Journal of Bioterrorism & Biodefense

DiEuliis et al., J Bioterror Biodef 2018, 9:1
DOI: 10.4172/2157-2526.1000159

Short Communication Open Access

Biodata Risks and Synthetic Biology: A Critical Juncture

Diane DiEuliis[1*,#], Charles D Lutes[1] and James Giordano[2,3,#]

[1]US Department of Defense, Center for the Study of Weapons of Mass Destruction, National Defense University, Fort Lesley J. McNair, Washington, DC, United States
[2]Departments of Neurology and Biochemistry, Neuroethics Studies Program-Pellegrino Center for Clinical Bioethics, Georgetown University Medical Center, Washington, DC, United States
[3]Center for Policy on Emerging Technologies, Washington, DC, United States

[#]Equally Contributed to the Article

Abstract

The tools of synthetic biology and the life sciences are rapidly advancing, as the ability to apply classical engineering to biological systems creates increasing possibilities for innovations in health and medicine, materials science, energy and agriculture. Intrinsic to these capabilities is the mounting 'digitization of biology', as the genetic code and its related metadata (including translated proteins, associated functions, herein referred to as "biodata") are amassed in order to engineer biology for specific purposes. The full spectrum of risks associated with the compilation and use of a wide range of biodata has not been fully identified or comprehensively understood. Further, divergences in traditional attitudes about security among disciplines, namely, biological sciences, engineering, information technology, and data science, complicate discussions on approaches to risk mitigation. To provide a more unified perspective and clarity, we propose that there are unique risks associated with the digitization of biology, represented by overlapping concerns of biosecurity and privacy. We discuss these in three categories of risk: 1) pathogen risks; 2) manufacturing risks, and 3) risks to individual privacy that can allow human harms. Further, we note that there is insufficient address or treatment of these risks in the formulation of ethics, policy and governance. Mitigation of risks will require characterization of all three spheres of risk, acknowledgement that they may require different solutions, and engagement of divergent disciplines and stakeholders to design solutions.

Keywords: Biodata; Biosecurity; Cybersecurity; Privacy; Synthetic biology; Genomic data

Introduction

The expanding repository of genomic data and its associated metadata, such as translated proteins and their functions (herein referred to as "biodata"), is enabling the digitization of biology - and will stimulate innovative development of new chemicals, pharmaceuticals, and biologics, as part of a robust bio economy, both in the United States and worldwide. Increased understanding of genotypes and related phenotypes affords concomitant insight to possibilities for manipulating the genomes of organisms for specific purposes. Although the assumed intent of amassing and using such biodata is toward achieving positive societal benefits, these capabilities also generate a number of risks, if not threats.

Risks associated with pathogen/host biodata

We and others have noted that access to pathogen biodata can facilitate purposeful engineering of modified and/or novel pathogens, thereby expanding the risk of both extant and new biological weapons programs. Pathogens have already been created de novo [1,2]. And the risk of developing such organisms is fortified by recently available, more efficient gene editing technologies [3,4]. These tools enable both engineering of pathogens, as well as more detailed understanding – and perhaps modification – of pathogen/host relationships. Currently, pathogen genomic data exist primarily in the public domain in open databases (e.g. GenBank, GeneCards, BioProject, and GeneLab) [5-8]. Other components of pathogen biodata, and information about the relationship of certain pathogens to the hosts they infect, are found within research papers published in the international scientific literature. In order to employ these data to develop pathogens that could be used for harm, it remains incumbent upon an actor to know what information to access and how such information can be utilized. While access to pathogen biodata may not be a security risk, per se, the compilation of information that could be employed toward developing agents to incur harm(s) has been, and should continue to be acknowledged as enabling possible misuse, particularly given the iterative sophistication and capability of available gene editing technologies.

