

Facilitating Biomedical Data Science – *what different actors in the biomedical research enterprise can do to help attract data scientists to biomedical data and issues:*

For repositories and data providers promoting data use and reuse

- Provide raw data in a non-proprietary format (such as comma separated value or plain text)
- Foster trust in a given dataset by giving information indicating that data are complete, clean, and responsibly collected
- Include specific use case examples for datasets to provide context for potential users who are not experts in biomedicine
- Establish benchmarks on datasets using standard tools to enable data scientists to compare the performance of their tools

For subject matter experts seeking collaborations with data scientists

- Ensure collaborations between data scientists and biomedical subject matter experts are interactive and rooted in effective communication around the limitations and opportunities of a given dataset
- Set expectations for a successful collaboration (such as team members' roles and responsibilities, where the final research product will be published or presented, etc.) in advance to minimize conflict and ensure collaborators are on the same page
- Hold back an adequate amount of data from analyses to validate data science models and conclusions

For trainers and educational institutions

- Define and develop core skills required of a data scientist, ensuring that they are sufficiently broad to allow aspiring data scientists to choose a training path tailored to their needs
- Integrate ethics and responsible conduct of research early in data science training to ensure these principles are integrated into data scientists' practices
- Promote equity, diversity, and inclusion to ensure data science has maximal impact for the greatest public good

For NIH

- Facilitate awareness of opportunities in biomedical data science and NIH interest in data science through, for instance, a “Data Science @ NIH” web portal that describes data science activities, interests, and opportunities across all of NIH and its institutes, centers, and offices (for one-stop shopping for those who do not understand NIH organization)
- Create and communicate opportunities for joint leadership of data science research and training that allow data scientists and subject matter experts to serve as equal partners (e.g., as principal investigators on the same grant)
- Create multiple pathways for discovering funding opportunities at NIH, particularly for data scientists unfamiliar with NIH (e.g., including, but not limited to, pointing to data science-relevant funding opportunity announcements on a Data Science @ NIH portal)
- Create opportunities for scientists from non-traditional backgrounds, such as those from industry or those who do not hold a doctoral degree to explore biomedical data – aim beyond biomedical academic audiences