

シリーズ研究集会、「D2K サイエнтиストを緊急に養成すべき疾患領域」について
～トランスレーショナル・リサーチと D2K サイエンスの視点から～
神沼二眞、NPO 法人サイバー絆研究所

まえがき

日本バイオインフォマテックス学会 (JSBi) の支援による表記の研究集会シリーズは、
生物学 (Biomedical) 領域における新しい潮流としてのデータサイエンスあるいは D2K
(From Data to Knowledge) サイエンスへの関心の高まりと、その担い手として期待さ
れる新しい専門家としての D2K サイエнтиストの養成について、討議することを目的
としている。シリーズの第 1 回ではがんの精密診療を取り上げ、第 2 回では「腸内細菌と
健康と疾病」を取り上げた。第 3 回となる今回は、脳神経系疾患に焦点を合わせ、「計測
分析技術および ICT の進歩を踏まえた D2K サイエнтиスト」のイメージを探ることを
目的として、異なる専門家から多角的な話題を提供していただき、深い討議を展開するこ
とめざしている。

一方 ICA (NPO 法人サイバー絆研究所) は、「次世代ヘルスケアと薬づくり」をテーマ
とした研究集会 Visionary Seminar を開催してきたが、JSBi 支援の上記のシリーズ研究
集会が、ICA が企画してきた 研究集会にも関係があるため、その開催にも協力してい
る。ただ、これまでの集会では肝心のデータサイエンス、あるいは D2K サイエンスや、
その担い手である D2K サイエнтиストのイメージについては、あまり紹介や討議の時
間がとれなかった。以下の小論は、そのことを多少とも補うことが目的である。ただし、
その議論はあくまで執筆者個人の見解であり、このシリーズの世話人や関係者を代表する
ものではないことをお断りしておきたい。

生物学におけるデータサイエンスの黎明期

生物学、あるいは医療に計算機を活用しようという試みは、我が国を含めて 1960 年
代に遡る。その最初は計量診断学とも呼ばれた*。その後、1980 年代の中頃には、医学や
医療で遭遇するあらゆるデータを計算機で扱える環境を構築するという試みも、絵空事で
はなく、ある程度の格好がつくようになってきた**。例えば当時、私の関係していた研究
グループでは、数値だけでなくカルテのような記述文書、塩基配列のような記号、眼球運
動のような波形、X線写真や眼底写真のような画像、線虫 (*C.elegans*) の胚発生を追尾す
るための 2 次元画像が連続した 3 次元画像の時間的な変化 (4-D) のような多様なデータ
を扱える環境を整備してきた。また疫学的な地図や、分子や細胞のグラフィクスのような
可視化システムも開発していた。

それらのシステムは、基礎生物学の研究者から臨床や疫学の研究者まで、幅広い生物医
学系の研究者と、統計学や数学や情報計算を専門とする若手の研究者たちとの共同研究の
中で開発されていた。責任者である私は、その全体を「医学、医療で遭遇するあらゆるデ

ータを計算機で扱えるようにする」活動と捉えていた*。

そこで学んだことは、そうした研究における最大の障害は、研究対象となる現象の記録を、計算機で扱えるように変換、整えることであった。その意味では、もっとも迅速に結果がだせたのは、写真などデジタル化が自動的に行える仕事であった。

(参考文献)

*増山元三郎、少数例のまとめ方—特に臨床医学に携わる人達の為に、河出書房、1953.

*高橋暁正、新しい医学への道、紀伊国屋書店、1964

**神沼二真、倉科周介訳:診療コンピュータシステム,文光堂、1981。(E.H. Shortliffe: Computer-Based Medical Consultations: MYCIN, Elsevier,1976)

**神沼二真、医療革新とコンピュータ、岩波書店、1985.

ゲノム生物学とビッグデータ時代のデータサイエンス

今世紀に入ってからゲノム解読技術 (Genomics) とそれに随伴するようなタンパク質や代謝物の網羅的解析技術 (-omics) と、AI を含む ICT との相乗的な進歩が、生物学 Biomedicine と関連する分野を大きく変革しようとしている。その変革を支える基盤のひとつが半導体の進歩である。それによってアナログデータが容易にデジタルに変換されるようになり、あらゆる分析計測機器 (Wearable/Wireless Sensors、あるいは単に Wearables) が容易に計算機に接続されるようになった。