Risks associated with synthetic biological manufacturing

The biotechnology industry engages synthetic biological manufacturing by applying the categorical "design, build, test" engineering cycle to traditional biology. Manufacturing processes for engineered biological organisms and/or cellular pathways that generate outcome products rely on organisms best suited to engineering, such as yeast or *E. coli*, but increasingly, synthetic biologists are exploring the genomes of organisms known to synthesize interesting or pharmacologically useful compounds. For example, a comprehensive sequencing study of hundreds of fungal genomes is being pursued to identify those that naturally express antibiotics or anti-neoplastic compounds, so that these compounds could then be engineered into yeast [9]. Other examples include the direct engineering of algae or organisms that secrete inorganic or other materials that could have industrial uses [10]. These manipulations of biodata to synthesize pathways require complex informatics and laboratory automation

***Corresponding author:** Diane DiEuliis, US Department of Defense, Center for the Study of Weapons of Mass Destruction, National Defense University, Fort Lesley J. McNair, Washington, DC, United States, Tel: 202-685-4700; E-mail: diane.dieuliis.civ@ndu.edu

James Giordano, Departments of Neurology and Biochemistry and Neuroethics, Studies Program-Pellegrino Center for Clinical Bioethics, Georgetown University Medical Center, Washington, DC, United States, E-mail: jg353@georgetown.edu

Received January 06, 2018; **Accepted** January 31, 2018; **Published** February 07, 2018

Citation: DiEuliis D, Lutes CD, Giordano J (2018) Biodata Risks and Synthetic Biology: A Critical Juncture. J Bioterror Biodef 9: 159. doi: 10.4172/2157-2526.1000159

– which could each (and all) incur vulnerabilities that generate biosecurity risks. Like other industries, the theft of intellectual property for corporate and economic espionage is an inherent risk, and presumably well established mechanisms are in place for protection of such property. However, it should to be noted that many synthetic biology companies are small, startup ventures that may not have the experience and/or resources of larger corporations, which are required to develop and/or apply standard protections against corporate espionage. Furthermore, as biological manufacturing becomes ever more automated, numerous steps in the automation process may be considered to be potential weak points for disruption or diversion of manufacturing so as to generate biosecurity risks or threats. For example, inputting manipulated biodata to an automated system could compromise the safety of outcome products, as well as the facility and the environment (and its populous). Similarly, cyber attacks could render a facility inoperative, which while clearly an economic concern, could also incur security concerns if the manufactured product is important to critical ecology and/or infrastructure. Pecoud et al. [11]. Have recently published a study of these risks, and define "biocybersecurity" as a novel sphere of hazards that surrounds the generation, use, and misuse of biodata. The authors identify novel risks by mapping some of the intricate relationships that exist between computational and experimental workflows in biotechnology. We concur with the observations of Pecoud et al. and advocate for the need to examine if and how such risks can and should be prevented or at least mitigated. As well, we note that compromise or mishaps within these systems could foster societal distrust, and create social instability that would augment disruptive effect(s), and thus harm a burgeoning and necessary industry within the global economic enterprise. We additionally note that some manufacturers outsource information and information-based tasks (to cloud-based laboratory automation) in order to increase efficiency, reliability and reproducibility of research, and to accelerate time to discovery [12]. This too could be problematic, because despite extant controls, legitimate cloud-based online systems remain vulnerable to purloinment due to the increasing sophistication of both state and non-state actors' cyber-infiltrative and -disruptive capabilities. At present there is a relative paucity of encryption or other data safeguards in biological laboratory settings. As noted, many new companies are either start-ups originating from academic institutions (which emphasize cultures of sharing and openness), or ventures created by engineers (and thus may be lacking background or resources in biosecurity), thereby rendering these enterprises susceptible to infiltration. Moreover, manufacturers of analytic and automation equipment that is used in the biotechnology industry may not currently be incentivized to prioritize cyber security processes in their products. Thus, even if the synthetic biology industry becomes fully informed and aware of incumbent risks, products available to mitigate such risks might not be available.

