

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 29 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26350359

研究課題名(和文) ビックデータ型バイロジの生成とその影響についての科学社会学的研究

研究課題名(英文) On the development and the influence of big-data based biology

研究代表者

福島 真人 (FUKUSHIMA, Masato)

東京大学・大学院総合文化研究科・教授

研究者番号：10202285

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究は、近年のライフサイエンスにおける、ビッグデータ化の諸影響を分析した。主要な論点として、そうした分野における、期待のダイナミズム 知識インフラという性格が生まれ出す、価値振動という問題、領域専門性/社会的注目度の違いによる対応の差、である。という座標軸が作り出す、異なる四象限における、その状態の差異である。こうした要素が生命科学とデータサイエンスの間の軋轢の原因となっている。

研究成果の概要(英文)： This project has focused on the three aspects of the relation between the rapid growth of data and the practice of biological research, in terms of 1) the rise and fall of expectations related to big data approach to biology; 2) the issue of value oscillation concerning the infrastructural aspect of such data science vis-a-vis the life sciences at large; 3) the analytical axioms of the degree of domain specificity of the concerned knowledge and the degree of social attention to the field. These three factors are found to be pivotal to understand the existing problems and obstacles concerning the complex relation between traditional biology and emerging big data approach, where the fluctuating expectation on the topic is influenced by the contradictory image of what data science can do to life sciences at large. It is also concluded that this phenomenon is also observable in the larger realm about which both researchers and policy makers are advised to pay closer attention.

研究分野：科学社会学・科学技術史

キーワード：科学社会学 バイオ社会学 科学技術社会論 科学人類学

1. 研究開始当初の背景

研究開始時の重要な論点の一つは、その前段階で調査を行ってきたケミカルバイオロジーあるいは構造ゲノムといった分野において、一方でゲノム研究によって急速に開発が進んだハイスループット(高速処理)のデータ処理技術の影響が現場の実践に広くみられるのに対し、他方でそこから生み出される大量のデータの取り扱いをめぐって、生物資料を扱う研究者(いわゆるウェット系)と、そうしたデータ等を取り扱う研究者(いわゆるドライ系)の間に、少なからぬ意図、目的、そして価値評価の齟齬が見られたという点である。

ケミカルバイオの場合は、どちらかという物である天然物をどうやって高速処理をするか、という点で現場の天然物化学系が四苦八苦しているという側面が強かったが、構造ゲノムについては、タンパク構造をこうした形で高速処理することが可能なのかという点について、伝統的なウェット側の抵抗感がめだち、その対立が長く尾を引いているという点が観察された。そうした対立が今回の研究における問題設定の大前提である。

2. 研究の目的

本研究の目的は、(1) まずこうしたバイオの情報化、あるいはビッグ・データ化にまつわる諸問題にかかわる理論的な概念を整理し、一般的な理論的枠組を準備する点である。こうした過程は単に産学を巻き込んだイノベーションの新たな領域を形成するというだけでなく、新たなテクノロジーの形成に関する「期待」の問題や、データベース等を含む一般的な知識インフラ形成の過程という問題とも密接に関係している。事実、こうした理論的な枠組は、生命科学だけでなく、創薬過程やイノベーション一般の問題とも深くかかわる。(2) 以上にあげた問題について、バイオ・インフォマティクス、プロテオミクス、あるいは創薬過程といった分野を中心に観察し、その具体的な特徴を探るという点である。

3. 研究の方法

研究の方法は、(1)については概念的、理論的な分析、文献解析、(2)については文献解析、部分的フィールド調査、インタビューその他等を用いた。

4. 研究成果

(1) 「期待」の動態学 - こうした急速に発展する技術に関しては、それを促進する推進力である期待(expectation)の働きと、それを押しとどめる現実のさまざまな制約という観点から問題をみる必要がある。その最初