これは社会に広く浸透している音楽の録音と再生技術の進歩と同じである。したがって 1980 年代には、まだ研究者を悩ませていたデータを計算機で扱える (すなわち計算に使える) ようにするためのデジタルに変換に関わる作業や労力が、大幅に軽減されることになった。インターネットに各種の機器をつなぐ IoT (Internet of Things) もこうした変化の延長線上にある。

そこでは数々の創造的な破壊 (Innovation) が起こせるという期待が高まっている*。とくに携帯型の生体計測機器とスマートフォンを介したクラウド計算環境とのつながりは、生活者や患者が健康の維持や疾病対策に主体的に関わる「参加型ヘルスケア (Participatory Healthcare)」の広がりを加速すると期待されている。

だが、ゲノム研究がそうであるように、こんどはデータの量が膨大かつ多様になったことから、これまでの生物学あるいはヘルスケアの専門家だけの力では、それらを活用することは、とうてい不可能であることも明白になってきた。つまり人工知能 (AI) の応用はその一例であるが、それを含む数々のイノベーションが期待されているが、そのためには D2K サイエニストと呼ぶべき新しい専門家の参加が必須の前提であることが認識され始めた)。

* E. Topol, The Creative Destruction of Medicine, Basic Books, 2012.

NIH の BD2K 事業

このことは、そうした変革のリーダーシップをとってきた米国の NIH には、よく理解されていたようだ。彼らの対応は素早かった。その最初は、「ビッグデータから知識へ (the Big Data to Knowledge (BD2K) initiative)」を標榜する活動を 2012 年に発足させたことだ*。ここでいうビッグデータには、もちろん爆発的に増大するゲノム (配列解析) や随伴するオミックスのデータも含まれている。NIH は傘下に 27 の研究所を擁しているが、BD2K 事業は、それらを横断するだけでなく、NIH の外部の教育研究機関や団体などにも、この事業の協力センターを設置し、さらにそれらを連携したネットワークあるいは研究コミュニティを組織することをめざした。

*R. Margolis et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. JAMIA, 21:957-958. 2014.

NIH はその事業のために副ディレクター格の Chief Data Strategist というデータマネジメントの責任者の席を設けた。そこに着任したのが薬学の専門家、ボルネ (Philip E. Bourne) である。彼は自分の使命を、NIH を “生物医学のデジタル帝国 (Digital Enterprise)” にすることだ、“To foster an ecosystem that enables biomedical research to be conducted as a digital enterprise that enhances health, lengthens life and reduces illness and disability”、と表現している*。

*これは、NIH の副ディレクターとして、データサイエンス部門を任されたに任命された Philip E. Bourne 作成の下記の資料による

https://www.genome.gov/pages/about/nachgr/sept2014agendadocs/presentation_nih_assoc_director_data_science_sept2014.pdf.

*P. E Bourne et al. The NIH Big Data to Knowledge (BD2K) initiative, J Am Med Inform Assoc, 22:1114, 2015.

BD2K から「戦略的なデータサイエンス」への取り組み

NIH の BD2K 事業に関しては、NIH の公式のサイトやいくつかの短い文献がある*。しかしデータサイエンスがどのような科学であるかという肝心なことについては、あまり明確な説明がされていない。この点に関して Bourne は (自分の専門でもある?) 構造生物学の最近の進歩であるクライオ電顕 (cryo-electron microscopy) とデータサイエンスの相似と相違を例として説明している*。ただ我々のように 1960 年代からこの領域を知る者からすると、BD2K 事業は、基礎生物医学の成果を臨床医学やヘルスケア全体に移転する橋渡し研究 (Translational Research, TR) が、デジタル化とネットワーク化、さらには人工知能 AI、機械学習 (Machine Learning, ML)、深層学習 (Deep Learning, DL)、携帯型の簡便な生体計測機器 (Mobile Sensors) の活用という大波を受けて、「前例のない脱皮を迫られている状況に対応するための活動」を象徴しているように思われる。

* C. Mura, E. J. Draizen, and P. E. Bourne, Structural biology meets data science: Does anything change? *Current Opinion in Structural Biology*, October 2018.

