

**Biological data are the foundation of biotechnology innovation.** Artificial intelligence and machine learning (AI/ML) are being used to extract useful insights from biological datasets to better design and use biotechnology products. However, strategic competitors and military adversaries could also use biological data in ways that threaten U.S. national and economic security. This White Paper is the first in a series about biological data and its importance to biotechnology and national security.

---

## What are biological data?

Biological data describe living systems. They include information about whole organisms, from humans and communities of complex organisms, down to cells and microbes, and the molecules that comprise them. These data can describe the internal components of organisms, such as DNA and proteins, and how these components work together. Biological data can also describe an organism's traits, how the organism interacts with their surroundings, and how they interact with each other.

---

## Biological data for the American economy

Within the biotechnology sector, certain types of biological data are more frequently used than others. These data types, including DNA, RNA and proteins as illustrated in the figure on the next page, inform how living systems can be used for biotechnology applications. These data can be collected from the environment, whole organisms, tissues, cells, and many other places in order to increase a researcher's knowledge about biology.

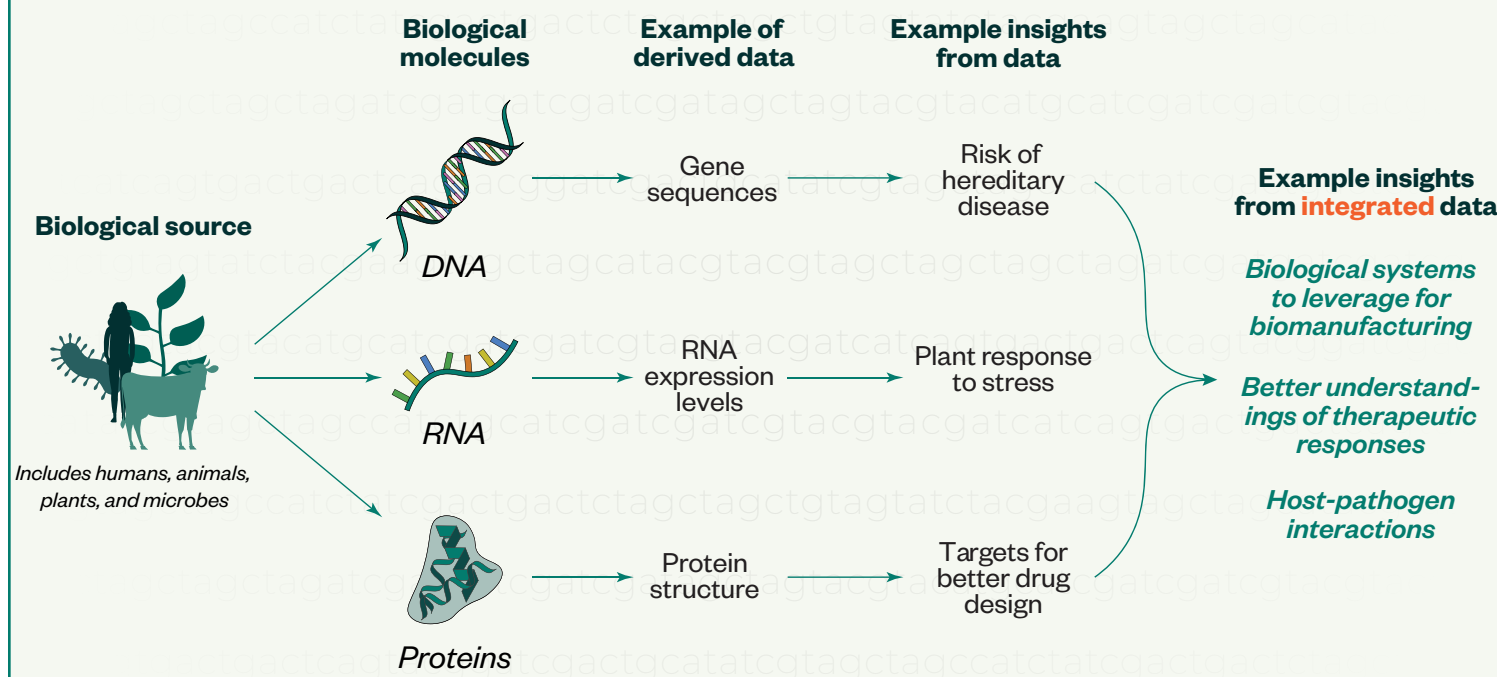
Biological data are important because they are used as the blueprint for biotechnology. As biological datasets expand and are increasingly made useful by AI/ML, they can tell researchers the necessary information to design, engineer, build, and manufacture entirely new products and capabilities based on biology. For example, researchers can use biological data:

