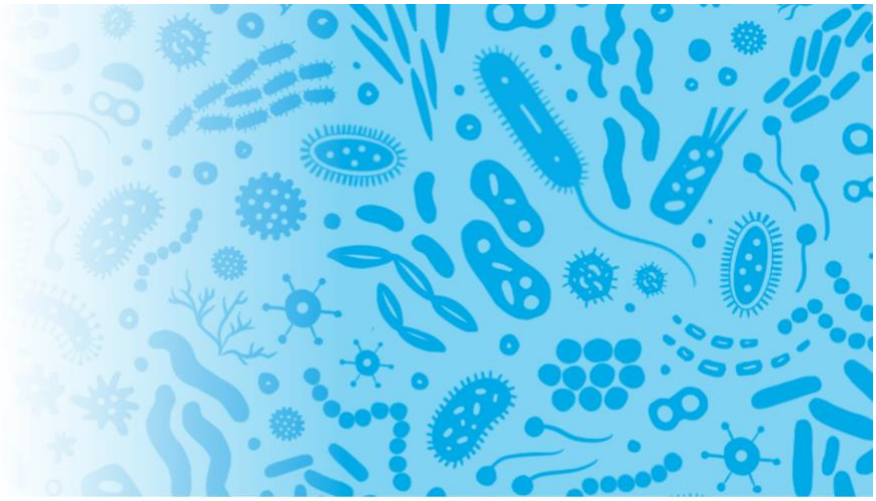

DOE BSSD Performance Management Metrics Report Q2

March 28, 2024

Investigators: Emiley A. Eloe-Fadrosh[1] (Lead, eaeloefadrosh@lbl.gov), Patrick S. G. Chain[2], Shreyas Cholia[1], Kjiersten Fagnan[1], Douglas Mans[3], Lee Ann McCue[3], Christopher J. Mungall[1], Nigel J. Mouncey[1]

Participating Institutions: [1]Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [2]Los Alamos National Laboratory, Los Alamos, NM 87545; [3]Pacific Northwest National Laboratory, Richland, WA 99354.

Deliverable Q2: Describe the mechanisms developed to engage the broader community to assist in the development of a community-based resource for microbiome research.

Executive Summary

The vision of the National Microbiome Data Collaborative (NMDC) centers on the concept of ***connecting data, people, and ideas*** to advance microbiome innovation and discovery. Building data infrastructure, while key to NMDC's ability to execute on our vision, can only go so far in creating scientific impact. By fostering strong community partnerships and developing a set of robust community outreach and training programs, we are able to turn our products – the Submission Portal, NMDC EDGE, and the Data Portal – into tools that empower the scientific community. Our multi-pronged community building approach spans individual researchers, research teams, consortia and scientific societies, and institutions and federal agencies.

To foster a collaborative and inclusive community-centered environment, we have identified three strategic objectives to promote an inclusive and connected community: (1) recognize and support the diverse research needs and perspectives of the microbiome research community; (2) promote best practices across the microbiome community, from researchers to funders, through community-driven practices (FAIR, CARE, and TRUST); and (3) build a microbiome ecosystem that enables scientific discovery and innovation across stakeholders. These strategic objectives allow our team to focus on impact across a diverse range of activities, from launching the American Society for Microbiology (ASM) Microbiome Data Prize to supporting the Ambassador and Champions programs fostering learning and building a collaborative network. We broadly communicate our work through social media (X/Twitter, LinkedIn, and Instagram), *The Microbiome Standard* (our quarterly newsletter), and Annual Reports. All our work is underpinned by a strong commitment to diversity, equity, and inclusion as articulated in our Action Plan that tracks progress towards key metrics.

A core component of our engagement strategy is user research. User research ensures the Submission Portal, NMDC EDGE, Data Portal, and the new Field Notes mobile app are designed with and for the scientific community. Our user research efforts consist of asking researchers exploratory questions to collect information on researcher priorities, methodologies, and perceptions to ensure that we are aware of the current state of microbiome research. Our usability testing provides researchers with prototypes or test environments of the NMDC products, and we capture valuable information on how users interact with the products to make improvements. Given the diverse nature of microbiome work, we acknowledge that we are not aware of all pressing data challenges and thus rely on the research community to help us identify the most important issues to prioritize. To date, we have conducted 24 interviews and one beta-testing call with 10 participants across all NMDC products, which have generated 321 insights and 120 action items. Herein, we describe the ways we engage with the microbiome research community to advance the NMDC mission.