実際、実験的な色彩が濃かった BD2K の活動を踏まえて NIH は、生物医学研究とその成果を臨床やヘルスケア一般に活用するための、本格的なデジタル対応戦略構想を昨年 (2018 年)6 月に発表した*。それは NIH が資金を出している生物医学に関わるデータサイエンスの環境をより根本的に刷新することをめざしたロードマップである。その骨子は、NIH が支援しているデータの所在、相互の関係などをわかりやすくして活用を促進すること、データを扱う既存の技法と新しく開発されたツールなどを統合的に活用できるようにすること、大学などで開発された革新的な技法を、企業などを介して広くかつ効率よく活用できるようにすること、データを扱う経費の増大に配慮することなどである。もちろん人材育成も重要なテーマになっている**。

* NIH のデータサイエンスの事業のサイト (<https://datascience.nih.gov/>)

NIH STRATEGIC PLAN FOR DATA SCIENCE.

** Annual Report on Investments in Training and Education for the NIH Big Data to Knowledge (BD2K) Initiative, 2017.

** M. C. Dunn, P. E. Bourne, Building the biomedical data science workforce, *PLoS Biology*, 15(7): e2003082, 2017.

デジタルメディシンの先行事例

以上の話を要約すれば、「医療がデジタル化されれば、いわゆる人工知能とか機械学習などはもちろん、数学や情報計算のあらゆる技法が活用できる」ということである。現在、研究であれ臨床であれ、計測や分析や診療機器からのデータのデジタル化は当たり前になってきている。今日話題になっている AI の応用などは、大変化の最初の波でしかない。したがって個別の課題よりもっと根源的な課題は、医療行為全体を ICT からどう効果的かつ効率的に支援するかである。

現在のところ ICT や AI が得意とするのは、認知的な課題に関するデジタル化されたデータが容易に集められ、計算の結果をヒトが容易に評価できる問題である。ある種のクイズ、将棋、囲碁、ある種のゲームなどは、その例である。こうした判断問題では、AI はその分野の専門家や熟練者を打ち負かし始めている。

診療の領域でも、画像や波形などの (現在では) デジタル化しやすいデータを扱う課題は、AI 向きである。事実、米国の医療革新の論客であるトポル (E. J. Topol) による最近の調査研究によれば、AI の研究がもっとも盛んなのは放射線診断、眼底写真診断などである*。脳神経系の場合は、PET のような (画像) 診断である。それに続くのは、眼球運動や体に装着した機器 (Mobile/Wearable/Digital Devices) による、時系列的なデータ収

集とその解析である。後者の例としてとくに関心が高いのが痛みや心の障害 **Mental Disorders** である。現在でも、もっとも遅れているのは医師自身が記載するカルテ（診療記録）の情報を基盤とした判断であろう。

さらに言えば、現在の AI、とくに画像解析における成功要因は、多くの研究者が（場合によっては自由に）使えるように、デジタルデータやそれを活用するソフトウェアのツールを整備していることにある。医学的なデータの場合は、患者のプライバシーや権利との関係で、同じようなリソースの整備は簡単には進まないにしても、そのような基盤環境整備の重要性は、すでに広く認識されるようになってきている。

* E. J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nature Medicine*, 25: 44-56, 2019.

* L. A. Celi et al. The PLOS ONE collection on machine learning in health and biomedicine: Towards open code and open data, *PLoS ONE* 14(1): e0210232, 2019.