の要素である期待、あるいはそれを扱う期待の社会学(sociology of expectation)は、こうした期待が技術革新の初期には重要な役割を果たすものの、そうした期待が現実の制約によって急降下するという時間的なダイナミズムを分析しており、初期の期待が息切れして、社会の熱狂が消えた状態で、どうやって研究を維持していくかといった興味深い問題を呈示している。論文は、こうした期待の乱高下にさらされた伝統的なライフサイエンスの分野(ここでいうウェット側に相当する)が、そうしたライバルへの期待の高まりに対して、自分たちもちょうど(カトリックの対抗宗教改革のように)ライバルの要素を取り入れつつ、自らの伝統を革新するという側面を分析しており、それを「研究過程のレジリエンス」と呼んだ。

他方、より革新的なライフサイエンスの分野、たとえば(ポスト・)ゲノム研究等においては、こうした期待の急激な高まりと、それへの反動としての失望というサイクルは歴史的に繰り返し観察されてきた(論文参照)。その顕著な例はいわゆるゲノム医療のケースで、ヒトゲノム解析の初期から、その知識を直接用いて、医療に革命をもたらすといった言説は繰り返し喧伝されてきたが、他方その実現には多くの隘路があるため、失望が常態化してきたという社会学的な分析がある。

この点に関していうと、特にバイオ・インフォマティクスにおいて、そうした期待の上昇と低下が過去において観察されており、2000年代のバイオ・インフォマティクスに対する期待の高まりと相反するように、2010年代を超えると、教育プログラム等が整備されつつある一方で、学会等の活動がやや停滞気味になり、世間の期待が失速している様子がみられた。この背後には、バイオ・インフォマティクスに限らず、こうした知識インフラに関係する領域に固有の問題があるが、その点については次章で解説する。

本研究では当初プロテオミクスも比較研究の対象にいていたが、ビッグ・データ化という面では、創薬全体の方がより劇的な側面があるため、創薬におけるコンピュータ化、情報化といった問題も同時変動的に観察、資料収集した。こちらは最近になってデータベース整理やAI化といった側面で期待が急速に高まり、一部では現状の日本の創薬における隘路や生産性の低下をカバーする最も重要な路線であるといった評価もなされる。実際、特に理研の京コンピュータに代表されるスーパーコンピュータを用いた創薬という政策がそうした急速な期待の上昇の中心にある。

(2) 情報インフラ-価値振動という問題

こうしたバイオ一般あるいは創薬過程における情報化、ビッグデータ化等々に対する急激な期待の高まりに対しては、反動の可能

性が常に存在しており、それはこうした知識インフラが「インフラ」として持つ特定の複雑な社会的要因(あるいは矛盾)と深く関係している。

その矛盾を分析したのが、論文の主要テーマであるが、簡単に言えば、インフラというのは、それがどんなものであれ、社会を基盤から支えるという意味で大きな力を発揮するというポジティブな面と、それがインフラ(すなわち、ラテン語の原義からいえば、劣ったとか裏の、の意味)にあるような)、表面にあらわれたものに対する比較劣位というネガティブな面という、相矛盾した二重構造を持つという点である。科学社会学者たちは、バイオ・インフォマティクスやデータベースの作成や維持について多くの研究をまとめてきたが、それらを通読すると、そうした知識インフラをつくったり、既に出て上がったインフラを維持したりするという作業を一体誰がやるのか、という問題が繰り返し生じるという点を指摘している。論文はそれを「価値振動」(value oscillation)となづけている。これは、インフラ的な装置の重要性は十分認識しつつ、自分がそうしたインフラ的な作業に埋没するのは避けたいという矛盾した感覚のことを示す。

この問題は、たとえばバイオ・インフォマティクスのケースでは、この分野が、1)生物学者に対する情報環境を整備する補助的な仕事なのか、2)それとも既に大量化しているバイオ情報を独自に分析し、それから生物学的な知見をえる研究者のことをいうのか、3)あるいは最新鋭の情報理論の開発をバイオ領域で行っている専門職なのか、一体どれを示すのか、実はハッキリしないという形であらわれてくる。この曖昧さは、当該学会等でも指摘されており、そのイメージの混乱が関係する学会の活動の停滞につながっているという指摘もある。これはバイオ・インフォマティクスという学際領域に潜在的に存在している価値振動の問題が、異なるベクトルとして具現化したものだが、こうした問題は、学際的な領域ではどこでも恒常的に起こりうる問題でもある。

