集団を想定し，正規分布などの指数分布族を前提とすることで，各種の統計量が容易に推計できる。しかし，対象が多いだけで，本当にサンプルが無限であるとは言えない場合まで無限母集団として扱うのは近似である。無限母集団として扱えば，データの扱いは非常に簡単になる。多数ではあっても有限な対象を観測するとすれば，そこでの分析はビッグデータの扱いとほとんど変わらない。テーラー近似やフーリエ近似も，複雑な関数を現実的に扱える形に近似するものだ。確率過程における無記憶性，時系列データに関するエルゴード性も，過去の履歴を捨象することで，計算を簡略化する。

これらはいずれも，計算量を節約して，計算を簡便なものにするための苦し紛れの方法と言えな

くもない。極端なことを言えば，コンピュータのない時代に，紙の上で計算できるまで簡略化できたのである。しかし，今日のようにコンピュータの能力が格段に高まれば，あらかじめ観念的な真実または真理を想定し，それにもとづいて近似しなくても，かなりの規模のデータを直接扱える。オーソドックスな手法が切り捨ててきた諸要素を考慮して分析することが可能になる。例えば，過去の履歴を含む規則性もコンピュータの計算能力によって発見できる可能性がある。このように，真理が存在するという前提を棚上げすれば，仮説検証型の研究様式とデータ駆動型研究とは似通ったものになる。

ビッグデータ分析は理論やモデルの正しさを検証しないが，それに代えて，過去の履歴やネットワークの動態を含む，文脈依存的な解を求めることができる。それはまさにストーリーの発見である。それは必ずしも十分に安定したストーリーではないかもしれないし，データ依存的でデータセットが変わると別のストーリーに書き換えられるかもしれない。これは文学的である。データ駆動型研究様式は，データの中から適切な物語(ナラティブ)を発見するので，バベルの図書館の小型版とも言えるかもしれない。

仮説なき研究は，人文社会科学分野の本質である。その意味でも，データ駆動型研究では，研究様式の文系，理系の違いは小さいのかもしれない。最近，理工系の研究者がナラティブが重要だというのを何度か聞いて，驚いたことがある。たしかに，現実社会との関連がある研究の場合や，例えば，SDGs(持続可能な開発目標)に科学研究が参画するような場合には，ナラティブが科学と社会を結びつけ，活動の方向づけをするのに有効かもしれない。

5. 統計的アプローチと計算論的アプローチ

ビッグデータほどではないにしても，統計的手法に慣れた者にとって気持ちが悪いのが，計算論とか計算主義と呼ばれる手法である。統計学は，伝統的には統計的に記述されたモデルを設定し，収集されたデータに統計的手法を適用して，モデルが成立するかどうか(厳密に言えば，否定されないかどうか)を検証することで，真理や理論に迫ろうという方法である。これは仮説検証型研究の核心部分である。一方，計算論では，あらかじめモデルを決めず，ある事象に関する多種多様なデータを収集し，データからモデルを推定しようという考え方である。何らかの方法でモデルを導出して，そのモデルの再現性や正確性が高ければ，そのモデルは現象を説明していると考える。可能性としてありうるモデルを多数想定し，その中で再現性や正確性が高いモデルを選択するという方法(モデル選択)もある。いわば，理論的に導かれたモデルでなくとも，データに対する適合度が高いモデルがあれば，それを採用すればよい，現実に適用するモデルとしては相対的に優れていると考えるわけである。ただし，当然ながら，データが変われば，まったく異なるモデルが選ばれる可能性もあり，不安定な面もある。