Advancing microbiome science, together

The tools and resources developed by our team aim to advance how scientists create, use, and reuse data in a collaborative and inclusive manner. The research community is deeply involved in the creation and execution of all aspects of the NMDC. We engage partners across research teams, organizations, federal agencies, scientific societies, and the international sphere. We have formed strong partnerships with complementary data resources like DOE's Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) and DOE's Systems Biology Knowledgebase (KBase), along with working closely with DOE User Facilities, the Joint Genome Institute (JGI) and the Environmental Molecular Sciences Laboratory (EMSL). Working across existing user communities and DOE research groups, we have identified many of the existing "pain points" for data sharing and interoperability. Through our User Research program, which was formally launched in June 2023, we focus on understanding the needs of the research community to support BER's bioenergy and environmental research goals.

To advance microbiome research in a meaningful and sustainable way, we outlined an extensive outreach, engagement, and communications strategy that prioritizes collaborative efforts. Our engagement strategy includes information on how we can best communicate, collaborate with, and share the NMDC products with microbiome researchers regardless of career stage, expertise, resource availability, geographic location, background, and institution. This past year, we outlined ten user personas as a means of better recognizing and understanding the diverse needs and goals of microbiome researchers so that the NMDC products can better support their work.

User research underpins infrastructure development

We apply user-centered design methodology to collect feedback from the research community in an iterative manner for ongoing improvements. Early and continuous implementation of feedback ensures that the Submission Portal, NMDC EDGE, the Data Portal, and the Field Notes app evolve with the changing needs of our diverse user community. For example, in September 2023 we performed a beta-testing call for NMDC EDGE. Aligned with our user-centered design process (**Figure 1**), our team incorporated two new workflows, 'Viruses and Plasmids (GeNomad [1])' and 'Metaproteomics.' Our beta-testing call focused on usability of these two workflows in NMDC EDGE, asking for feedback on their experience with data upload and download, ease of use, feature requests for future development, and utility of the training materials and user guides. From the collated feedback, we generated 40 action items to improve how users can download data, run workflows, and view their results. These action items were discussed internally by our team to assess feasibility and determine prioritization within NMDC EDGE's product development roadmap. As a direct result of this process, we implemented support for batch processing of data, graphs to visualize processed data compared to tabular outputs, and a new data import mechanism using the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) accession numbers to directly process publicly available data without the need to download and re-upload.

To ensure a positive feedback loop for our user-centered design, we recognize participants through an ORCiD professional service acknowledgement analogous to peer review service recognition. This enables all contributors to have their work officially recognized using their ORCiD researcher profile. We also list acknowledgements for interviewees and beta-testers on the NMDC website.

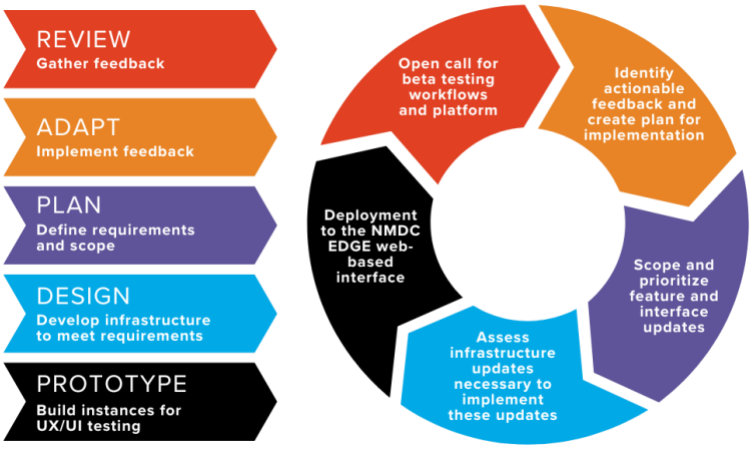

Figure 1. An example of the NMDC's user-centered design process for iterative development of NMDC EDGE through beta testing. The open call collected feedback from 10 respondents and generated 40 action items to improve how users can upload and download data, run workflows, and view their results.

Workshops, presentations, and events generate valuable real-time feedback

We prioritize conference and workshop attendance to promote the NMDC Program while collecting valuable live feedback from attendees. While the pandemic disrupted in-person participation, from 2021 onwards, our team collectively presented at over 75 conferences and events, reaching over 5,000 researchers (**Figure 2**). We host and contribute to workshops, panel discussions, seminars, webinars, special sessions, and presentations at many of the major microbial sciences and bioinformatics conferences including the annual ASM Microbe meeting, the Genomic Standards Consortium (GSC) meeting, Sequencing to Function: Analysis and Application for the Future (SFA²F), and the International Symposium on Microbial Ecology (ISME) meeting.

