

Cohort-based learning for microbiome research community standards



In the past decade, microbiome researchers have produced a wealth of sequence data and associated microbial outputs and activity¹. Most microbiome studies to date have analysed microbiomes in order to answer a specific research question, with little emphasis placed on data reusability. However, increasing microbiome data reuse would result in a better return on investments and an increase in accessibility to microbiome science for researchers unable to generate their own microbiome data. It may also enable researchers to address scientific questions beyond the scope of the original study². Recognizing these advantages, the scientific community has repeatedly advocated for findable, accessible, interoperable and reusable (FAIR) data³. We launched the National Microbiome Data Collaborative (NMDC) Ambassador Program in April 2021 to increase FAIR data adoption and metadata stewardship among microbiome researchers. Herein, we provide information and insights regarding the implementation of a community learning model to assist in the advancement of other iterations of this type of programme.

The campaign for FAIR data advocates for sufficient metadata (for example, geographic location, sampling date, host information, and so on) to be available for microbiome datasets, because the utility and reusability of data sharply decrease when crucial metadata fields are missing, insufficient or improperly formatted^{4,5}. Training events, ideally with a practical or hands-on component, are essential to increase recognition and adoption of metadata standards.

The NMDC was founded in 2019 as a Department of Energy project with the intention of making multi-omics microbiome data FAIR and open to all⁶. Core to this mission is the implementation of standardized metadata fields linked to microbiome datasets, leveraging the minimum information about any (x) sequence (MIxS) standards developed by the Genomic Standards Consortium⁷. During the initial stages of the NMDC programme, it was confirmed that the microbiome research community had varied knowledge, or were unaware, of the benefits of metadata standards

Table 1 | Ambassador Program requirements and expectations

Contribute to metadata training materials	Update content for one NMDC-provided template presentation on sample metadata to reflect an environment relevant to their research community
Engage with their research community	Co-host one four-hour workshop with the NMDC to gather feedback to improve sample metadata standards training
Spread the NMDC mission	Partner with the NMDC to present at one conference session, webinar, panel or other event of their choice
Share their perspectives	Share two updates on FAIR microbiome data activities on any NMDC online platform (social media, website, newsletter)
Learn about metadata standards	Attend all NMDC trainings

as well as options for available metadata templates⁸. In order to address this knowledge gap, we set up the NMDC Ambassador Program as a community learning model.

Community learning models enable a cohort of researchers to receive targeted training and materials to disseminate and present to their own research communities, and have been shown to increase the overall reach of the training content⁹. The NMDC advertised the Ambassador Program through social media, mailing lists, the NMDC website and professional networks. The aim was to support a culture change within microbiome research towards increased metadata stewardship by recruiting, training, and supporting early career researchers to host metadata standards workshops.

From 56 total applications, 12 early career researchers, representing 11 institutions across the United States, were selected for the pilot scheme. Ambassador backgrounds, areas of expertise and microbiomes of interest were varied. Ambassadors received an honorarium to cover costs associated with workshop implementation. Six unique training sessions for the Ambassadors were hosted by the NMDC team in partnership with the Center for Scientific Collaboration and Community Engagement (CSCCE). These sessions covered best practices in data management, metadata standards, community engagement and effective practices to serve as community liaisons.

After completion of all six training sessions, each Ambassador was asked to adapt NMDC-created metadata training templates to include specifics regarding their microbiome

of interest. They also prepared hands-on guides for MIxS metadata spreadsheets. Ambassadors were responsible for hosting two workshops, or training sessions, centred around metadata standards for their respective research communities. Table 1 outlines the full list of Ambassador Program expectations.

To measure the impact of these workshops, post-workshop surveys were conducted to assess how attendee knowledge and familiarity with metadata standards increased after Ambassador events¹⁰. In less than a year, a total of 23 Ambassador-hosted workshops and presentations reached more than 800 researchers, and demonstrated improvement in participant recognition and practical experience with metadata standards (Fig. 1).

Due to the COVID-19 pandemic, the programme's outreach strategy had to pivot from the envisioned in-person workshops to predominantly virtual options. Training was provided to Ambassadors by CSCCE in virtual event organization to enable switching to online-only training formats. Ambassadors implemented several CSCCE-proposed strategies including real-time polls, digital whiteboards and live chat, all of which helped to make virtual events more engaging and informative.

