



# National Microbiome Data Collaborative

Progress Report 2022  
U.S. Department of Energy

# Vision

To connect data, people, and ideas to advance microbiome innovation and discovery.

---

# Mission

To support a Findable, Accessible, Interoperable, and Reusable (FAIR) microbiome data sharing network, through infrastructure, data standards, and community building, that addresses pressing challenges in environmental science.



Cover credit: Photo courtesy of Jeremy Snyder  
This page credit: Roy Kaltschmidt, Berkeley Lab

# Enabling inclusive, interdisciplinary microbiome science

2022 was a time of renewal and reflection for the National Microbiome Data Collaborative (NMDC) team. As we prepared to transition from our three-year Pilot phase and drafted an ambitious plan for the next three years, we reflected on the ways that the NMDC has made significant strides towards our goal of creating a more equitable and dynamic microbiome research community. Our product initiatives—the [NMDC Submission Portal](#), the [NMDC Data Portal](#), and [NMDC EDGE](#)—have helped researchers gain new understanding into their data and opened up new workflows. By taking our learnings from the Pilot and integrating them with innovations in distributed data infrastructure and linked data technologies, the NMDC supports new ways for researchers to ask questions about how genes, individual microbes, and microbial communities are associated with environmental processes.

The NMDC Renewal Proposal was a major undertaking, and I commend my teammates for their incredible work as we re-envisioned how the NMDC will support a sustainable community-driven data infrastructure. We presented our work to an external review panel and our colleagues at the Department of Energy (DOE) on

June 29 and June 30. The review panel recognized our team's effort on the NMDC Pilot and provided key recommendations. The NMDC's renewed vision is to drive a microbiome data sharing network connecting data, people, and ideas to advance microbiome innovation and discovery. To move towards this vision and beyond the NMDC Pilot, we proposed a transformative approach to microbiome science that leverages lessons learned and laid out an ambitious proposal framework for collaborative, interdisciplinary data infrastructure to support integrative science. Over the next three years, our focus will be on supporting a FAIR (findable, accessible, interoperable, and reusable) microbiome data sharing network, through infrastructure, data standards, and community building, that addresses pressing challenges in environmental science.



**Emiley Elloe-Fadrosch**  
National Microbiome  
Data Collaborative Lead

The first section of this report examines the ways that the NMDC continues to build tools to enable research and discovery. We frequently ask for community feedback so that we can improve our [Submission Portal](#), [Data Portal](#), and [NMDC EDGE](#). Since it was launched in 2021, the Data Portal has been used to conduct research into soil metadata from areas as diverse as boreal forests, wetlands, and botanic gardens. Similarly, NMDC EDGE continues to be an important tool for researchers to access standardized bioinformatics workflows. Our newest product initiative, the NMDC Submission Portal, was launched in April 2022. Together, our product initiatives continue to be used for sample submissions, training activities, and research collaborations.

The second section of this report describes the NMDC's numerous engagement activities. This past year, the NMDC team also worked to create new opportunities for virtual learning and networking. The NMDC and the Environmental Molecular Sciences Laboratory (EMSL) teamed up for a "Summer School", where attendees heard lectures from subject-matter experts in the fields of metadata and FAIR data, metagenomics, natural organic matter (NOM) characterization, and statistical analyses. The NMDC team also attended and presented at several events over the past few months, including two (ASM) events, [ASM Microbe](#), the [18<sup>th</sup> International Symposium on Microbial Ecology \(ISME\)](#), and the [2022 All Scientists' Meeting for the Long Term Ecological Research \(LTER\) Network](#). The team also hosted a hands-on workshop at the fourth annual meeting of the [Metabolomics Association of North America](#). Finally, to meet the