Risks of human harms related to privacy

Last, although certainly not least is a third dimension of risk associated with human biodata: namely, the risk of specific human harms. The ability to identify individuals based on portions of their DNA is possible [13]. Individuals can be identified through health records and physical samples provided to their health practitioners. Of course, regulations and laws are in place to insure privacy of such information (e.g. the Health Insurance Portability and Accountability Act –HIPAA; the Genetic Information Non-discrimination Act, GINA), but the security of hardware, software and cloud ware have been questioned with regard to vulnerability to hacking. Still, the stringency of institutional dedication to cyber security is such that these concerns can be assuaged with relative confidence. However, greater apprehensions are fostered by individuals' increasing use of web-based genomic screening [14] and sharing of unstructured clinical data that are obtained by health- and activity-monitoring devices (e.g.- fitbits or other "wearables"). Specific queries of recreational genetic genealogy databases have also been shown to reveal surname identifiers [15]. Privacy issues may not be fully understood or acknowledged by participant sharers in these open data systems, which have few cyber security controls, and human biodata can be sold to third parties [14,16] for other uses of which participants are not aware. This incurs two domains of risk. First, individual human biodata are crucial to the effective articulation of precision medicine, and as such may afford considerable economic benefit to pharmaceutical and biotechnologic innovation. Economic incentives drive increasing attempts to competitively access biodata by international entities seeking to effect a growing presence, if not dominance, in global biomedical markets. This is evidenced by unprecedented hacks and compromise of certain healthcare entities' biodata [17] and represents a risk of losing market competitiveness in precision medicine. Second, nefarious (dual) use of the tools and techniques of precision medicine could enable creation of "precision maladies" by nations or actors seeking to harm specific individuals or groups. Similarly, as healthcare systems rely more directly on biodata for more precise clinical assessment and personalized therapeutic interventions, purposeful corruption of human biodata could hamper diagnoses and/or prevent effective treatments. Taken together, these new capabilities create situations in which the identification of human biodata can be used to harm both the privacy, as well as the physical health and security of individuals or groups.

The risk domains associated with biodata, as described above, are depicted in Figure 1. In the upper sphere, the diagram depicts the convergence of the biological sciences with information and cyber technology tools to enable the physical engineering of organismal genomes. As biological information and tools become further digitized, risks are emerging and overlapping in security and privacy domains, as depicted in the lower spheres of the diagram. Importantly, ethical and policy approaches have traditionally been relatively siloed within biosecurity and privacy realms, however emerging capabilities in biodata access and use render these separations inaccurate and inadequate. Rather, we call for renewed appreciation and address of risks that now entail both biosecurity and privacy domains as consequential to the digitization of biology (Figure 1).

Current Governance and Recommendations

The operational use of large volumes of genomic data, paired with data from the transcriptome, proteome, metabolome, and other (extensive, multimodal) metadata still requires employment of sophisticated computational bioinformatics and hardware; but this will not always be the case, as technologies and techniques continue to advance. The demands, call, and implements for progress in the life sciences will make biodata acquisition and use more attractive, necessary, and facile [18]. Thus, there is clear and present need to address biosecurity, prior to what may become a watershed of high-risk and/or threatening developments and applications. Given that biodata are information, their misuse to affect biosecurity has not been fully apprehended by extant guidelines and policies of dual-use. Institutional policy addressing Dual Use Research of Concern (DURC) [19]. Includes "knowledge and information" in its definition, suggesting categorical inclusion of biodata. However, the scope of the policy is limited to select agents (and a few other pathogens), which would therefore not entail regulation of novel organisms or substances

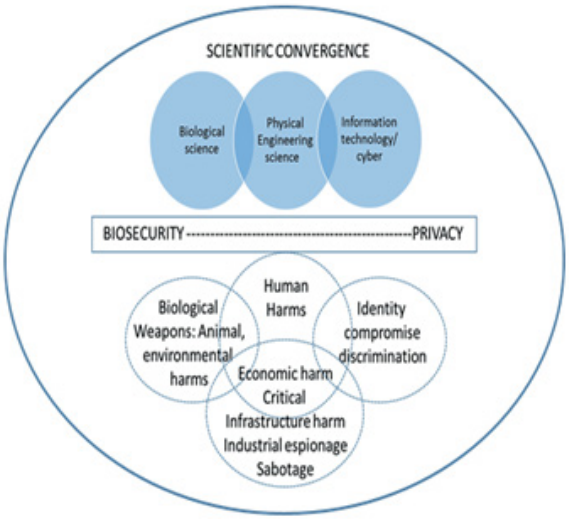

Figure 1: Complexity of converging scientific disciplines and spheres of risks along a continuum of biosecurity and privacy.