The NMDC continues to recruit 10–20 Ambassadors through this programme annually. Learnings from the first cohort provided the NMDC team with valuable insights for how to improve the programme moving forward. Feedback and recommendations collected from the Ambassadors during post-programme surveys and debriefing

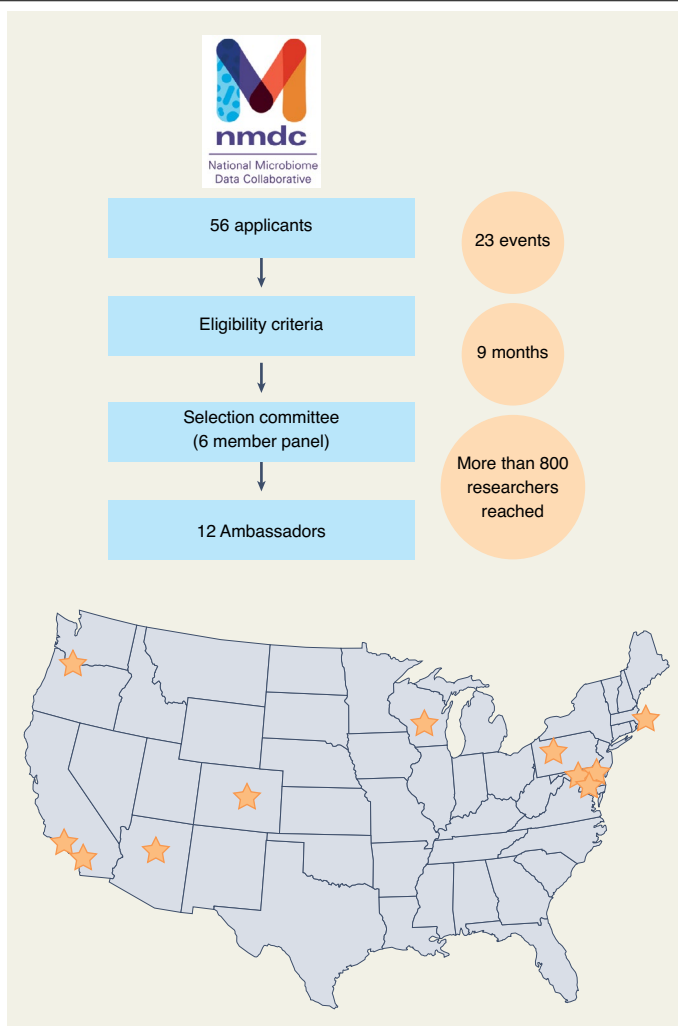


Fig. 1 | The selection of the pilot cohort of NMDC Ambassadors, the locations of their affiliated institutions and the quantitative impact of the Ambassador Program. Eligibility criteria are described [here](#). Publ. note: Springer Nature is neutral about jurisdictional claims in maps.

calls were summarized in a final report and presentation by CSCCE to the NMDC team. The Ambassadors wanted more guidance on choosing venues for workshops. Therefore, more comprehensive language was drafted and a list will be provided to future Ambassadors as a starting point for venue choices. Several Ambassadors communicated the benefits of having an NMDC team member present at their events to answer questions, especially regarding how metadata standards fit into the overall NMDC mission, and how metadata requirements form the basis of the NMDC Data Portal¹¹. This feedback led to the NMDC team creating supplemental training materials for Ambassadors focused on how to talk about and answer questions relating to the NMDC. The Ambassadors also recommended having

more variety in workshop content, therefore the programme scope was amended to include material on microbiome bioinformatics workflows and data reuse through the NMDC Data Portal, while continuing to include content on metadata standards.

The pilot cohort of the Ambassador Program demonstrated the advantages and practicality of a community learning model in microbiome research and biological data stewardship. Ambassadors were provided with extensive training from industry experts, they were able to hone their presentation and hosting skills, and they were given key opportunities to network and present at diverse venues. Equally, the Ambassadors advanced the NMDC mission and broadened the NMDC community network by reaching more researchers

than would have been possible by NMDC team members alone.