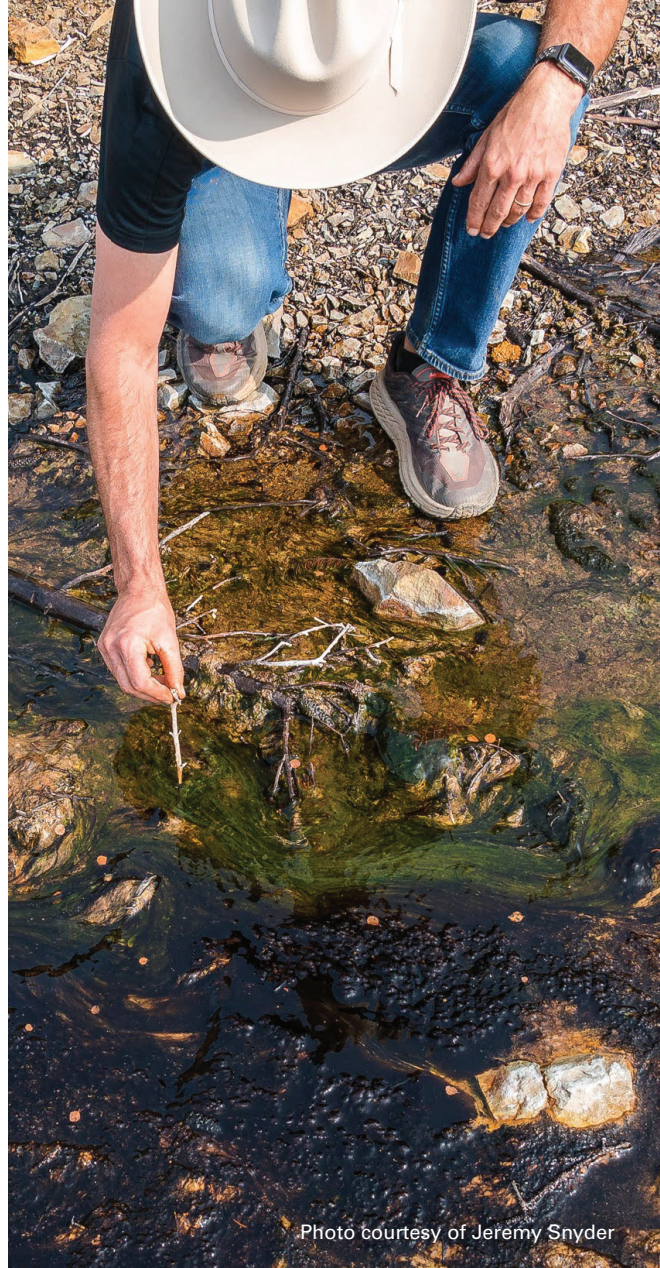


Photo courtesy of Jeremy Snyder

the continued need for remote activities, the NMDC team successfully organized a [virtual series](#). These five webinars, running from January 2022 to May 2022, featured 20 different subject-matter specialists and focused on many topics, including the benefits of using standardized workflows, the features of user-centered design, the data-citation landscape, the development of metadata standards, and, finally, the barriers to microbiome research. This also marked the second year of the NMDC's [Champions](#) and [Ambassadors](#) programs. Each year, the NMDC sponsors a cohort of early career researchers (Ambassadors), who help us test our products and train other researchers in metadata best practices. We also encourage researchers to join our Champion program and use their feedback to improve our projects. From in-person workshops to hybrid lectures, the NMDC team has brainstormed and implemented creative ways to bring the research community closer together.

The NMDC continues to establish itself as a center for collaborative, integrative microbiome data science research. Our team provides unique services to tackle new integrative science questions. We advance new community partnerships to enable high-impact science. We seek to leverage partnerships and resources to support a sustainable infrastructure ecosystem.

We invite you to read on to learn more about our team's accomplishments this past year, and we hope that you will engage with the NMDC to shape the future of microbiome data discovery.

A handwritten signature in black ink, reading "Emiley Eloé-Fadrosh". The signature is fluid and cursive, with a large, stylized 'E' at the beginning and a long, sweeping tail that loops back under the name.

Emiley Eloé-Fadrosh  
National Microbiome  
Data Collaborative Lead

# Creating needed infrastructure to answer tomorrow's research questions

The NMDC is tackling infrastructure challenges through iteratively developing an open-source, integrated ecosystem for multi-omics data. Metadata standards and consistently processed multi-omics data are essential to successfully developing a FAIR microbiome data infrastructure. The NMDC launched the [Submission Portal](#) in April 2022 and continued our user-centered design process to improve the [Data Portal](#) and [NMDC EDGE](#). This year marked an important milestone for the NMDC: NMDC products are now able to support environmental microbiome data, as well as multi-omics data from the Earth Microbiome Project. These new capabilities will aid the NMDC as it transitions from the Pilot phase to broadly supporting the environmental research community.

Credit: Andrea Starr,  
Pacific Northwest  
National Laboratory



# Lowering Barriers to Data Collection through User-Centered Research

The NMDC is a co-created resource: it is implicitly designed with the scientific community in mind, and its utility is further enhanced through user-generated feedback. We have engaged in extensive user research through interviews and direct collaboration with the scientific community that have informed the design, functionality, development, and display of data through NMDC products.

Our user-centered design methodology enables the scientific community to provide feedback leading to iterative and continuous improvement of our systems and ensuring that our systems enable a high level of scientific productivity. In 2022, we conducted three rounds of user interviews (17 user interviews total) from three target groups: general microbiome researchers, potential data submitters, and metadata generators. This user-centered design process resulted in the generation of 116 user insights that have led to the creation of key features in NMDC products, as well as changes to the NMDC website and engagement activities.

As we gather feedback, the NMDC looks for opportunities to create tools that will lower barriers to data collection. In April 2022, the NMDC launched a beta release of

## ACCOMPLISHMENTS

- Launched the NMDC [Submission Portal](#) to assist the collection of multi-omics studies and biosample information supporting requirements for Joint Genome Institute (JGI) and EMSL sample submission
- Together with the [Genomic Standards Consortium](#) (GSC), rendered the Minimum Information about any (X) Sequence (MIxS) standards in Linked Data Modeling Language (LinkML) and released the latest version
- Added support for all available MIxS environmental packages
- Completed a pilot program to support metadata submission for five Facilities Integrating Collaborations for User Science (FICUS) projects
- Generated 116 insights from user research and translated them into actionable tasks

the NMDC [Submission Portal](#). This portal provides researchers one place to submit study and biosample information compliant with existing microbiome standards as defined by the GSC. The portal has clear descriptions, support for all available MIxS environmental packages, and an intuitive, real-time validation function

to ensure compliance with standards. Subsequently, the NMDC team hosted several workshops centered around the portal and taught users how to complete and submit metadata. The NMDC also created a user guide to assist in user submissions and also in users' understanding of portal functionality.



Credit: Roy Kaltschmidt, Berkeley Lab



# Facilitating New Ways to Access and Integrate Multi-Omics Data

The [NMDC Data Portal](#) was launched in 2021, and the NMDC team has continued to add studies, biosamples, and standardized data products from our multi-omics bioinformatics workflows to it. The portal allows users to browse all multi-omics analyses by functional annotation, environment, and associated metadata. Users can explore the portal using the investigator name, -omics processing information, KEGG Orthology (KO), module and pathway terms, and a suite of environmental descriptors. Users may also search through two systems for ecosystem classifications, including the GOLD ecosystem classification paths and using EnvO classification terms (Environmental Broad Scale, Environmental Local Scale, Environmental Medium).