1. **To make products:** Biological data can be used to engineer organisms, like yeast and bacteria, to produce chemicals or materials faster and more efficiently (biomanufacturing).<sup>1</sup>
2. **To design specific functional capabilities:** Biological data help inform the creation of engineered plants that can fend off pests without the use of pesticides, or plants that are more resilient to different weather conditions.<sup>2</sup>
3. **To improve health:** Researchers and doctors can use biological data to better understand diseases and develop more targeted cell and gene therapies.

*“In the age of advanced computing, having access to biological data provides a critical strategic advantage. The U.S. must make use of and maintain the data it has, incorporate new data into existing systems, and design systems with adaptability for future uses in mind.”*

— Commissioner Alexander Titus

## From biological data to integrated findings



### Biological data and potential misuses

Researchers need substantial amounts of high-quality biological data to continue innovating in biotechnology, especially when applying artificial intelligence capabilities.<sup>3</sup> In order for the U.S. to continue leading the world in biotechnology innovation, there is a critical need to continue generating and assembling biological data across the public, academic, and private sectors. However, these data tend to be siloed, so the size and utility of these datasets are limited. The ability to generate more useful data that can be better accessed and securely shared would support broader applications of biotechnology. **White Paper #2** in this series will discuss data generation and standardization to promote the creation of biological data as a strategic asset.

Certain types of biological data are essential for research and development but could also be intentionally misused to harm the United States and its interests. For example, biological data that describe pathogens are important for basic biological research and can be leveraged to improve health, but the same data could potentially be used to engineer more harmful versions of pathogens than those that naturally occur. Additionally, when biological data describe a specific group of people, the data can be used to produce more precise medical treatments, but also to blackmail or exploit said groups. Concerningly, the Chinese Communist

Party continues to collect, aggregate, and store biological data in a government-funded Chinese National GeneBank<sup>4</sup> and controls access to and use of these data.<sup>5</sup> **White Paper #3** will discuss measures related to protecting biological data from being misused.

The Federal Government could enact policies to promote the generation of biological data and to codify the measures required to maximize the value and use of biological data within all sectors. Additionally, legislation could be enacted to protect U.S. biological data assets from being misused. **White Paper #4** will provide policy ideas to both promote and protect U.S. biological data.

### Sources

- 1 BCG. "[How Synthetic Biology can make a Materials Difference](#)"
- 2 Biology (Basel). "[Biotechnology Engineered Plants](#)"
- 3 Congressional Research Service. "[Artificial Intelligence in the Biological Sciences: Uses, Safety, Security, and Oversight](#)"
- 4 Bo Wang et al. "[\[The China National GeneBank-owned by all, completed by all and shared by all\]](#)"
- 5 Ropes&Gray. "[China Unveils Implementing Rules on the Administration of Human Genetic Resources](#)"

For any questions about this white paper, or related work at the National Security Commission on Emerging Biotechnology, please contact us at [ideas@biotech.senate.gov](mailto:ideas@biotech.senate.gov).

Staff at the National Security Commission on Emerging Biotechnology authored this paper with input from the expert Commissioners. The content and recommendations of this white paper do not necessarily represent positions officially adopted by the Commission.

