

The National Microbiome Data Collaborative community survey on microbiome research data

Pajau Vangay^{1*} (pvangay@lbl.gov), Faiza Ahmed², Anubhav³, Jeffrey Baumes², Jonathan Beezley², Mark Borkum³, Lisa Bramer³, Shane Canon¹, Patrick Chain⁴, Danielle Christianson¹, Yuri Corilo³, Karen Davenport⁴, Brandon Davis², Meghan Drake⁵, William Duncan¹, Kjersten Fagnan¹, Mark Flynn⁴, David Hays¹, Bin Hu⁴, Marcel Huntemann¹, Julia Kelliher⁴, Sonya Lebedeva¹, Po-E Li⁴, Mary Lipton³, Chien-Chi Lo⁴, Douglas Mans³, Stanton Martin⁵, Lee Ann McCue³, David Millard³, Kayd Miller¹, Nigel Mouncey¹, Chris Mungall¹, Paul Piehowski³, Elais Player Jackson⁴, Anastasiya Prymolenna³, Samuel Purvine³, TBK Reddy¹, Rachel Richardson³, Migun Shakya⁴, Montana Smith³, Jagadish Chandrabose Sundaramurthi¹, Deepak Unni¹, Bruce Wilson⁵, Donny Winston⁶, Elisha Wood-Charlson¹, Yan Xu⁴, **Emiley Eloe-Fadrosh¹**

¹ Lawrence Berkeley National Laboratory, Berkeley, CA; ² Kitware, Clifton Park, NY; ³ Pacific Northwest National Laboratory, Richland, WA; ⁴ Los Alamos National Laboratory, Los Alamos, NM; ⁵ Oak Ridge National Laboratory, Oak Ridge, TN; ⁶ Polyneme LLC, New York, NY

<https://microbiomedata.org/>

Project Goals: The National Microbiome Data Collaborative (NMDC) is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an open-source, integrated data science ecosystem that leverages existing data standards, data resources, and infrastructure within the DOE complex.

Abstract

The NMDC is building a science gateway to enable access to multidisciplinary microbiome data and standardized, reproducible data products by leveraging unique capabilities, expertise, and resources available across Lawrence Berkeley National Laboratory (LBNL), Los Alamos National Laboratory (LANL), Pacific Northwest National Laboratory (PNNL) and Oak Ridge National Laboratory (ORNL). To address the data challenges of the microbiome research community, the NMDC team is using a community-centered design approach, which involves seeking feedback from the community throughout its phases of iterative development. The NMDC team is also applying best practices in user experience design through consultation with the Science Gateways Community Institute (SGCI), an NSF-funded initiative to provide services, resources, community support, and education for creating and sustaining science gateways.

To understand the needs of the microbiome research community, the NMDC team conducted a community survey during December 2020 to January 2021 on how researchers work with data,

access and share data, and use existing resources. The anonymous online survey was distributed broadly with NMDC partner groups, and redistributed within the community.

We present preliminary results from n=768 survey respondents and share a preliminary qualitative analysis by SGCI. We also present several core user groups that emerged from the survey data and briefly discuss how these results will inform the development of the NMDC science gateway.

This survey is an important first step toward understanding how the microbiome research community works with data, and toward establishing an inclusive, bidirectional communication channel between the NMDC and the broader community.

Funding statement

This work is supported by the Genomic Science Program in the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC05-76RL01830 (PNNL).