The [Data Portal](#) currently contains more than 19,000 data files available for download, which are associated with 754 biosamples from a breadth of environmental microbiomes, spanning river sediments, subsurface shale carbon reservoirs, plant-microbe associations, and temperate and tropical soils. In 2022, the NMDC team added several new datasets, including large amounts of data from the multi-omics [Earth Microbiome Project](#) (EMP500). Many of the datasets available through the

## ACCOMPLISHMENTS

- Added nearly 646 processed metagenomes and 996 NOM analyses
- Coordinated with the Integrated Microbial Genomes & Microbiomes (IMG/M), Genomes OnLine Database (GOLD), and Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) teams to support data sharing across complementary resources
- Implemented user feedback to improve data discovery and display in the portal, including improved page navigation and faceted search functionality, improved annotation search, and intuitive menu buttons and help links
- Released new streamlined NMDC [documentation](#) that supports four modes of documentation, including tutorials, how-to guides, technical references, and explanation

portal are also now available as a narrative within the DOE's [Systems Biology Knowledgebase \(KBase\)](#).

A key feature of the [Data Portal](#) is enabling the research community to discover multi-omic data through a variety of search functionalities, such as faceted search and interactive visualizations using both the environmental metadata and functional annotations. Our team has developed a detailed [Data Portal User Guide](#) to help users understand the portal's full functionality. This guide was incorporated into the new streamlined

[documentation](#) that the team developed in this past year using the Diátaxis technical documentation framework. Diátaxis specifies four modes of documentation: tutorials, how-to guides, technical references, and explanation.

The NMDC [Data Schema](#) supports all aspects of the NMDC infrastructure, including development of the [Submission Portal](#) and the [Data Portal](#). The schema has been developed to define how metadata elements (e.g., biosamples and environmental descriptors) are related, and supports community-driven standardized



Credit: Emily Davenport

Figure 1: NMDC workshop participants learn about the NMDC Data Portal

metadata formats. This past year, we pioneered a new modeling approach called the LinkML for developing the schema. Through our unique collaboration with the GSC, we used the LinkML modeling approach as the basis for the MlxS standard and collaboratively released the latest version. This is a considerable achievement, moving from a spreadsheet-based reporting standard to a machine-actionable standard with automatic validation.

Going forward, the NMDc team will work closely with staff at the JGI and EMSL to develop a streamlined sample submission system. This new work, initiated towards the end of 2021, is in close collaboration with the [DataHarmonizer](#) team at Simon Fraser University. Adoption of the DataHarmonizer tool will advance our efforts to lower barriers to sample metadata submission and support community-driven standards.

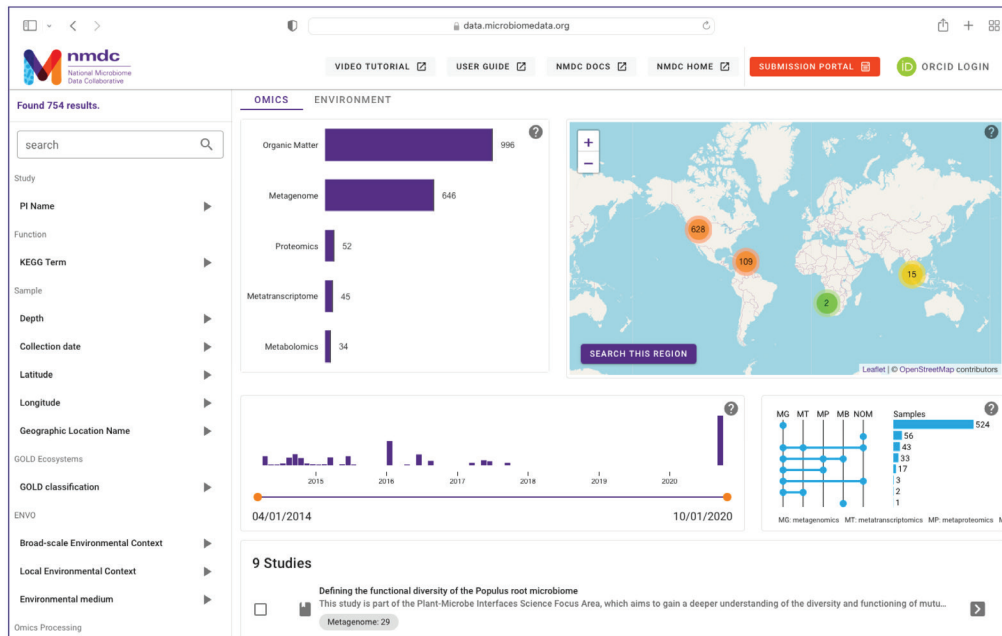


Figure 2. Data Portal landing page

# Accessible Bioinformatics Workflows for Data

The lack of standardized bioinformatics methods can impede a researcher's ability to compare microbiome data generated by different groups of researchers. To address this, the NMDC team has integrated existing production-quality open-source bioinformatics workflows used by the JGI and EMSL to process raw multi-omics data and produce interoperable and reusable annotated data from metagenome, metatranscriptome, metaproteome, metabolome, and NOM characterizations. To support broad availability of these workflows, the NMDC team has made these workflows available as [stand-alone downloadable software containers](#).

To advance genome-based characterization of environmental viruses and plasmids, the NMDC is leveraging our close partnership with the JGI to support high-throughput systematic identification and classification of viruses and mobile genetic elements. As part of our work beyond the Pilot phase, we have partnered with the JGI to integrate its innovative new virus and mobile genetic elements identification tool that allows concurrent discovery and characterization of both viruses and plasmids from sequence data. This new tool, [geNomad](#), which supports the JGI's IMG/VR database

## ACCOMPLISHMENTS

- Advancing virus discovery in collaboration with the JGI, the NMDC supports a newly developed workflow for virus and mobile genetic element identification as both a standalone software container and available through [NMDC EDGE](#)
- NMDC EDGE includes metagenomic, metatranscriptomic, and the NOM characterization workflows with improved documentation and tutorials
- NMDC EDGE has undergone beta testing with the NMDC Champions, and more than 80% of suggestions have been successfully implemented
- NMDC staff hosted several training workshops centered on using the platform, including at the National Summer Undergraduate Research Project (NSURP) and the EMSL 2022 Summer School

for genomes of cultivated and uncultivated viruses, is now at the core of a new containerized NMDC virus and plasmid workflow, which also includes automated quality estimation and taxonomic classification of predicted viruses, for broad access to the research community.

