



ISSUE BRIEF



Expanding Access to Data Science Careers

Panel 1: The Data Science Workforce Challenge

Executive Summary. Life sciences employers need a workforce that can use large data sets and significant data analysis to solve complex challenges. Yet in Massachusetts and nationwide, these employers are experiencing critical shortages of qualified data scientists. Within the field of life sciences, the bioscience industry—a branch of life sciences that focuses on developing biological solutions to address pressing issues—is facing a particularly acute shortage.

In this brief, we share highlights from the first in a series of expert panels co-led by Education Development Center, Mass. Life Sciences Center, and Mass. Biotechnology Education Foundation. During the panel, four life sciences leaders weighed in on the critical skills and knowledge that data science workers need to succeed in their industry. They agreed that aspiring data scientists must have strong computational thinking skills, especially an understanding of the underlying logic of programming. They also concurred that prospective candidates need to possess a strong set of non-technological skills—including critical thinking, problem-solving, communication, and self-motivation. As this field grows, there will be increasing opportunities for young people to pursue data science careers in life sciences. However, we believe a more intentional integration of bioscience and computer science curricula—as well as stronger programs to increase awareness of such careers—is crucial to growing and diversifying the field and addressing workforce shortages.

Key Takeaways:



What Is the Role of a Data Scientist in the Life Sciences?

- In the life sciences, data science is the union of epidemiology, informatics, omics, statistics, and computer science—all analytical and quantitative domains. It is a broad umbrella term, ranging from data collection and examination to statistical analysis and visualization of the data in ways that engage and inform key audiences.
- Titles of data scientists in the life sciences include junior roles, such as bioinformatics assistant or computational associate, and more senior roles such as computational biologist or computational chemist.
- Data scientists identify, collect, organize, clean, annotate, inspect, visualize, test, and manage key datasets to research a biological question.
- A data scientist typically extracts actionable insights and knowledge out of complex data sets. Working with a biologist, the data scientist learns how the data was produced, what it is measuring, and how the research questions can be explored with the data.
- Data scientists are communicators. They must be adept at listening to their colleagues to understand the phenomena they are studying, working collaboratively in a team, and being able to explain their strategies and processes to their team and external audiences.



What Kinds of Education and Experience Do Data Scientists Typically Have?

- Data scientists who are researchers or computational biologists require graduate degrees. As teams grow, there are more opportunities for specialized roles that may not require such advanced degrees.

- Many positions do not require a PhD. Positions that focus on data analysis and interpretation require a strong math or statistics background. For positions that require organizing, combining, and comparing data sets, programming skills are key. In the life sciences industry, individuals with a solid understanding of biology tend to be more successful than pure computer scientists, mathematicians, or statisticians.
- Junior positions are available in some companies for undergraduate biology students. Internship opportunities exist for college and high school students who have taken biology courses and have done coursework or self-learning on data science.
- Communication skills are vital. Data scientists must understand their audience and what they need to take away from the data, produce effective data visualizations, and write compelling narratives that tell a clear and coherent 'story' as borne out by the data.
- Few candidates will "check all the boxes." Opportunities exist for candidates from a variety of backgrounds to land jobs.



What Do Biotechnology Companies Look for When Hiring Data Scientists?

- A strong background in biology is important. Not all data scientists are biologists, but data scientists who understand the biological questions they are trying to answer make more informed choices when analyzing data sets.
- A core set of critical thinking skills, effective communications skills, and intellectual curiosity are essential. Employers often insist that these traits are paramount—specific skill sets can be learned on the job, but employees with these inherent traits are teachable, and can succeed in a range of environments.

- Some experience coding is necessary, but which language a candidate has experience in is less relevant. Understanding logic and being able to repurpose existing code and troubleshoot problems are highly desirable.



What Are Some Key Elements of a Strong Pathway Toward a Career as a Data Scientist in the Life Sciences?

- An early-stage pathway is essential, offering younger students the opportunity to develop relevant skills in data competency, basic algorithmic logic, and problem solving.
- Ideally, students will engage in programs that offer real-world and real-life experiences, especially with larger data sets, using a problem-based approach.
- For high school and college students, internship programs, such as those offered by the Mass. Life Sciences Center, engage students in unique and valuable explorations of what a career in data science looks like.
- Efforts to broaden participation should seek to spark the interest of students from groups that are under-represented in the current workforce, and make it clear how a career in the life sciences is important and can help them make a difference in the world.

- Greater efforts are needed to recruit, retain, and support more diverse candidates from earlier grades (middle and high school) through undergraduate studies and into graduate programs. This includes young women, young people of color, young people from low socioeconomic status communities, and others underrepresented in the workforce.



What Are Some Key Technologies that Data Scientists Use?

- The panelists emphasized that data scientists in the life sciences do not need to know one specific programming language. Instead, they need to have experience with any language (such as R or Python) and understand the underlying logic of programming.
- Many data scientists in the life sciences and other fields use open-source tools. For example, Stack Overflow—a free “question and answer site for professional and enthusiast programmers”—is a favorite resource.
- The panelists expressed enthusiasm for ggplot, an open-source package for data visualizations. Tableau is a commercial product that has grown in popularity. All panelists underscored the importance of data scientists in the life sciences being able to use data visualization tools to clearly convey findings.



Conclusion

This first panel in the 4-part series focused on the shortage of skilled data science workers in the life sciences industry and examined the competencies workers must have to succeed. Panelists discussed the needs of the bioscience industry for workers with data science and life sciences knowledge, as well as the skills and aptitudes required for data science jobs in the life sciences. The following implications arose from the discussion:

- Guidance counselors can play a pivotal role in helping steer high school students to consider further studies and careers in this field.
- Supporting K-12 students in building strong data literacy, critical thinking, and communication skills can give them an early head start to data science careers.
- Experiential learning is key: “Real-world” experience working with large and complicated data sets is key to preparing students of all ages for data science careers.
- Technology changes rapidly. The specific language students learn, especially in high school, is not as important as the experience they gain with critical thinking and coding.
- Life sciences employers are prepared to train new data science workers in the specific tools and technologies they will need to use.

The next panel in the series will focus on data science education policy support and how it can help expand access to data science careers.

Panelists

A special thank you to our panelists for their time and thoughtful contributions to the discussion.

- **Sudeshna Das**, Director, MGH Biomedical Informatics Core, Harvard Medical School, Massachusetts General Hospital
- **Shuba Gopal**, Data Scientist, Broad Institute
- **Iain McFadyen**, Chief Data Officer, LifeMine Therapeutics
- **Sally Trabucco**, Data Scientist, Foundation Medicine
- *Moderator: Karla Talanian*, Director of Talent and Workforce Development, Massachusetts Biotechnology Education Foundation

Education Development Center (EDC) is a global nonprofit that advances lasting solutions to improve education, promote health, and expand economic opportunity. Since 1958, we have been a leader in designing, implementing, and evaluating powerful and innovative programs in more than 80 countries around the world.