この計算論的アプローチは，現実のデータを再現できる，いわば予想が当たるモデルがいいモデルだと考える点で，極めて現実的でアジャイルな性質をもっている。もう一つの特徴は，仮説駆動型ではないが仮説非依存というわけではなく，むしろ仮説探索型のアプローチだということである。想定されるモデル群の再現性を推定するためには，モンテカルロ・シミュレーションなどによってモデルの挙動を検証することも可能である。実験，観察，測定といった実社会でのデータ収集ではなく，コンピュータ実験ともいうべき新たなデータ収集，検証の手段を生み出したとも言える。マルチエージェント・シミュレーションは，コンピュータ空間の中に，固有の行動モデルをもつエージェントを多数設定し，さまざまな条件の下で相互作用の時間変化をシミュレートするものである。それにより，模擬的に社会や組織の動態を分析し，モデルを探索する。

20 年くらい前に，統計学者と計算論にもとづ

いて研究する各分野の研究者たちが集まって議論をする場に参加したことがある。小規模な研究会だったが，モデルを前提として分析をする統計学者と，データから入ってモデルは結果的にできればよく，そのモデルは統計理論に従うことよりも，現象をよりよく再現できればいいのだと考える計算論学者の考え方は鋭く対立し，交わることはなかった。新興勢力である計算論のほうが威勢よく，統計学者は押され気味だったように感じた。

現実には，統計学と計算論は必ずしも対立するだけではなく，相互に影響しあっている。統計学でも構造方程式モデリング(共分散構造分析)のような統計的モデル選択の手法は今日一般的に使われており，統計学的アプローチといっても，事前に唯一のモデルを想定しないアプローチもあるのだ。コンピュータの発展の結果，統計における最尤推定その他に計算論的アプローチを活用することもある。また，これも古くからあるアプローチであるが，ブートストラップという方法もある。これは簡単に言えば，モデルの前提となる確率分布が事前にはわからない場合，大量データの中から，一定規模のデータをランダムサンプリングして統計量を推計するという手続きを，非常に多くの回数繰り返すことで，統計量の分布そのものをシミュレーションで推計するという手法である。このように統計学的アプローチでも，計算論的アプローチ，データ駆動型のアプローチとミックスすることが一般的になっている。一方，データマイニングやビッグデータ分析などのデータ駆動型研究や，人工知能における機械学習などでは，統計学的手法が多用されているという現実もある。

6. データ駆動科学の今後

このように，今日では仮説を前提として検証するという伝統的な仮説検証型の研究様式と，少なくとも事前には仮説に依存しない，場合によっては仮説のないデータ駆動型の研究様式が，ともに急速に発展し，また両者は相互に補完関係を強めている。こうした研究様式の変容が頻繁に議論されるようになってきているが，現段階ではその姿を表す語で定着したものはない。最近では，いくつかのコンセプトが登場している。データサイエンス，eサイエンス，デジタライゼーション，デジタル・トランスフォーメーション，X-informatics(informaticsという語尾の学問分野群)，アジャイルな科学，科学の第4のパラダイム，等々。これらは単なるハイプ(大げさな物言い)やバズワード(流行)なのか，それとも新しい研究様式の出現を正しく捉えたものかは，いずれ明確になるだろう。どの言葉も，現在進行中の大量データの利活用にもとづく科学研究の変容のある側面を表現しようとしていることは確かである。

伝統的には仮説検証型の研究が王道であり，仮説のない研究や，見通しもなくとにかく大量のデータを収集するといった研究は邪道だったわけだが，コンピュータの性能向上に伴って発展してきたデータ駆動型研究様式は，無視できないものとなってきた。また，仮説検証型の伝統的研究様式との混交も進んでおり，仮説のない研究は研究ではないと言って済ますことができない状況になってきている。そこで新たな政策的課題が生じる。

第一は，科研費その他の研究資金配分制度の下で行われるプロジェクトの事前評価に際して，そのような研究計画をいかにして適切に評価するかである。AI，ビッグデータ，大量データの活用を前提とする仮説なき研究を伝統的な研究観の中で振興できるかである。とくに，仮説もないのに大量データの収集を実施するために行われるデータバンク事業などには大量の研究資金が必要であり，新しいタイプのビッグサイエンスとも言える。本稿冒頭の医療ビッグデータは，仮説なき研究の推進を法律によって支えていこうとしているものだとも言える。これも一つの選択肢であろう。いずれにしても，ビッグデータ時代の新しいタイプの研究活動をいかにして組織化し，それをいかに支援するかは，今後の科学技術イノベーション政策にとって重要な課題であり，また重点的に取り組むべき可能性を秘めた研究活動である。