Figure 2. Examples of conference participation. (Left) ASM Microbe 2023 with CEO Stefano Bertuzzi, NMDC Program Lead Emiley Eloe-Fadrosh, ASM Microbiome Data Prize Awardee Benjamin Callahan, and NMDC Ambassadors Reid Longley, Ishi Keenum, and Saraí Finks. (Right) SFA²F with NMDC Engagement Lead Julia Kelliher and Ambassador Reid Longley.

A focus on documentation and learning materials

To promote accessibility and understandability of the NMDC products, we continuously develop technical documentation, user guides, video tutorials, and reference material for a broad range of experience levels. This documentation is organized according to the Diátaxis framework and made available as a single documentation site. Researchers who participate in NMDC's user research efforts are asked for feedback about these materials so that they can be continuously improved based on the real needs of the microbiome research community.

Ambassadors & Champions: embracing data stewardship

We developed the Ambassador and Champions programs to foster learning, support networking, and translate feedback to actions through our user research program. Many organizations have effectively utilized Ambassador programs focused on building relationships within and across teams and institutions, along with facilitating partnerships, like the ASM Ambassador Program or the NEON Ambassador Program. We have tailored this model to create two distinct, but synergistic programs, that leverage different aspects of how the NMDC builds community.

First, the Champions program was launched in October 2019 during a community workshop focused on metadata standards [2]. The Champions program is open to any researcher regardless of career stage or location, with applications open year round and a rolling review process. During the pilot phase, we worked closely with interested researchers to further refine the goals and expectations of the program and clarified benefits. In early 2021, the Champions program underwent an internal evaluation to assess bidirectional benefits and was remodeled to set priorities on "early access" to NMDC beta-testing and user research, quarterly "meet ups" to support networking, and opportunities to contribute on an ad-hoc basis. This remodeling helped to expand the program in mutually beneficial ways and has seen steady growth in the number of researchers interested in supporting the range of NMDC engagement and development activities (**Figure 3**). To date, 68 Champions are actively participating in the program. Champion paper writing subgroups were launched in February 2024, with the first active group focusing on reusing and publishing on microbiome data from the NMDC Data Portal. The Champions also frequently contribute to user research efforts, the development of training materials, and they advocate for data stewardship within the microbiome research community.

The Ambassador program was launched in Spring 2021 to complement the Champions program. The focus was on a year-long cohort-based learning experience for US-based early career researchers interested in promoting metadata standards. The program aimed to build a network of early career researchers spanning diverse areas of expertise and microbiome science from human health to agriculture to aquatic ecosystems. Through a competitive application process, Ambassadors commit to work with the NMDC team to promote the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles and also demonstrate a strong commitment to inclusion, diversity, equity, and accountability (IDEA). Following training in metadata standards and data stewardship, the Ambassadors develop a set of two events or workshops to host within their respective

research community. The initial cohort of Ambassadors included 12 early career researchers who collectively managed 23 events that reached more than 800 researchers [3]. Since the pilot, we have expanded the scope of the Ambassador training and activities to include standardized bioinformatics tools and a broader view of data stewardship. Ambassadors contribute to the development of training and event materials, spread the NMDC mission, provide feedback to the NMDC team about the program, and work with each other and the NMDC team to plan and host events relevant to diverse areas of microbiome research. They are also provided with networking opportunities, credit for their work as an Ambassador, a $1,000 honorarium for event and travel costs, co-authorship opportunities, extensive technical and professional trainings, logistical and administrative support from the NMDC team, and continual recognition for their work through the NMDC communications channels. The 2023 cohort of 13 Ambassadors hosted 21 events, reaching over 550 researchers. We have seen a nearly two-fold increase in our current Ambassador cohort (**Figure 3**).