これをウェット側の生物学者からみると、次々と開発されるデータベースや、新たな分析ソフトの展開に追いついていけないという、使用面での不満にもつながってくるが、他方、データベース作成側でも、ゲノムとタンパクではそのデータベースの性質がかなり異なり、前者に比べ後者は何かすっきりしない特性を持つのに、それをを用いるユーザー側がそうした違いを理解してくれない、といった形での不満として表明されることがある。

こうした価値振動の問題は、普通政策レベルではあまり理解されていないため、場当たりの対応に終始することになる。たとえば5年計画で大量の資金がついたさまざまなデータベースによる知識インフラの設立

は、それが、のちに続く維持、補修の努力によって継続されるという点が看過されやすい。こうした施策の初期の段階、つまりその技術への世間の注目が集まる時期を超えると、その後の活動は地味で脚光が浴びにくくなるため、それが政策に反映されるのがむずかしくなる。

バイオ・インフォマティクスに代表されるドライとウェットの関係において、最も先鋭的に出てくる問題が、こうした価値振動をどうやって現場レベルで調整するかという点である。筆者が聞き取りを行った複数の研究者のうち、こうした知識インフラの維持や保全に全員が関心を持っているわけではなく、人によっては自分の役割をその立ち上げのみに限定している場合もあった。また複数のケースにおいて、そうした保全や維持にかかわる人員を確保するのがむずかしいと感じている研究者も少なくない。

こうした分野は、一般のラボではテクニシャン(技官)の仕事とされることも多いが、フルの研究員とテクニシャンの間の関係は流動的で、実際、データベースの保全維持に関する諸問題、たとえば項目のこまかな修正といった問題は、研究者そのものが個別に対応している場合が少なくない。当然そうした地道な維持と、論文につながるような、より花々しい活動との間にある矛盾をどう解決するかは、日々の調整に依存するという形になる。

他方、知識インフラ的な作業を含めて、情報学的な側面を単なる生物学の下請けではなく、それ自体が分野を主導するものとみなす視点も存在するが、そうなると今度はウェット系とドライ系の主導権争いという面も表面化しうる。その極端な例は、論文で分析した、米の構造ゲノム学における、タンパク質構造分析をめぐる、伝統的なウェット系の構造生物学者と新規参入の情報学中心の構造ゲノム的アプローチの間にあった厳しい対立である。

もちろんすべての分野でこうした対立が表面化しているというわけではなく、研究室によっては、同一研究室の中でドライとウェット系の研究者を意図的に共存させ、両方の異なる「認識的文化」(epistemic culture)をうまく克服しているケースも少なくない。しかし日米のバイオ・インフォマティクスの研究状況に詳しいある研究者によれば、この二者の共同という点では、日本の方がかなり遅れをとっており、米(彼の場合は特にスタンフォード大に代表される西海岸)では情報系の研究者がウェット系の生物学者を大量に雇ったラボを率いるようなことは珍しくはないが、他方日本では、こうしたバイオ・インフォマティクスの会合を開こうとすると、その冒頭から「ウェットとドライの対立」という話が、最も中心的な関心として飛び出してくるという。

さらにこうした二つの領域が関係する領

域として、システム生物学といった分野も挙げられるが、この分野も情報学、計算機科学と生物学の融合という形で近年政策的にも関心を集めている。もともとこの概念そのものが日本の北野宏明によって提唱され、国際研究集会等を中心に精力的に推進されてきたにもかかわらず、現実点での政策的な強調点は、どちらかというところスパコンによる情報処理のハード面での革新の強調に偏っているような印象をうけ、必ずしもデータ解析における革新性という方向にいてないようである。