One way that our users apply this workflow is through the NMDC product [NMDC EDGE](#). The platform was developed to provide an intuitive, user-friendly graphical interface for users to run the NMDC workflows. It currently supports the metagenomic, metatranscriptomic, and NOM workflows. Detailed [workflow documentation](#) and [tutorials](#) enable both novice and experienced users to understand the various components of the data-processing steps. The first round of beta testing took place throughout 2021 and 2022, and resulted in more than 80% of all recommendations being implemented. NMDC Champions continue to beta test and perform usability testing on the NMDC workflows and NMDC EDGE.

NMDC team members have run several successful workshops this past year using [NMDC EDGE](#). These workshops included a pilot NMDC EDGE virtual workshop, which was attended by 40 users. When surveyed afterwards, all of the workshop attendees

reported that they found NMDC workflows useful, and they would use the platform in the future. In June, the NMDC ran an NMDC EDGE workshop for the [NSURP](#).

NMDC team members also had the unique opportunity to co-host the EMSL Summer School, "[Soils Exposed](#)," which educated early career researchers on multi-omics data analyses of soil samples. This summer school taught participants about metadata completion and analysis of metagenomics and NOM data via Fourier-transform ion cyclotron resonance (FTICR). The NMDC team led training sessions on metadata standardization and its importance for data sharing and reuse. Students learned how to analyze this data, perform statistical analyses, and visualize their data. The [Submission Portal](#) and [Data Portal](#) were extensively featured in interactive hands-on activities. Additionally, detailed training on metagenomic analysis spotlighting [NMDC EDGE](#) was provided. The summer school was very successful, with more than 450 registrants and 150 attendees participating daily during the open presentation sessions.

# NMDC EDGE

The National Microbiome Data Collaborative (NMDC) is an effort to make microbial data (FAIR) by providing search functions and access to standardized data and data products. NMDC EDGE is an open-source bioinformatics platform with a user-friendly interface.

**Upload Files**  
 Max single file size is 10.0GB. Max server storage space is 150.0GB. Files will be kept for 180 days.  
 Allowed file extensions are: fasta.gz, fa, fasta, fastq, fastq.gz, fastq, fastq.gz, contigs.gz, gbk, vcf, tsv, bed, coverage, raw, raw.gz, bam, bam.gz

Storage space usage: Object/150.0GB  
 Uploading size: Bytes

Drag Files or Click to Browse

**Run Multiple Workflows**

Project Name:  (Required, at 3 but less than 50 characters)

Description:

**Input File Reads**

Input File Search:

Is Indexed?

Input Interleaved Pairs:

Interleaved FASTQ #1:

**Choose Workflows**  
 All of the 168DC Metagenomic workflows are connected in EDGE, so that the output of one workflow can automatically be the input for the next workflow if several metagenomic workflows are selected.

- ReadQC Workflow**
- Read-based Taxonomy Classification Workflow**
- Metagenome Assembly Workflow**
- Metagenome Annotation Workflow**
- Metagenome MLAs Workflow**

Input Map File:

Input Domain File:

**Navigation Menu:**  
 The basics | User Guides | CLI Information

- Introduction
- Metagenomics
- Metatranscriptomics
- Metaproteomics
- Metabolomics
- Organic Matter

Figure 3. The left menu bar in NMDC EDGE allows users to upload data using an intuitive drag and drop feature, access video and written tutorials, and select desired bioinformatic workflows to run

# Building community by connecting data, people, and ideas

The cross-cutting nature of microbiome research, from environmental sciences, agriculture, and energy, to human, natural, and built environments, creates a diverse community united by a common challenge: the need for seamless data discovery, exploration, and access. To develop inclusive solutions to address this challenge, the NMDC team works closely with the research community to learn and better understand specific needs across the diverse array of microbiome research.