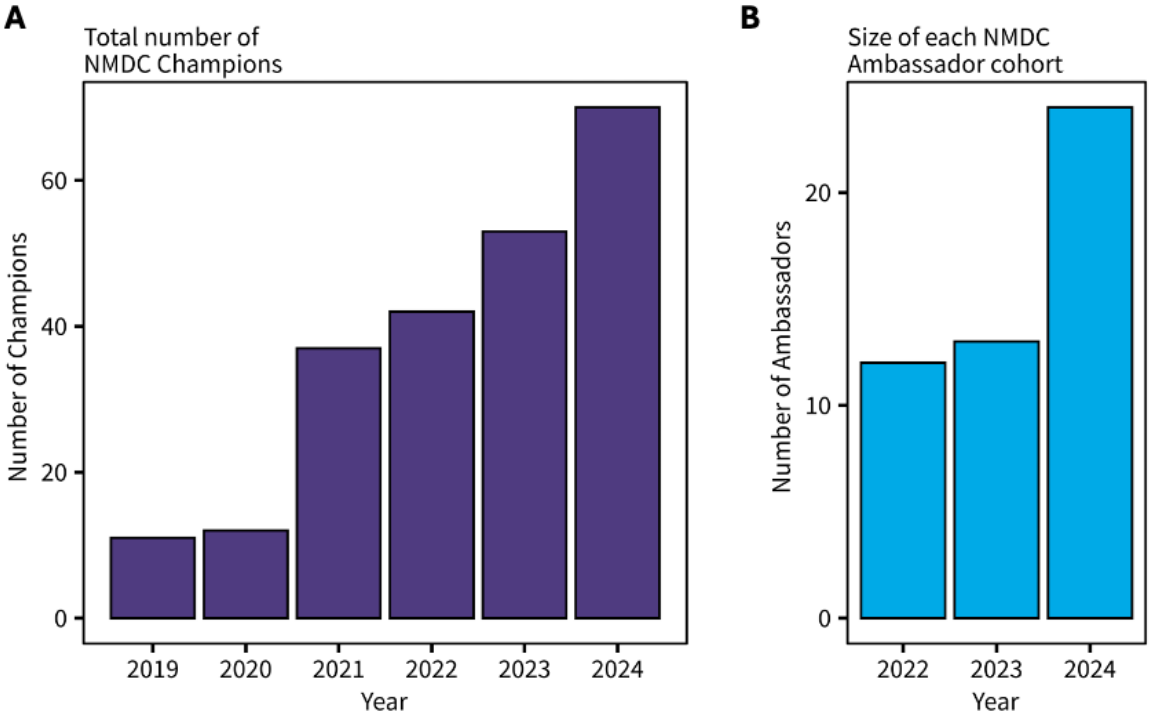

Figure 3. Overview of the number of early career researchers participating as (**A**) Champions and (**B**) Ambassadors since the launch of each program.

Institutional partnerships: paving a collaborative future

In addition to our work with individual researchers and research teams, we have worked across consortia, scientific societies, and institutions both nationally and internationally to extend our reach. Strategic institutional partnerships allow the NMDC to identify and foster scientific and organizational priorities that are mutually beneficial to move microbiome research forward. Our three longest and most impactful partnerships are with ASM, the Genomic Standards Consortium (GSC), and the Microbiome Centers Consortium (MCC).

ASM leadership has been heavily involved from the very beginning through two separate program development and ideation workshops, one convened in early 2017 at the JGI and the second hosted in late 2017 by ASM in their Washington D.C. headquarters. This fruitful collaboration has resulted in continued engagement, from the involvement of ASM's CEO Stefano Bertuzzi on the NMDC Scientific Advisory Board (SAB) to the launch in 2021 of the ASM Microbiome Data Prize. Since 2019, we have worked with ASM staff to host town hall events and scientific sessions at the ASM Microbe conference, which is the largest single meeting across the microbial sciences, to engage with the research community and showcase the work of NMDC Ambassadors. Further, we aligned strategic priorities in 2021 around the Climate Change and Microbes five year scientific portfolio from the American Academy of Microbiology, the honorific arm of ASM. NMDC leadership was involved in the launch of this important scientific portfolio through the Microbe and Climate Change - Science, People, and Impacts report [4] and mSystems commentary [5]. This past year, we have bolstered our partnership with ASM to advance an open science ecosystem and will be hosting an extended session on 'Microbial Data and Tools without Borders' (EEB-MC-001) as part of ASM Microbe 2024.

Our team similarly has a longstanding and productive partnership with the GSC to drive technical implementation of the GSC's community standards and synergistically engage with the research community. Since the initiation of the NMDC, GSC's President Lynn Schriml has served as an active member of the NMDC SAB. Our team is heavily involved in both the GSC's Compliance and Interoperability Working Group and Technical Working Group, along with membership on the GSC Board. In October 2023, we worked closely with the GSC to successfully transition the Minimum Information about any (x) Sequence (MIxS [6]) standard to use the Linked Data Modeling Language (LinkML) which provides a robust, yet flexible solution to data modeling. As part of this release, our team was instrumental in launching the new MIxS documentation website and we also have a forthcoming book chapter in the *Comparative Genomics* volume in *Methods in Molecular Biology* [7].