これを気象学モデルと比較して考えてみると、そこに必要なのは、1)気象の数学的モデル、2)それを計算するためのスパコン、3)そしてデータを収集するための地球全体の観測スポットの完備、4)データベースの維持管理、といった四つの領域が同時に成立する必要があるが、システム生物学のような領域では、その対象がそもそも細胞なのか臓器なのかといった点からまず関心の差があり、さらにそれに応じた一般的モデルの欠如、そしてそれを精密に計るための技術の未発達、さらには異なる分野でのデータ統合といった諸問題のために、システム生物学という分野そのものの存続がやや疑問視される傾向もあり、海外では名称の変更や予算の縮小を経験し、本邦では、むしろもっと一般的な、生命動態システム科学といった言い方を採用することで、そうしたアンバランスに対応しようとする傾向すらみえる。この背後にもさまざまな価値振動の問題が内在しているのは想像に難くない。

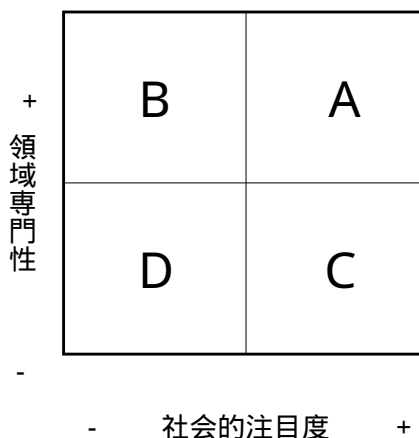
(3) ビッグデータ化の一般的な問題

こうした、情報とバイオをめぐる関係の複雑な関係は、当然バイオだけではなく、その他さまざまな領域で同様に発生しうることが想像できるが、その一例は、近年におけるデータサイエンスという用語の流行にある。これは(1)でのべた期待の動態学の典型的な事例であるが、興味深いのは、ここでいうデータサイエンスという言葉が、その職業の中核的なイメージとして何を（認知心理学者であるロッシュの意味での）「プロトタイプ」と見なしているかという点である。

少なくとも現在の政府によるデータサイエンティストの倍増計画といった数値目標は、その目的とする職種の具体的なイメージがハッキリせず、問題が多い。つまり本研究が示すように、ここで期待されているのが、1)特定の分野での先端的な分析を牽引するというイメージ(いわば社会学者ゴフマン流に言えば、ステージ上のイメージ)と、2)他方データベースのような情報基盤を維持・管理する裏方的な役割(これがインフラ研究に近いが、ゴフマン流に言えば、舞台裏のイメージ)という二つの領域に分裂しているという点である。しかもそこでいう情報あるいはデータというのが、特定の領域に本質的に固

有な専門性を持つものなのか、それともただ一般的なデータ解析ツールを使えば、それなりの結果がでるものなのかも一般的にハッキリしないのである。メディアに登場するデータサイエンスは、しばしば初歩的な統計学的解析を会社のデータ管理に応用するといったイメージに近いが、このような手法だけで、これでエピゲノムの解析ができるはずもない。さらにこうしたデータサイエンティストが、果して特定領域の研究に指導的な役割を示すのか、それともむしろインフラ的、研究補助的な役割に徹するのかも、実はハッキリしていない。これがまさに本研究が明らかにしてきた「価値振動」の問題である。

こうした観点を図式化すると次のようになる。



ここで示されているうち、横軸の「社会的注目度」というのは、筆者のインフラにおける価値振動のうち、インフラ、つまりある種の下請け作業として理解される傾向の逆であり、それにスポットライトがあたる度合いを+ - で示している。つまりその値が低い程、インフラとしての性格を強く持つという意味である。他方縦軸の領域専門性とは、その内容がどれだけ特定領域の専門性を要求されるかの意味である。

このような四象限からみると、たとえば情報科学やデータサイエンスと生命科学といった領域間の関係を考えても、基本的に四つの異なるレベルが存在するのが分る。すなわちAは特定の情報科学が専門性の高い分野(たとえばバイオ)において先導的な役割を果たす場合で、これは情報科学者がバイオ情報を分析して、そこからたとえばゲノムについての独自の法則性を見出すようなケースである。Bは領域専門性が高いが、下請け的な要素が強い場合で、これは特別な目的に特化したデータベースを開発、あるいは維持管理する様な場合である。Cは領域専門性があまりハッキリしないが世間の注目を浴びているようなケースで、これは世間で喧