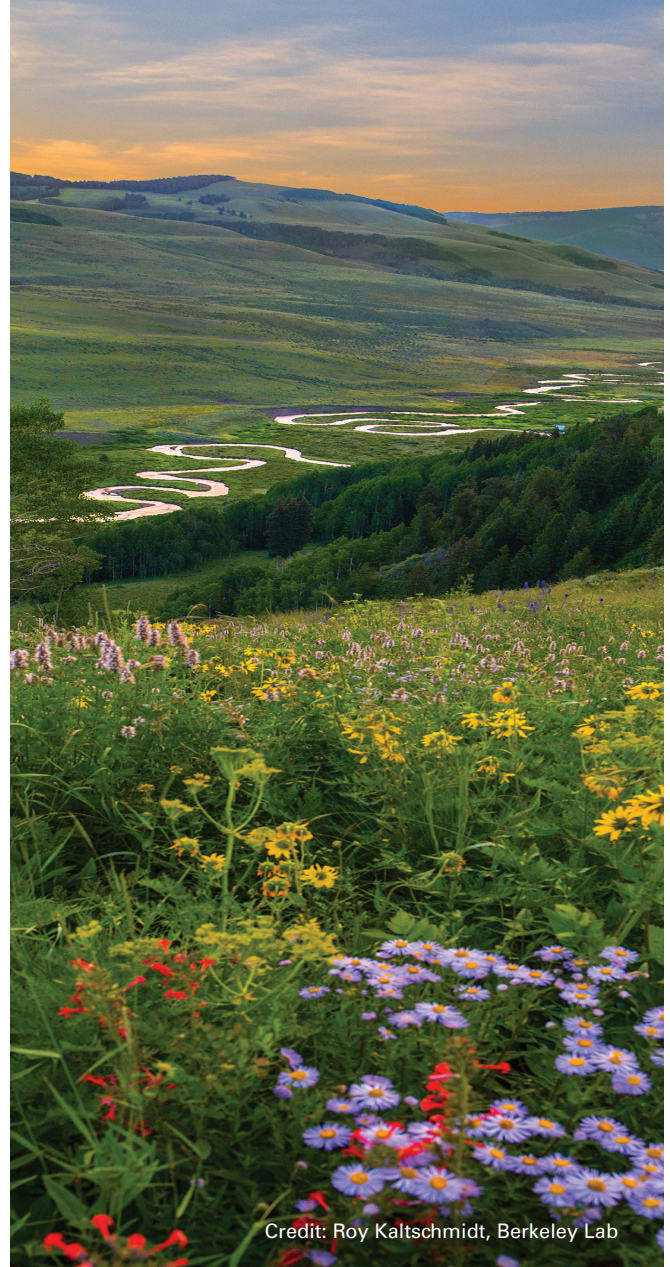
Credit: Roy Kaltschmidt, Berkeley Lab

# Inclusion, Diversity, Equity, and Accountability Plan

We are committed to supporting the diversity of experiences, expertise, backgrounds, needs, and perspectives of the microbiome research community, and to actively working towards an inclusive culture at a programmatic and individual level. We have developed a strategic plan to hold ourselves accountable to the goals listed below.

## Our Goals

- Goal 1:** Promote transparency and accountability within NMDC's team and operations.
- Goal 2:** Promote transparency and accountability within NMDC's governance structure.
- Goal 3:** Engage and support diverse stakeholders and users.





# Building the NMDC Community

One of the NMDC's primary objectives is to encourage a collaborative, community-centered approach. The NMDC team has been working towards this goal by creating a strategy that combines researcher engagement with institutional strategic engagement. To appeal to individual researchers, the team has trained community members to increase accessibility and knowledge of multi-omics microbiome research. Our team has encouraged users to process data through [NMDC EDGE](#) and use the [Submission Portal](#) for metadata collection and submission. To support our institutional partners, the team has created a strategy that focuses on building strategic partnerships and integrating the NMDC into the larger microbiome research landscape. Recently, we have embarked on new partnerships with the [National Ecological Observatory Network \(NEON\)](#) (NEON) and [NASA GeneLab](#). The NMDC and NASA GeneLab teams have provided feedback on each other's bioinformatic workflows, and NASA GeneLab team members attended the pilot NMDC workflow workshop. The teams also continue to pursue joint training efforts and collaborative opportunities. The NMDC has also planned joint training opportunities with NEON.

## ACCOMPLISHMENTS

- Presented at 29 events in 2022, reaching over 2,000 people
- The inaugural cohort of 12 NMDC Ambassadors hosted 23 workshops focused on metadata standards over the course of 2022, reaching more than 800 researchers
- Collaborated with more than 25 NMDC Champions to improve NMDC products through beta testing and user feedback
- Developed a new approach to user research to inform product development and conducted 49 user interviews, leading to 129 actionable insights
- Grew our online community to more than 1,500 followers

In June 2021, the NMDC team developed our IDEA (inclusion, diversity, equity, and accountability) [strategic plan](#) to ensure our commitment to supporting the diversity of experiences and perspectives within the microbiome research community. Our plan seeks to promote transparency and accountability within the NMDC team, operations, and governance structure. Additionally, the NMDC team looks for new ways to engage and support our diverse stakeholders and users. We actively work towards an inclusive culture at a programmatic and individual level. Aligned with our IDEA strategic plan, this past year, the NMDC organized a [session](#) at the annual meeting of the American Association for the Advancement of Science ([AAAS](#)) about how scientific institutions can build quality public data infrastructure that will make all data FAIR and people-centric.

We support early career researchers through our NMDC [Ambassador program](#). Our Ambassadors receive training and support to engage with their respective communities to advance the NMDC mission in a one-year cohort-based program. The inaugural cohort of 12 Ambassadors hosted metadata standardization workshops for their home institutions. Conversations with the Ambassadors led

to the creation of new supplemental materials, including communication guides for NMDC outreach that can be used at conferences. Additionally, our Ambassadors were able to help us beta test products, leading to improvements in specific MIxS environmental packages and in the GOLD ecosystem classification system. Collectively, their presentations and workshops reached more than 800 microbiome research community members. Feedback from these Ambassadors has been implemented in the plans for the 2023 iteration of the program.

We define community broadly, from those who sign up for our quarterly newsletters to those who attend our events at conferences. The NMDC encourages all individuals who want to be even more involved to join the NMDC [Champions program](#). Champions help us beta test new initiatives, improve our products, and talk to others about metadata standards. We re-evaluated and improved our Champions program this past year based on feedback from those involved with the program. Relunched in April 2022, the Champions program has included more regular updates through newsletters and virtual events, more networking instances, and opportunities for joint authorship on scientific publications.