The NMDC Ambassador Program will continue to evolve, and we hope that it will continue to increase awareness of, and proficiency with, data stewardship, leading to better use of microbiome datasets over time.

Julia M. Kelliher , **Marisa Rudolph** ², **Pajau Vangay** ², **Arwa Abbas** ³, **Mikayla A. Borton** ⁴, **Emily R. Davenport** ⁵, **Karen W. Davenport**¹, **Natalia G. Erazo**⁶, **Chloe Herman**⁷, **Lisa Karstens** ⁸, **Brandon Kocurek** ⁹, **Holly L. Lutz**¹⁰, **Kevin S. Myers** ¹¹, **Ingrid Ockert**², **Francisca E. Rodriguez**¹, **Camille Santistevan** ¹², **Jaclyn K. Saunders**^{13,14}, **Montana L. Smith**¹⁵, **Emily Vogtmann** ¹⁶, **Amanda Windsor**¹⁷, **Elisha M. Wood-Charlson** ², **Lou Woodley** ¹² & **Emiley A. Eloefadrosh** ²

¹Los Alamos National Laboratory, Los Alamos, NM, USA. ²Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ³Children's Hospital of Philadelphia, Philadelphia, PA, USA. ⁴Colorado State University, Fort Collins, CO, USA. ⁵Pennsylvania State University, State College, PA, USA. ⁶Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA. ⁷Northern Arizona University, Flagstaff, AZ, USA. ⁸Oregon Health and Science University, Portland, OR, USA. ⁹Center for Veterinary Medicine, U.S. Food and Drug Administration, Laurel, MD, USA. ¹⁰Scripps Research, La Jolla, CA, USA.

¹¹Wisconsin Energy Institute and Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA.

¹²Center for Scientific Collaboration and Community Engagement, Oakland, CA, USA.

¹³Woods Hole Oceanographic Institution, Falmouth, MA, USA. ¹⁴University of Georgia, Athens, GA, USA. ¹⁵Pacific Northwest National Laboratory, Richland, WA, USA. ¹⁶National Cancer Institute, Bethesda, MD, USA.

¹⁷Center for Food Safety and Applied Nutrition, Office of Regulatory Science, U.S. Food and Drug Administration, College Park, MD, USA.

e-mail: jkelliher@lanl.gov; aeloefadrosh@lbl.gov

Published online: 17 April 2023

References

1. Jansson, J. K. & Baker, E. S. *Nat. Microbiol.* **1**, 16049 (2016).
2. *Nat. Microbiol.* **2**, 1573 (2017).
3. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
4. Cernava, T. et al. *Environ. Microbiome* **17**, 33 (2022).

5. Dempsey, W. P., Foster, I., Fraser, S. & Kesselman, C. *Harvard Data Sci. Rev.* <https://doi.org/10.1162/99608f92.44d21b86> (2022).
6. Wood-Charlson, E. M. et al. *Nat. Rev. Microbiol.* **18**, 313–314 (2020).
7. Yilmaz, P. et al. *Nat. Biotechnol.* **29**, 415–420 (2011).
8. Vangay, P. et al. *mSystems* **6**, e01194–20 (2021).
9. Stevens, S. L. R. et al. *PLoS Biol.* **16**, e2005561 (2018).
10. Kelliher, J. *NMDC Ambassador Event Survey* (2023); <https://doi.org/10.6084/m9.figshare.22226329.v1>

11. Eloë-Fadrosh, E. A. et al. *Nucleic Acids Res.* **50**, D828–D836 (2021).

Author contributions

J.M.K. and E.A.E.-F. conceptualized and wrote the majority of the manuscript with substantial contributions from M.R. and P.V. A.A., M.A.B., E.R.D., N.G.E., C.H., L.K., B.K., H.L.L., K.S.M., J.K.S., E.V. and A.W. were the Ambassadors from this pilot cohort and contributed to the content of the

manuscript. C.S. and L.W. contributed to the programme and synthesis of feedback and results. P.V. also ran the Ambassador Program during the pilot cohort and K.W.D., I.O., F.E.R., M.L.S. and E.M.W.-C. contributed to running the Ambassador Program, creating training materials and editing the manuscript. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.