伝されているデータサイエンティストのような類の人々、つまり企業等で率先してデータを解析し、それが経営戦略上も明らかリーダーシップを握っているようなケース、Dはそうした汎用データ解析者であるが、基本的に裏方として表に出ないようなケースである。

こうやってみると、たとえばバイオのような特定領域においては、バイオ・インフォマティクスの専門家に期待される役割が、AとBの間のモデルに分裂しているのはしばしば見聞する点であり、ウェット系の生物学者にとってはしばしば情報研究者はBとして裏方に徹することが期待される一方、本人はむしろAとしてその役割を全うしたいわけである。

他方メディアや政府の政策に散見するデータサイエンスという言い方はどちらかというところC Dの系列に近いものを想定しているようにも見えるが、A B的な側面も否定できず、その実態は曖昧模糊としたものである。

結論

ビッグデータ化に伴うさまざまな社会的な影響をバイオのさまざまな領域との関係で分析してみると、そこでいう情報化には、より一般的には、どういう領域においてそれが進行し、その情報の分析にどのような専門性が必要となるかといったレベルと、他方その情報化が他の領域に対してどのような関係性を持つか、という二つの座標軸があることが分かってきた。重要なのは、この二つの軸が交差する四象限において、情報化あるいはビッグデータ化の持つ意味が相互にかなり異なるという点であり、それに応じた教育や政策の内容も異なってくるという点である。実際ビッグデータ化をめぐる多くのメディア上の解説や、あるいは政策的対応も、そうした異なる領域、そしてそれに関係した目標を区別せず議論しているため、まるで汎用性を持つデータ解析者のような存在を大量に生産することがその解決であるように論じているものも少なくない。しかし現実には、それがどの領域で議論され、その中で情報系の人々にどのような役割が期待されているかは、四象限の分野においてかなり異なり、またその分野自身の中でも、ある種の潜在的な対立の可能性があるものもすくなくない。こうした分野の問題をより精密に分析していくことが、この複雑な分野のもつ隘路を克服するための重要な指針となると考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

Fukushima, M. (2017) The Experimental Zone of Learning: Mapping the Dynamics of Everyday Experiment, *Mind, Culture and Activity*. (Published on line <http://dx.doi.org/10.1080/10749039.2017.1293689>) (査読あり)

Fukushima, M. (2016) Value Oscillation in Knowledge Infrastructure: Observing its Dynamic in Japan's Drug Discovery Pipeline, *Science and Technology Studies*, 29(2): 7-25 (査読あり)

Fukushima, M. (2016) Blade Runner and Memory Devices: Reconsidering the Interrelations between the Body, Technology, and Enhancement, *East Asian Science, Technology and Society* 10:73-91 (査読あり)

Fukushima, M. (2016) Constructing "Failure" in Big Biology: The Socio-technical Anatomy of the Protein 3000 Program in Japan, *Social Studies of Science*, 46(1): 7-33 (査読あり)

Fukushima, M. (2016) Resilience in Scientific Research: Understanding How Natural Product Research Rebounded in an Adverse Situation, *Science as Culture*, 25(2):167-192 (査読あり)

[学会発表](計1件)

Fukushima, M. (2015) Infrastructural Esthetics, *Society for Social Studies of Science* (4S) 2015年8月31日~9月3日スペイン、パルセロナ市

[図書](計2件)

福島真人 (2017) 『真理の工場 - 現代科学の社会的研究』(印刷中) 東京大学出版会

福島真人 (2014) 「古道、秩序、系譜学 - 公共的制度設計についての断章」山下晋司(編)『公共人類学』141-154 東京大学出版会

6. 研究組織

(1) 研究代表者

福島真人 (FUKUSHIMA, Masato)

東京大学・大学院総合文化研究科・教授

研究者番号：10202285

(2) 研究分担者 なし