# Growing and evolving scientific communities

Over the course of 2022, NMDC team members attended 24 conferences and gave more than 60 presentations. Given ongoing COVID limitations, many of these conference presentations were virtual. Team members also strengthened ties with international collaborators by presenting at events outside of the United States, including a presentation at the [22nd Genomics Standards Consortium Workshop](#), a panel discussion during the [18th International Symposium on Microbial Ecology](#) and a hands-on workshop at the fourth annual meeting of the [Metabolomics Association of North America](#). We also hosted [a panel discussion](#) during the [AAAS annual meeting](#) focused on public data infrastructure for scientific discovery, bringing together experts to discuss indigenous data governance and partnerships across sectors to support data sharing. In June, the NMDC hosted a town hall at the [ASM Microbe](#) conference, featuring a keynote from ASM Microbe Data Prize Awardee Patrick Schloss and showcasing three NMDC Ambassadors. The [NSURP](#), a community-driven initiative to create summer research opportunities for Black, Indigenous, and People of Color (BIPOC), also featured a workshop on [NMDC EDGE](#).

NMDC members joined with other microbiome researchers throughout the year to imagine new directions for the future of microbiome research. In April 2022, the American Academy of Microbiology (AAM)

released its new Climate Change and Microbes Scientific Portfolio, where NMDC team members contributed to drafting the colloquium report "[Microbes & Climate Change: Science, People, & Impacts](#)" and the companion perspective, "[Microbes and Climate Change: a Research Prospectus for the Future.](#)" Following this, NMDC team members worked closely with the ASM to support two separate "Hill Days." During each Hill Day, they met with policymakers to discuss the current state of microbiome research and the importance of data stewardship. For the first Hill Day in July, NMDC team members talked with policymakers about the importance of microbes in climate change adaptation, mitigation, and resilience as part of the AAM's Climate Change and Microbes Scientific Portfolio. Then, for a subsequent Hill Day in December, NMDC team members met with the offices of several elected officials from multiple states to educate them on the importance of microbiomes and microbial research. This Hill Day was co-organized by the Microbiome Centers Consortium (MCC) and this particular opportunity strengthened relationships between the NMDC, the MCC, the ASM, and local elected officials.

# NMDC Ambassadors

[NMDC Ambassadors](#) are early career researchers who are motivated to engage with their research communities to support the generation of FAIR microbiome data. Ambassadors are a key part of NMDC's goal of creating culture change to encourage open science, data stewardship, and the use of standards in microbiome research.

The inaugural cohort of NMDC Ambassadors grew their technical knowledge about metadata standards by attending training workshops with the NMDC team and metadata experts. Over the course of nine months, the Ambassadors organized 23 events, including workshops and presentations. Through their engagement activities, they reached a combined audience of more than 800 people.

The Ambassadors also gave input that helped the NMDC team improve specific MlxS environmental packages and improve the GOLD ecosystem classification system. The Ambassadors broadened the NMDC community network by reaching more researchers than would have been possible by NMDC team members alone.

Feedback received from the Ambassadors in 2022 was collated by Center for Scientific Collaboration and Community Engagement (CSCCE) and shared with the NMDC team and Ambassador cohort. This feedback has been instrumental in developing new programmatic directions and implemented into the plans for the 2023 [Ambassador program](#), which will be launched in January of 2023.

## MEET OUR AMBASSADORS



Arwa Abbas



Mikayla Borton



Emily Davenport



Natalia Erazo



Chloe Herman



Lisa Karstens



Brandon Kocurek



Holly Lutz



Kevin Myers



Jaci Saunders



Emily Vogtmann



Amanda Windsor

# NMDC Champions

[NMDC Champions](#) understand and appreciate the value of well-curated data. They are willing to advocate for the importance of FAIR microbiome data. NMDC Champions are the first in line to provide feedback on the usability of NMDC platforms, training materials, and many other external-facing project initiatives. In 2022, our Champions' contributions resulted in important features and enhancements on the NMDC [Data Portal](#) and [NMDC EDGE](#). Several Champions

were also invited to participate in a number of opportunities in 2022, including the addition of a Champion to the NMDC Scientific Advisory Board and the inclusion of Champions in workshops and the NMDC [Community Conversations](#). The program was relaunched in April of 2022 to build upon the existing opportunities with clearer goals and greater benefits to the Champions, including additional networking and publishing opportunities.