もうひとつの潮流、簡便な携帯型の生体センサー

医療あるいは、より広く予防や予後対策や介護を含んだヘルスケア領域で、存在感を高めているのがウェアラブルズ（Wearables）である。これは身につけられ（Wearable）、計測結果（データ）をスマートフォンなど無線（Wireless）で、遠隔にある計算施設（Cloud Computing）に送れる仕組みの、最初のセンサー部分を意味する。ウェアラブルズは、モバイル（Mobile）、デジタル（Digital）などとも呼ばれている。それらを活用するのがモバイルヘルス（Mobile Health）、あるいはデジタルヘルス（Digital Health）である。そこにはスマートフォンから入手できるソフトウェア（Apps）の活用も含まれている。

ウェアラブルズの計測対象は脈拍、皮膚温度、身体の位置や姿勢、歩行数その他の活動量などに広がっており、それらのデータを活用するソフトウェア開発も活発になっている。ウェアラブルズに微量な血液や唾液や尿や汗などのサンプル採取を組み合わせると、消費者あるいは生活者が主体的に自らの生体（健康）状態を把握する効果的な手段になる。そのような技術や環境の進歩は、利用者や患者のいるところで診断したり、治療したりするサービス、すなわち POC (Point-of-Care) や POP (Point-of-Procedure) などと呼ばれるサービスへの道につながってくる。インターネットの利用から見れば、それらは IoT (Internet of Things、インターネットへの機器などの接続利用) となる。

そうした機器の普及は、当然、これまで受身であった生活者や患者が、より積極的に自分たちの健康の維持や改善に関与する可能性、すなわち参加型ヘルスケア（Participatory Healthcare）の可能性を広げている。そうした流れとして FDA はこれまでは医療機器と認めなかったウェアラブルズを医療用機器に認定し始めた*。

米国では、もともとゲノムを含めて自分の体を自分で測定して何がしかの対処をしようという気風がある**。ウェアラブルズの進歩と関心の高まりによって、今やそれはより勢

いのある潮流になってきた。それに呼応するように NIH の BD2K プロジェクトでも、とくにウェアラブルズ（モバイル）のためのデータサイエンスの2つの拠点、the Mobilize Center and the Center と Mobile Sensor Data-to-Knowledge (MD2K)を整備している***。

* E. J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nature Medicine*, 25: 44-56, 2019.

** Larry Smarr, Quantifying your body: A how-to guide from a systems biology perspective, *Biotechnology, Journal*, 7, 980-991, 2012.

** E. Topol, *The Patient Will See You Now*, Basic Books, 2015.

*** S. Kumar et al. Center of excellence for mobile sensor Data-to-Knowledge (MD2K), *J Am Med Inform Assoc* 2015; 0:1-6. doi:10.1093/jamia/ocv056

*** A. Gelfand, MOBILE HEALTH BD2K CENTERS HARNESS SENSOR DATA, Mobilize Center, *Biomedical Computation Review*, Summer, 2017

脳神経系疾患領域に特徴的なデータサイエンス

データサイエンスの視点から、脳神経系疾患への取り組みを考えると、いくつかの特徴的な研究領域が目につく。それらは例えば、(1) 脳の解剖学的な知見と各部の機能との関係の解明、(2) ヒトにとっては情報と考えられる生体内での信号伝達の分子的レベルでの実態と、経路 Pathway/Network の探索とその働きの解析、(3) 生体内の情報処理の分子的な機序の解明、(4) 上記の構造や機能の年齢による（発達段階に対応した）変化、男女差、個人差、疾患による変化などの把握、およびそうした変動の原因の解明、などである。

そうした目的のためのデータの収集やそれらの解析の基盤になるのは、解剖学的 (Brain Anatomy あるいは Neuroanatomy) な所見、外科的な措置状況下における応答の電気信号、生化学的な計測、NIRS (Near-Infrared Spectroscopy, 近赤外血流計測)、無侵襲あるいは造影剤を使った X 線撮影、CT (Computer Tomography)、MRI (Magnetic Resonance Spectroscopy, 核磁気共鳴撮像) とその派生技術というべき PET (Positron Emission Tomography, 陽電子消滅を使った断層撮影)、DTI (Diffusion Tensor Imaging, あるいは diffusion MRI, 水分子の動きを捉える撮像法) など、多様な分析計測機器と活用技術である。

それらに付随するのは、このような機器から生成されるデータを、臨床関係者にどのように表現し提示するかという視覚化する (Visualization) 技術である。ただし、そのような技法や技術の多くはそうした機器を開発、製造している企業の課題になっていることが多く、一般に公開され入手できるようにはなってはいないようだ。