created by synthetic biology. Of note is that the "Companion Guide" [20]. Provides instructions for implementing the DURC policy in practice, and offers useful suggestions for managing sensitive data, thus affording a template for development of a similar framework or decision-tree that could be used to assess biodata risks. But while useful, these guidelines do not afford governance of biodata utilized by synthetic biology manufacturing companies. To the extent that companies use commercial providers of synthetic DNA, most of those providers screen orders for their similarity to dangerous pathogens [21]. However, this is voluntary guidance, and while providing useful constructs and functions, is not without limitations – and defined weaknesses [22]. Protection of human subjects (in the United States) is guided and governed by The Common Rule [23] and HIPAA [24]. As per the most recent update, the Rule does not mandate any kind of encryption or protection of human genomic data collected during research; nor does it regulate data technologies that identify patients. HIPAA regulation affords privacy for individuals across 18 different characteristics considered to be personally identifiable information (PII). However, human genomic data are not included. Furthermore, while HIPAA enforces privacy within the US, such requirements no longer apply once patient files leave national custody. This may incur privacy risks if/as countries create partnerships with US-based institutions and health records are transferred beyond US borders – and US regulatory oversight and control. As noted, common cyber security tools have not been rigorously applied, or may be insufficient, to safeguard all types of biodata in the contexts described herein. But this also establishes opportunities for both innovation in cyber tools that can be used for different types of biodata protections, and the adaptation of such tools for use in academic and commercial settings on a global scale. To be sure, current encryption methods may provide some protections; for example, fragmented encryption could be applied during genomic sequencing, and continued when transferring data. Other options may be to leave raw genomic data unencrypted and only encrypt associated metadata, or to adopt guidelines to ensure that no single entity simultaneously possesses all components of a type of biodata.

Moreover, we have argued, and re-assert here that emerging biotechnologies– such as biodata systems - demand equivalent (if not equal) dedication to development of ethically-informed policies to guide and govern their use(s)-in-practice [25,26]. Key ethico-policy questions include: Should unrestricted access to, and/or "tampering" with biological data be regulated by international law? If so, what mechanism(s) must be in place to sustain global enforcement? Should the Biological and Toxins Weapons Convention (BTWC) consider biodata compromise as potentially contributory to bioweapon development? If so, how should the BTWC be revised to reflect those methods and contexts of use that fall under this purview [27]. Because the types of biodata described here are used in different sectors for varying purposes (and can be digitally available on a global scale), mitigation of risks poses challenges, and it is unlikely that a single solution will be effective. Instead, we propose a number of approaches that if taken in combination, may be viable and of value. Important to this stance is recognition of the disparate experience(s), cultures and perspectives of the bioscience and cyber-science communities, which to date may have impeded addressing such questions with the granularity adequate and necessary for risk mitigation solutions. Thus, while existing cyber solutions may be employed to address such risk(s), these tend to be under-utilized due to a lack of multi- and inter-disciplinary discourses that would be required to foster additional innovative approaches to resolve specific issues with greater precision [28]. We also posit the need for further dialogue – and policies—addressing dual-use that highlight information as a dual-use entity, and which develop models for realistic assessment of benefit(s), risk(s), and threat(s) in key biodata domains. Best practices for other types of data are being developed internationally (e.g. norms proposed for financial data) [29]. Which could provide a framework, and a number of governance structures would be useful if biodata were more widely recognized as PII (e.g. - perhaps if/when considered as a specialized case of consumer protection). Simply put, information is knowledge, and knowledge is power that can be leveraged toward benevolent and /or malevolent purposes. Availability of biodata may enable state and non-state actors to acquire information to develop novel pathogens and other substances that are currently not identified or registered on international regulations and treaties governing biological and chemical weapon, and/or to manipulate PII for nefarious purposes (e.g. - to misinform health records, disrupt health care services and subsidies, etc.,). We assert that given the multi-national development and global utility of biodata, international discussion of novel risks associated with dual-use implications of biodata are, and will be increasingly necessary. Both domestic and international standards for the use of biodata should be developed, with acknowledgement of the potential for such data to be utilized in the creation of biological weapons. These possibilities could be highlighted in existing venues (e.g. BTWC Review Conferences; World Health Organization and United Nations' meetings) for ongoing international deliberations.

Conclusion

The current momentum of 'big data' approaches in the biosciences to enable research, manufacturing, and clinical enterprises and outcomes prompts the development of biodata methods, capabilities, and applications. This confers ever greater power to modify and/or create new organisms, expand biological manufacturing platforms, and provide therapeutics for a variety of human diseases. These capabilities can also compromise the biosecurity of individuals, groups and populations, which elicits our queries of how such risks should be assessed and mitigated, and underscores our call for more thorough recognition of risk, and development of regulations to guide and govern trajectories of use and misuse.