第二はデータの質の問題である。ビッグデータ

時代の研究にとってデータが決定的に重要であることは言うまでもない。いくらビッグデータ分析といっても，質のよいデータが揃っていないと意味はない。しかし，現実にはデータにはノイズが多く，研究に必要なデータセットを準備することは容易ではない。インターネットから収集するデータには，とくに注意を要する。そこにはフェイク情報もあれば，意図的にバイアスをかけたデータも存在する。論理学の基本的な法則として矛盾律(無矛盾律)がある。ある命題が真であると同時に偽であることはない，すなわち矛盾する二つのことが同時に成り立つことはないという当たり前の法則である。逆に，矛盾の存在を前提とすれば，あらゆる命題を導くことができる，またはあらゆる命題が真であるという帰結(何でもあり)にいたることも矛盾律から導かれる。つまり，ネットの中に相互に矛盾するデータが存在していれば，ビッグデータ分析の場合には，アルゴリズム次第で，導かれる分析の結果は不安定になったり，かなり怪しいものになったりする可能性が高い。ノイズを消去すること，フェイク情報を識別して排除することなどを通じて，良質なデータセットにすることが必要になる。それ自体を AI で処理することなども行われているが，一方ではフェイク情報を本物の情報のように見せかける AI 技術なども存在しており，イタチごっこの状況を呈している。

実験結果や測定値などの研究データの場合には，基本的には統制されたデータなので，質の悪いデータが含まれる可能性は低いと思われる。しかし，いわゆるデータの捏造や改ざん，といった質の悪いデータが紛れ込む可能性は否定できない。オープンデータ化の動きもあり，今後は大量の研究データが集積され，広く利用されていくと予想されるが，捏造や改ざんが紛れ込むと，せっかくのオープンデータを活用したデータ分析でも，品質が低下する原因になるだろう。データの質の維持・向上と，そのためのデータマネジメントやデータキュレーションなどの果たす役割が大切になる。

第三は倫理的問題である。医療ビッグデータの場合には，データの扱いに関して細かい条件が定められている。仮説なき研究を推進する一方で，倫理的扱いについても法的に規定しているわけである。医療データの扱いは個人情報保護とのバランスの中で決まってくるものだが，インターネットやその他のデータソースから大量に収集されるビッグデータに関しても，データを収集する者，そのデータを利用する者はデータの扱いについて慎重になる必要がある。意図しないで収集したデータの場合でも，さまざまな情報を連結することで個人を特定できる可能性はある。データ収集者が，特定のデータを人為的に紛れ込ませる，意図的にバイアスをかけたデータ収集をする，といった可能性もある。データの分析の結果，機微な情報が得られる可能性もある。

科学技術研究と民間におけるビッグデータの利活用とそれを支える研究の推進は，目前の課題であり，政策的にも重点的に振興すべき分野の一つである。しかし，関連する分野，検討すべき課題は広範に及ぶ。バランスよく目配りする必要がある。同時に，最近，AI の倫理が頻繁に議論されるようになってきているように，データ倫理，ビッグデータの倫理といった課題も，研究開発の推進と同時並行で取り組む必要がある。

小林信一 こばやし　しんいち
1956 年生まれ。1986 年東京工業大学工学部助手。文教大学国際学部講師，電気通信大学情報システム学研究科助教授，文科省科学技術政策研究所総括主任研究官(併任)，筑波大学大学研究センター助教授，同ビジネス科学研究科教授などを経て，国立国会図書館調査及び立法考査局専門調査員。2017 年 3 月定年退職。2018 年 7 月から広島大学高等教育研究開発センター特任教授。同年 10 月にセンター長就任。専門は科学技術政策，高等教育政策，科学技術論など。