## MEET OUR CHAMPIONS



Yigal Achmon



Kai Blumberg



John-Marc Chandonia



Sean Cleveland



Sneha Couvillion



Joan Damerow



Justine Debelius



Bill Duncan



JP Dundore-Arias



Cassie Ettinger



Hannah Freund



Alexis Garretson



Sean Gibbons



Buck Hanson



Judson Hervey



Alex Honeyman



Bonnie Hurwitz



Ulas Karaoz



Marie Kroeger



Kate Lane



Jessica Audrey Lee



Ryan McClure



Nancy Merino



Jason Rothman



Ahmed Shibl



Venkat Subramanian



Luke Thompson



Geizecler Tomazett



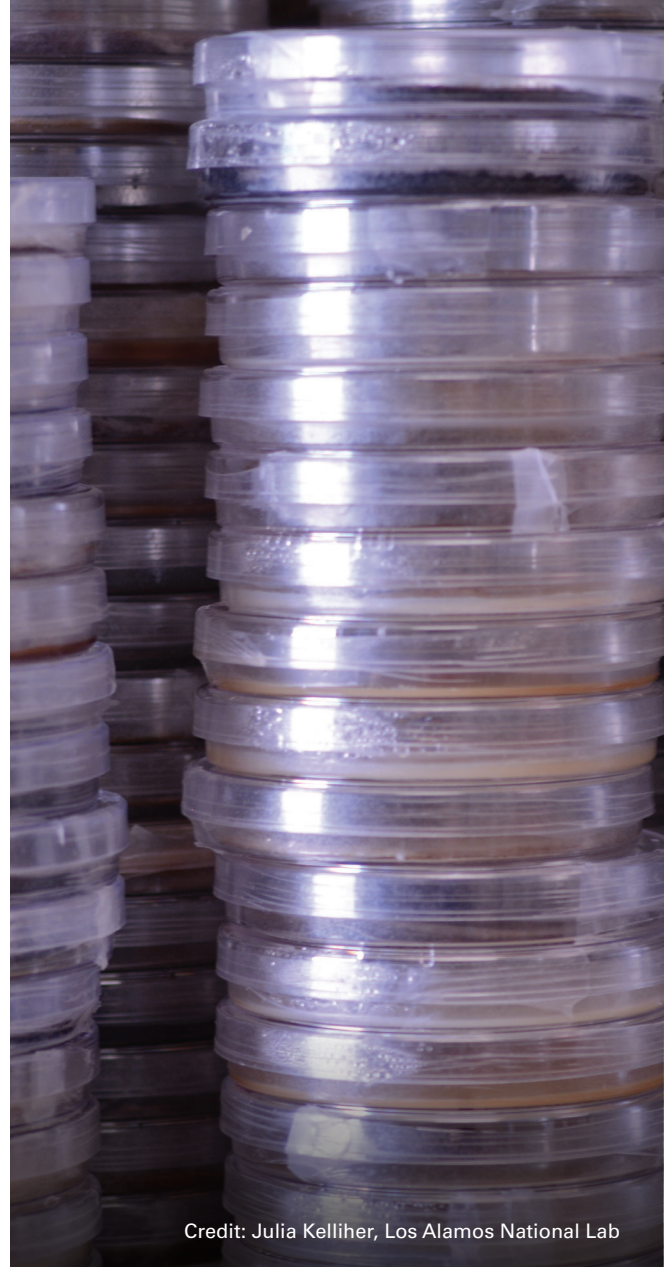
Alonna Wright



Ying Zhang

# NMDC Community Conversations Series

The [Community Conversations](#) series aimed to connect with the scientific community on a wide-ranging set of topics, including FAIR data and its applications in microbiome research, user-centered design for science gateways, data stewardship and inclusive data resources, and best practices for engagement in the scientific community. This monthly webinar series invited the broader research community to engage with the NMDC, learn about the breadth of NMDC program activities, and contribute feedback. Between 2021 and 2022, the NMDC team hosted eight online discussions with more than 20 different subject-matter specialists. These webinars reached nearly 550 community members, who learned about the usage of scientific data resources. Additionally, since they were uploaded onto the [NMDC's YouTube channel](#), these webinars have been viewed nearly 300 times.



## Integration Across DOE User Facilities and Computational Resources

Resource	Mission, Scope, and Capabilities	Integration with the NMDC
<p><a href="#">(IMG/M)</a> Joint Genome Institute, Berkeley Lab (LBNL)</p>	<p><b>Integrated Microbial Genomes &amp; Microbiomes:</b> A JGI resource for the integration of sequence data to support the annotation, analysis, and distribution of microbial genomes, metagenomes, and metatranscriptomes. <a href="#">Read more</a></p>	<p>The NMDC uses IMG/M's production-quality annotation workflow along with the JGI's quality control and assembly components. Processed metagenome and metatranscriptome data is accessible from IMG/M under the <a href="#">Collaborations menu</a>. There are 15 studies with shared data across the IMG/M platform and NMDC <a href="#">Data Portal</a>, which are accessible by navigating to the <a href="#">Study List page</a> within IMG/M.</p>
<p><a href="#">GOLD</a> Joint Genome Institute, Berkeley Lab (LBNL)</p>	<p><b>Genomes OnLine Database:</b> A JGI resource for comprehensive access to information regarding genome, metagenome, and metatranscriptome sequencing projects, and their associated metadata. <a href="#">Read more</a></p>	<p>The NMDC has directly contributed to the enhancement of GOLD metadata through updated EnvO curation to complement GOLD's five-level Ecosystem Classification to be fully compliant with the GSC's MlxS standards. More than 63,000 biosamples have been curated with at least one EnvO term. The NMDC programmatically accesses data through the GOLD API.</p>
<p><a href="#">EnvO</a> Open Biological and Biomedical Ontology Foundry, Berkeley Lab</p>	<p><b>Environment Ontology:</b> A community-driven ontology for the concise, controlled description of environments. <a href="#">Read more</a></p>	<p>The NMDC search functionality uses a faceted approach including EnvO annotation. The use of EnvO terms in the <a href="#">Data Portal</a> facilitates dataset aggregation from different sources and increases the findability and interoperability of the data. The NMDC does programmatically access the GOLD API.</p>
<p><a href="#">KBase</a> Berkeley Lab (LBNL)</p>	<p><b>DOE Systems Biology Knowledgebase:</b> A software and data science platform designed to meet the grand challenge of systems biology: predicting and designing biological function. KBase integrates data and tools in a unified graphical interface. <a href="#">Read more</a></p>	<p>Many of the metagenome datasets from six studies available through the <a href="#">Data Portal</a> are also now available within <a href="#">KBase</a> for downstream advanced analyses. Users can access NMDC data to create a KBase Narrative, an interactive digital notebook that can be shared with the scientific community. Future work will support streamlined data exchange to enable custom analyses in KBase, including metabolic modeling.</p>