これらの多様な計測分析機器を駆使した臨床と研究の目標を簡略に表現すれば、「構造と機能」の解明ということになる。そこではマクロ的な解剖学的 (Brain Anatomy あるいは

Neuroanatomy) な知見から、個別神経細胞 (Neuron) の分子機序に至る、さまざまなレベルでの構造と機能の関連が解明されようとしている。ゲノム解読が進んでからは、構造や組織は生来のゲノムの影響を受け、その結果が機能に影響するという発想が基盤になっている。

これに対して、例えば IT 事業で成功したポール・アレン (Paul G. Allen, 1953-2018) が設立した Allen Brain Institute (<https://alleninstitute.org/>) では、最初はマウスであったが現在ではヒトを含む脳の地図 (Brain Atlas) の作成と公開に挑戦している。それらは 2 次元の画像、3 次元の構造、あるいは 4 次元の動画などである。

ICT の専門家の間でとくに関心が高いのは、脳神経系の構造と機能を理解するために使われる抽象的な概念である。それらは結合性 Connectivity、結合の強さ、ネットワーク、回路、3 次元構造などである。機能としては、部位の活動度を反映した電位、水 (分子) の動き、酸素、血流などが計測の対象になる。この領域では多くの魅力的なモデルが提唱されているが、それを検証したり、臨床に役立てたりすることは極めてむずかしいようだ*。そこでは基本的な統計から始まり、より高次の数学的に表現されたモデルまで、絶えず検証の対象になっている。

* S. M. Smith and T. E. Nichols, Statistical Challenges in “Big Data” Human Neuroimaging, *Neuron* 97: 263-268, 2018

* K. J. Mitchell, Neurogenomics – towards a more rigorous science, *European Journal of Neuroscience*, Vol. 47, pp. 109–114, 2018.

* D. S. Novikov, On modeling, *Magnetic Resonance in Medicine*, 79: 3172-3193, 2018

ウェアラブルズの脳神経系疾患への応用

ウェアラブルズはスポーツやフィットネスの分野で話題を集め、ビジネスとしての関心を高めてきたが、臨床医の関心を惹くような問題には活用されていないと評されていた。だが最近、臨床への興味深い応用領域が 2 つ浮上している。その基盤になっているのは、医療機関以外の場所で、個人の健康に関連した指標を、繰り返し計測できるという利点である。そのことは個人の普段の生活における生理学的な計測項目の値 (Baseline) を推測できる可能性につながる*。そうしたデータからは、生体の本来のリズムや生活様式や近傍 (Ambient) 環境の影響を示す変動が把握されうると期待されている。そうした基準値がわかれば、疾患の予兆や健康あるいは慢性的な疾患の推移も、予測できる可能性がある。その応用としては心房細動、心筋梗塞、脳出血、脳梗塞、癲癇 (てんかん) の発作など、危険な変化を予測することがある。最近注目された事例は、市販の心拍センサーからのデータと深層学習 (Deep Learning) を組み合わせて心房細動の予測に成功したという報告である**。

* X. Li, et al. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLoS Biol* 15(1): 2017.

*J. Dunn, R. Runge and Michael Snyder, Wearables and the medical revolution, Personalized Medicine, 10.2217/pme-2018-0044.

**B. Ballinger et al. DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk, eprint arXiv:1802.02511, 2018.

もうひとつの応用は、慢性的な痛みを伴う疾患や、気分障害、あるいは運動や認知機能の衰えなど、経過を詳しく記録したい患者の状態変化を記録するのに活用することである。そうした記録の収集と解析は医療機関内だけの計測では不十分であり、把握できる患者が生活の場で記録できる簡便な携帯型機器の活用が期待される。その応用例として考えられるのは、繊維筋痛症、気分障害、パーキンソン病などである*。これらの領域では、センサーを開発している企業と大手の ICT 企業とが、臨床研究者の協力をえて医療のイノベーションをめざした競争を展開している。また FDA もそうした機器の認可に前向きである**。

*T. L. Hoskin et al. Longitudinal stability of fibromyalgia symptom clusters, Arthritis Research & Therapy, 20:37, 2018.

