



HOT TOPICS

NIDA vision for big data science to understand the biological underpinnings of substance use disorders

Susan N. Wright¹ and A. Roger Little¹*Neuropsychopharmacology* (2021) 46:262; <https://doi.org/10.1038/s41386-020-00850-1>

Addiction is a complex disease that involves neuropathological, psychiatric, and environmental determinants. As the primary federal agency supporting scientific research on drug use and its consequences, the mission of the National Institute on Drug Abuse (NIDA) is to advance science on the causes and consequences of drug use and addiction (<https://www.drugabuse.gov/about-nida/strategic-plan/nidas-mission>). NIDA strives to improve individual and public health by both strategically supporting and conducting basic and clinical research on drug use, its consequences, and the various underlying mechanisms involved (<https://www.drugabuse.gov/about-nida/strategic-plan/nidas-mission>).

Like many other fields, biomedical research aimed at understanding how drugs of abuse alter brain biology and function to engender a state of physical dependence and/or promote the compulsive behavior that characterizes addiction is generating substantial data of various types (imaging, genetic, physiological, electronic health records, etc.) The untapped power of data emerging from these studies lies in their integration, mining and analysis (e.g., effectively integrating genome-wide molecular profiling datasets for addiction with other datasets). The interdisciplinary field of data science evolved from the necessity to extract knowledge and insights from increasingly large and/or complex datasets using new quantitative and analytical approaches. Intra/inter-university and multi-disciplinary collaborations are encouraged for data science projects.

NIDA's vision for how big data science can be leveraged to reveal new aspects of addiction biology is closely aligned with the NIH Strategic Plan for Data Science (https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). Specifically, NIDA is focusing on: (1) the integration of existing datasets and tools with those that are being newly developed; (2) making datasets and resources findable, accessible, interoperable, and reusable (FAIR) [1]; (3) ensuring that the algorithms and tools being developed in academia meet industry standards for ease of use and efficiency of operation; (4) data storage and management; and (5) enacting appropriate policies to promote stewardship and sustainability. This approach allows for increased opportunities for secondary data analysis to answer new research questions as well as examine alternative perspectives on the original questions of previous studies, offering cost

and time benefits (e.g., reusing and integrating the Treatment Episode Data Set with other studies).

To carry out these goals, NIDA is continuously creating opportunities for the latest computational capabilities to be combined with biomedical research. The NIDA website lists the latest big data funding opportunities (e.g., NOT-DA-19-041). Big data science is inextricably linked with artificial intelligence, particularly machine learning approaches, and such algorithms have played an important role in recent years in making sense of addiction data. Advancements in supercomputing from the petascale to the exascale levels are helping to address issues of scale, complexity, and processing speed, and are allowing for even more opportunities to analyze and model complex biological systems to increase our understanding of addiction, and quantum computing is on the horizon; collaboration/stewardship are encouraged. With the incorporation of data science as a new tool for the study of substance use disorders comes the need for individuals with expertise in computer science, bioinformatics, and mathematics. The integration of data of many different types will enable scientific discovery with regard to the biological and behavioral complexity underlying addiction.

FUNDING AND DISCLOSURE

The authors have nothing to disclose.

AUTHOR CONTRIBUTIONS

Both authors (SNW and ARL) contributed equally to the design of the NIDA vision for big data science to understand the biological underpinnings of substance use disorders and to the writing and editing of this manuscript.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.

¹Division of Neuroscience and Behavior, National Institute on Drug Abuse, National Institutes of Health, Three White Flint North, Room 08C08 MSC 6018, Bethesda, MD 20892, USA

Correspondence: Susan N. Wright (susan.wright@nih.gov)

Published online: 11 September 2020