Acknowledgements

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 720270: HBP SGA1 (JG); by federal grant UL1TR001409 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through the Clinical and Translational Science Awards Program (CTSA), a trademark of the Department of Health and Human Services, part of the Roadmap Initiative, "Re-Engineering the Clinical Research Enterprise" (JG); from a grant by the AEHS Foundation, as part of the Neuro-HOPE Project (JG), and by funding from the Austin and Ann O'Malley Visiting Chair in Bioethics of Loyola Marymount University (JG). The views expressed in this article are those of the authors and do not reflect the official policy or position of the National Defense University, the Department of Defense or the U. S. Government.

References

1. Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. Science 297: 1016-1018.
2. Tumpey TM, Basler CF, Aguilar PV, Zeng H, Solorzano A, et al. (2005) Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310: 77-80.
3. Hearn A (2017) There are things worse than death: can a cancer cure lead to brutal bioweapons? The Guardian.
4. Pope SM (2017) Impact of gene editing tools like CRISPR/Cas9, on the public health response to disease outbreaks. Disaster Med Public Health Prep 11: 155-159.
5. https://www.ncbi.nlm.nih.gov/genbank/
6. Clark K, Pruitt K, Tatusova T, Mizrachi I (2013) BioProject. The NCBI Handbook [Internet].
7. http://www.genecards.org/
8. https://genelab.nasa.gov/
9. Talbot JM, Bruns TD, Taylor JW, Smith DP, Branco S, et al. (2014) Endemism and convergence across the mycobiome. PNAS 111: 6341-6346.
10. Tana Y, Adhikarib RY, Malvankarb NS, Warda JE, Woodarda TL, et al. (2017) Expressing the Geobacter metallireducens PilA in Geobacter sulfurreducens yields pili with exceptional conductivity. mBio 8: e02203-2216.
11. Peccoud J, Gallegos JE, Murch R, Buchholz WG, Raman S (2017) Cyberbiosecurity: From naive trust to risk awareness. Trends Biotechnol 36: 4-7.
12. Hayden EC (2014) The automated lab. Nature 516: 131-132.
13. Lippert C, Sabatini R, Maher MC, Kang EY, Lee S, et al. (2017) Identification of individuals by trait prediction using whole-genome sequencing data. PNAS 114: 10166-10171.
14. https://www.technologyreview.com/s/601506/23andme-sells-data-for-drug-search/
15. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) Identifying personal genomes by surname inference. Science 339: 321-324.
16. Brogan J (May 2017) Who owns your genetic data after a home DNA test?
17. Arndt RZ (2017) Healthcare data breaches caused by hacks are on the rise. Modern Healthcare.
18. Appleton E, Densmore D, Madsen C, Roehner N (2017) Needs and opportunities in bio-design automation: four areas of focus. Curr Opinion Chem Biol 40: 111-118.
19. https://www.phe.gov/s3/dualuse/Pages/InstitutionalOversight.aspx
20. https://www.phe.gov/s3/dualuse/Documents/durc-companion-guide.pdf
21. https://genesynthesisconsortium.org/wp-content/uploads/IGSCHarmonizedProtocol11-21-17.pdf
22. DiEuliis D, Carter SR, Gronvall GK (2017) Options for synthetic DNA order screening, revisited. mSphere 2: e00319-17.
23. https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996
24. Giordano J (2017) Toward an operational neuroethical risk analysis and mitigation paradigm for emerging neuroscience and technology (neuroS/T). Exp Neurol 287: 492-495.
25. DiEuliis D, Giordano J (2016) Neurotechnological convergence and "Big Data": A force-multiplier toward advancing neuroscience. In: Collmann J, Matei S (eds.), Ethical Reasoning in Big Data. Computational Social Sciences. Springer, Switzerland.
26. Gerstein D, Giordano J (2017) Rethinking the biological and toxins weapons convention? Health Security 15: 638-641.
27. Giordano J (2012) Integrative convergence in neuroscience: Trajectories, problems and the need for a progressive neurobioethics. In: Vaseashta A, Braman E, Susmann P (eds.). Technological Innovation in Sensing and Detecting Chemical, Biological, Radiological, Nuclear Threats and Ecological Terrorism. NATO Science for Peace and Security Series A: Chemistry and Biology. Springer, Dordrecht. pp: 115-130.
28. https://ccdcoe.org/research.html
29. Schmitt M, Maurer T (2017) Protecting financial data in cyberspace: Precedent for further progress on cyber norms? Just Security.