Resource	Mission, Scope, and Capabilities	Integration with the NMDC
<p><b>EDGE</b> Los Alamos National Laboratory (LANL)</p>	<p><b>Empowering the Development of Genomics Expertise:</b> A bioinformatics platform that can be installed anywhere providing a user-friendly web-based interface to enable novice microbiome and bioinformatics users to process and explore the analysis of raw multi-omics data. <a href="#">Read more</a></p>	<p>The NMDC workflows have been integrated into the EDGE web-based interface for point-and-click access to data processing, along with the newly added virus and mobile genetic element identification workflow. <a href="#">NMDC EDGE</a> is hosted at the San Diego Super Computer (SDSC) to allow any user to run their own multi-omics data through the available workflows.</p>
<p><b>NEXUS</b> Pacific Northwest National Laboratory (PNNL)</p>	<p><b>Network Execution of User Science:</b> A central data repository for semi-automated capture and storage of data generated by EMSL staff and users, including the capture of basic metadata associated with instruments and user proposals.</p>	<p>The NEXUS repository is the primary data provider of metaproteomics, NOM, and metabolomics data to the NMDC Data Portal. The data standards, ontologies, and standardized workflows developed by NMDC and EMSL for those data types will enable interoperability with additional facilities.</p>
<p><b>ESS-DIVE</b> Berkeley Lab (LBNL)</p>	<p><b>Environmental Systems Science Data Infrastructure for a Virtual Ecosystem:</b> A repository of DOE-generated earth science data, including observational, experimental, and modeling research.</p>	<p>The NMDC team has identified studies that are supported within the ESS-DIVE archive, and this past year implemented links from the <a href="#">Data Portal</a> to ESS-DIVE data to associate microbiome data and complementary environmental measurements. This coming year, we plan to work closely with the ESS-DIVE team to expand data exchange services and infrastructure links.</p>



# Partnership Highlights



AMERICAN  
SOCIETY FOR  
MICROBIOLOGY

## American Society for Microbiology (ASM)

The NMDC team has a longstanding partnership with the ASM, beginning with the ideation of the NMDC and program development through its program launch in 2019 at the ASM's annual conference. Our multifaceted partnership includes involvement in the ASM's Microbiome Stakeholder Group and participation of the ASM's CEO, Stefano Bertuzzi, on the NMDC Scientific Advisory Board. This past year, the NMDC partnered with the ASM for several key engagement activities, including hosting a town hall at the ASM Microbe conference. This portion of the program included a special welcome by the ASM's CEO and NMDC SAB member Bertuzzi and a keynote by ASM Microbiome Data Prize Awardee Patrick Schloss, and also showcased the work of three NMDC Ambassadors.



## Center for Scientific Collaboration and Community Engagement (CSCCE)

The CSCCE is a research and training center to support and study the emerging field of scientific community engagement, and has partnered with the NMDC team since the initiation of the NMDC program. CSCCE has provided valuable training for the NMDC team and Ambassadors on best practices for community engagement and has provided a framework for planning, coordinating, and implementing engagement activities. With a shared vision of creating an engaged scientific community, the CSCCE has been instrumental in the development and success of our Ambassador and Champion programs, and NMDC's continued work with the microbiome research community. At the end of the 2022 Ambassador Program, CSCCE collected feedback received from the Ambassadors. This feedback was shared with the NMDC team and was critical to enhancing the future plans of the Ambassador program.



## DMPTool California Digital Libraries (DMPTool)

In partnership with the [University of California Curation Center](#) of the [California Digital Library](#), the NMDC team has created a microbiome-specific DMPTool Template. [DMPTool](#) is an open-source application that assists researchers in the creation of data management plans compliant with federal funding requirements. The NMDC DMPTool template supports microbiome data management best practices with specifications unique to microbiome standards and data processing. This template will help researchers create data management plans that lead to data being FAIR.



## Genomic Standards Consortium ([GSC](#))

As standards underpin all NMDC work, our team has a longstanding partnership with the GSC, a community-based organization focused on developing and maintaining internationally recognized standards for genomic data. The GSC's President, Lynn Schriml, serves on the NMDC Scientific Advisory Board. Several members of the NMDC team are heavily involved in the GSC's Compliance and Interoperability Group to support development and integration, along with NMDC Lead Eloë-Fadrosh serving on the GSC Board. This past year, the NMDC team worked closely with the GSC to support adoption of the LinkML implementation of the MIxS standards and supported the new version release.



## International Microbiome and Multi-Omics Standards Alliance ([IMMSA](#))

IMMSA, led by the [National Institute of Standards and Technology](#) (NIST), supports coordinating cross-cutting efforts that address microbiome measurement challenges and broadly engage microbiome-focused researchers from industry, academia, and government. NMDC Lead Eloë-Fadrosh serves on the IMMSA Steering Committee, and several NMDC team members are actively engaged in the working groups on bioinformatics tools, workshop planning, and the newly formed metabolomics group. This past year, NMDC team members presented at the 2022 NIST-hosted IMMSA workshop.



## Microbiome Centers Consortium ([MCC](#))

The MCC is a cooperative and collaborative network of microbiome centers led by Jennifer Martiny at UC Irvine. Martiny serves on the NMDC Scientific Advisory Board. This past year, MCC and NMDC team members participated together during the ASM's Microbiome Hill Day to educate policymakers on the importance of microbiome and microbial research. The NMDC team also presented at the second MCC meeting in March 2022 to further collaborative discussions and explore how to increase cross-agency work.



For more information, contact:  
**Ingrid Ockert,**  
Communications and Outreach Specialist  
iockert@lbl.gov  
Design by Creative Services, Berkeley Lab



In November 2022, NMDC staff attended a two-day retreat in Napa, California. The 34 team members from PNNL, LANL, and LBNL brainstormed implementation strategies and planned 2023 milestones.  
Credit: Wendi Lynch



A series of photos from the National Microbiome Data Collaborative. Clockwise from top:  
Credit: Jeremy Snyder; Credit: Julia Kelliher, LANL; Credit: Jeremy Snyder.





**nmdc**

National Microbiome  
Data Collaborative



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

[microbiomedata.org](https://microbiomedata.org)  [@microbiomedata](https://twitter.com/microbiomedata) [#NMDc](https://www.linkedin.com/company/microbiomedata/)  <https://www.linkedin.com/company/microbiomedata/>