Lastly, our partnership with the MCC dates to 2019 when both the NMDC and the MCC were launched. The MCC is a collaborative network of microbiome centers [8] led by Dr. Jennifer Martiny at the University of California, Irvine who serves as the Chair of the NMDC SAB. We have actively participated in the MCC annual meetings, webinar series, and jointly involved in ASM's Microbiome Hill Day to educate policymakers on the importance of microbial research and a national microbiome network. Many of the NMDC Ambassadors and Champions are members of participating Microbiome Centers, and we have benefited from the thoughtful leadership the MCC has brought to advancing a coordinated and collaborative network.

FAIR microbiome data for all: promoting equity, accessibility, and inclusion

In 2021, we developed an initial Inclusion, Diversity, Equity, and Accountability (IDEA) Strategic Plan that publicly cemented NMDC's commitment to making microbiome research more equitable. This Plan was divided into three overall goals: (1) promote transparency and accountability within NMDC's team and operations; (2) promote transparency and accountability within NMDC's governance structure; and (3) engage

and support diverse stakeholders and users. Each goal had associated tangible actions for a total of 27 actions. As of March 2024, 22 of those actions have been completed, with 5 currently in progress.

To ensure our ongoing commitment to our IDEA principles, we decided to annually revise our goals and actions and release our accountability report for previously outlined action items. We released the 2024 NMDC IDEA Action Plan that includes new goals and 41 action items for the year aimed at promoting IDEA best practices in all NMDC activities. The IDEA Action Plan webpage also includes a link to provide the team with anonymous feedback. We encourage community members to join our IDEA Working Group that meets quarterly to discuss ongoing IDEA-related actions, upcoming events of interest, new research and reports surrounding IDEA topics, and how to best promote FAIR data and open science for all.

We consistently engage with historically underserved and underrepresented communities including hosting a diverse panel focused on providing data management tools and resources to the community at the National Diversity in STEM conference organized by the Society for the Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS); translating several NMDC user guides into Spanish and French; and working with a 2023 Ambassador to translate training slides into Spanish for an event hosted at the University of Puerto Rico.

Building robust outreach campaigns to foster community

To continually provide the microbiome research community with updates about the NMDC products and activities, we send out a quarterly newsletter, *The Microbiome Standard*. We also send out monthly emails to our Champions with updates, opportunities, and upcoming events. We engage with the microbiome research community on X/Twitter, LinkedIn, and Instagram, and have launched several social media campaigns to highlight our Ambassadors, Champions, collaborators, NMDC team members, outreach activities, and product updates. The NMDC Community Slack provides a means to discuss the current state of microbiome research with community members. We also publish blog posts, science highlights, and peer-reviewed publications to keep the community aware of NMDC activities and facilitate collaborations. We release an annual report every year to communicate the progress of the NMDC Program. Together, we have consistent engagement with over 2,100 followers on X/Twitter, more than 650 followers on LinkedIn, 100 followers on Instagram, and consistently see several hundred visitors to our Portals and well over 2,000 visits quarterly to the NMDC website. Through tracking our outreach and engagement metrics, we can observe growth and monitor across all our stakeholder groups.

References

1. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023. doi:10.1038/s41587-023-01953-y

2. Vangay Pajau, Burgin Josephine, Johnston Anjanette, Beck Kristen L., Berrios Daniel C., Blumberg Kai, et al. Microbiome Metadata Standards: Report of the National Microbiome Data Collaborative's Workshop and Follow-On Activities. mSystems. 2021;6: 10.1128/msystems.01194–20.

3. Kelliher JM, Rudolph M, Vangay P, Abbas A, Borton MA, Davenport ER, et al. Cohort-based learning for microbiome research community standards. Nat Microbiol. 2023;8: 751–753.

4. Microbes and Climate Change – Science, People & Impacts: Report on an American Academy of Microbiology Virtual Colloquium held on November 5, 2021. Washington (DC): American Society for Microbiology; 2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK580166/ doi: 10.1128/AAMCol.Nov.2021.

5. Tiedje JM, Bruns MA, Casadevall A, Criddle CS, Eloe-Fadrosh E, Karl DM, et al. Microbes and Climate Change: a Research Prospectus for the Future. MBio. 2022;13: e0080022.

6. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29: 415–420.

7. Eloe-Fadrosh EA, Mungall CJ, Miller MA, Smith M, Patil SS, Kelliher JM, et al. A Practical Approach to Using the Genomic Standards Consortium MIxS Reporting Standard for Comparative Genomics and Metagenomics. In: Setubal, J. (eds) Comparative Genomics. Methods in Molecular Biology, vol 2802. Humana Press, New York, NY. (*In Press*)

8. Martiny JBH, Whiteson KL, Bohannan BJM, David LA, Hynson NA, McFall-Ngai M, et al. The emergence of microbiome centres. Nat Microbiol. 2020;5: 2–3.