*B. Cao et al. DeepMood: Modeling Mobile Phone Typing Dynamics for Mood Detection, eprint arXiv:1803.08986, 2018.

*P. Schwab and W. Karlen, PhoneMD: Learning to Diagnose Parkinson's Disease from Smartphone Data, <https://arxiv.org/abs/1810.01485v2>, 2018.

**J. Dunn, R. Runge and Michael Snyder, Wearables and the medical revolution, Personalized Medicine, 10.2217/pme-2018-0044.

参加型ヘルスケアとベースライン研究

以上簡単にふれた脳の構造と機能を精密にしらべる高額な医療機器の進歩と、個人レベルで使えるウェアラブルズの進歩は、いずれ融合していくと予想される。その先駆的な事例としては、例えば英国バイオバンク UK Biobank を使った大規模なコホート集団をベースにしたゲノム (SNP) データと脳画像の識別を対比したフェノタイプの探索研究*がある。いまや日常生活においてウェアラブルズから産生される膨大な多次元の時系列データと、分子レベルから神経細胞や脳の解剖学的な部位に至る、さまざまなレベルにおけるネットワークに関わる知識とを統合的に活用して疾患を捉え、介入を考えるような研究が展開される時代である。

そこでは IoT を使った介入法 (Digital Interventions) あるいは情報ドラッグ (Info-Drug) が今より (とくに精神疾患治療に) 活用されるとも予想される。

*L. T. Elliott et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank, Nature, 562: 210-215, 2018.

*Z. Dezhina et al. A systematic review of associations between functional MRI activity

and polygenic risk for schizophrenia and bipolar disorder, *Brain Imaging and Behavior*,
Published online, May 10, 2018.

学会やNPOの視点

我が国の現状を見ると、データサイエンスあるいはD2Kサイエンスの重要性は認識されているものの、その担い手としてのD2Kサイエンティストの養成が本格的に始まるには、まだ時間が掛かるように思われる。その間隙を埋められるのは、関連する学会や専門性を有するNPO/NGO的な団体ではなかろうか。肝心なことは、ヘルスケアでいうTranslational Researchの視点に立つことである。この視点は、基礎生物学の研究者にとっては馴染みがないかもしれないが、医学や医療関係者にとっては、共感できる価値観だ。しかし研究者とくにICTの研究者にとって、これは一種の価値転換を受け入れることを意味する。いずれにしても、この視点で考えると挑戦すべき課題が見えてくる。

その始点は、「生物医学の研究が急激に拡大している領域に焦点を合わせ、そこで浮上して来る実践的な課題解決に取り組むことで、データサイエンスの新しい課題を発見できれば、イノベーションもそれに続く可能性がある」という作業仮説である*。新しい人材は、既存の領域の専門家あるいは先達が教えて育てるのではなく、実践課題に取り組む中で自然に育ってくる。新しい挑戦課題や解決策を発想できるか否かは、運に左右されることもある。だが、「データサイエンスとは、これこれの学問であるから、しかじかの予備知識や技法を習得しておくべきだ」、という方針は理想ではあるが、進歩と変化の激しいこの時代においては、現実には間に合わない。こうした実践駆動型、課題駆動型の試みを学会やNPOが先導して提唱、実践することが、現実的ではないだろうか。

そうした取り組みの出発点になるのは、医療系と情報計算系の学会やNPO/NGOとの対話ではないと考える。

*D. C. Mohr, H. Riper, S. M. Schueller, A Solution-Focused Research Approach to Achieve an Implementable Revolution in DigitalMental Health, *JAMA Psychiatry*,
Published online December 13, 2017.

お願い：この資料は、2019年2月19日JSBi支援の研究集会の当日配布による補助的な資料です。時間的な制約で、推敲や校正が不十分であり、引用されうる文献ではありません。ただ、ここで挙げた資料は、ほとんど、ネット（Google Scholar）より無料入手可能です。また、問い合わせ、ご意見などは mail@join-ica.org をお使い下さい。