Environmental Microbiome

MEETING REPORT Open Access

Microbiome data management in action workshop: Atlanta, GA, USA, June 12–13, 2024

Julia M. Kelliher[1*], Mashael Aljumaah[2,3], Sarah R. Bordenstein[4,5], J. Rodney Brister[6], Patrick S. G. Chain[7], Jose Pablo Dundore-Arias[8], Joanne B. Emerson[9], Vanessa Moreira C. Fernandes[10], Roberto Flores[11], Antonio Gonzalez[12], Zoe A. Hansen[13], Eneida L. Hatcher[6], Scott A. Jackson[14], Christina A. Kellogg[15], Ramana Madupu[16], Cassandra Maria Luz Miller[17], Chloe Mirzayi[18], Ahmed M. Moustafa[19,20,21], Christopher Mungall[22], Aaron Oliver[23], Nonia Pariente[24], Jennifer Pett-Ridge[25,26,27], Sydne Record[28], Linta Reji[29], Anna-Louise Reysenbach[30], Virginia I. Rich[31], Lorna Richardson[32], Lynn M. Schriml[33], Reed S. Shabman[34], Maria A. Sierra[35], Matthew B. Sullivan[31], Punithavathi Sundaramurthy[36], Katherine M. Thibault[37], Luke R. Thompson[38,39], Scott Tighe[40], Ethell Vereen[41] and Emiley A. Eloe-Fadrosh[22*]

Abstract

Microbiome research is revolutionizing human and environmental health, but the value and reuse of microbiome data are significantly hampered by the limited development and adoption of data standards. While several ongoing efforts are aimed at improving microbiome data management, significant gaps still remain in terms of defining and promoting adoption of consensus standards for these datasets. The *Strengthening the Organization and Reporting of Microbiome Studies* (STORMS) guidelines for human microbiome research have been endorsed and successfully utilized by many research organizations, publishers, and funding agencies, and have been recognized as a consensus community standard. No equivalent effort has occurred for environmental, synthetic, and non-human host-associated microbiomes. To address this growing need within the microbiome research community, we convened the *Microbiome Data Management in Action* Workshop (June 12–13, 2024, in Atlanta, GA, USA), to bring together key decision makers in microbiome science including researchers, publishers, funders, and data repositories. The 50 attendees, representing the diverse and interdisciplinary nature of microbiome research, discussed recent progress and challenges, and brainstormed actionable recommendations and paths forward for coordinated environmental microbiome data management and the modifications necessary for the STORMS guidelines to be applied to environmental, non-human host, and synthetic microbiomes. The outcomes of this workshop will form the basis of a formalized data management roadmap to be implemented across the field. These best practices will drive scientific innovation now and in years to come as these data continue to be used not only in targeted reanalyses but in large-scale models and machine learning efforts.

*Correspondence:
Julia M. Kelliher
jkelliher@lanl.gov
Emiley A. Eloe-Fadrosh
eaeloefadrosh@lbl.gov

Full list of author information is available at the end of the article

Keywords Microbiome, Environmental Microbiome, Standards, Data management, Data stewardship, Data reuse, Checklist, Guidelines, Workshop

Introduction

The field of microbiome research is rapidly growing with innovations spanning human, animal, plant, marine, and soil health, and implications for ecosystem resilience, nutrient cycling, food security, and environmental responses to extreme conditions [1–3]. Institutions, companies, and government agencies around the world have collectively invested billions of dollars into understanding human microbiomes and their relationships to health and disease [4–6]. The importance of environmental, non-human host, built environment, and synthetic microbiomes and their contributions to ecosystems is also broadly acknowledged within the research community, but the investment, infrastructure, and public recognition have historically lagged behind human-associated microbiomes [7]. However, this field is also experiencing a rapid rate of growth as the sense of urgency for preserving environmental health in the face of changing conditions becomes increasingly pertinent.

Researchers study microbiomes using techniques that interrogate their composition and function, for example, by examining their associated DNA, RNA, proteins, and chemical signatures. Combinations of these investigations, known as multi-omics, require the generation of large data files that are difficult to produce, store, process, and share. These data are also produced by a range of methods and instruments, such that data produced by different researchers or institutions may not be directly cross-comparable or interoperable. The generation of microbiome data has vastly outpaced the development of data management infrastructure and consensus reporting standards. This disconnect hinders data reuse, including in meta-analyses and large-scale modeling efforts. Despite efforts to improve microbiome data management [8–12], these data are often not generated, stored, and shared according to the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles that emphasize data reuse [13], which limits the growth of this field and compounds existing research siloes.

In 2021, a group of human microbiome researchers collectively released the *Strengthening The Organization and Reporting of Microbiome Studies* (STORMS, https://www.stormsmicrobiome.org) checklist which outlines reporting guidelines for human microbiome research publications [14]. This consensus checklist has been rapidly and widely adopted, and is already recognized as integral to promoting standardization across human microbiome and epidemiological data and studies. Data management across non-human host-associated microbiomes, the environmental sciences, and synthetic communities is in need of a STORMS-like reporting framework. However, across these diverse domains, it is currently challenging to consistently understand even basic study design or environmental parameters. While the existing Minimum Information about any (x) Sequence (MIxS) standard [15] supports sequence-specific metadata specifications including environmental ontologies, these do not encompass the minimal requisite information about microbiome datasets (e.g., study and sampling design, sample preparation or synthetic material construction, and other essential parameters to enable cross-study comparisons and data reuse). Significant gaps also remain in the adoption of existing standards and their application across diverse environment types and also multi-omics data [16].

The *Microbiome Data Management in Action* Workshop was convened from June 12–13, 2024, in Atlanta, GA, USA. The workshop was funded by the National Science Foundation (NSF) and was closely coordinated with the organizers of the subsequent *American Society for Microbiology Microbe* conference. The workshop aimed to address the increasing need for a robust data management ecosystem for microbiome research following the recommendation of the 2018 Interagency Strategic Plan for Microbiome Research [17] "to support the development of platform technologies to enhance data access and sharing." The workshop brought together researchers, publishers, funders, and data repository representatives to identify local and national priorities. The workshop sought to tackle the theme of data management broadly, with a more specific focus on standardized reporting guidelines. The workshop attendees began generating a roadmap for microbiome data standards across non-human host-associated microbiomes, the environmental sciences, and synthetic communities leveraging the framework and lessons learned from the STORMS reporting guidelines.

Microbiome Data Management in Action Workshop

Meeting goals and objectives

The workshop goal was to ultimately advance microbiome science through coordinated data management across researchers, funders, data repositories, and publishers. The workshop itself was designed to be the first step in leveraging the STORMS reporting checklist to develop a consensus roadmap to encompass non-human host-associated microbiomes, the environmental sciences, and synthetic communities. The workshop was designed to lead to actionable insights with concrete paths forward rather than open-ended discussions.

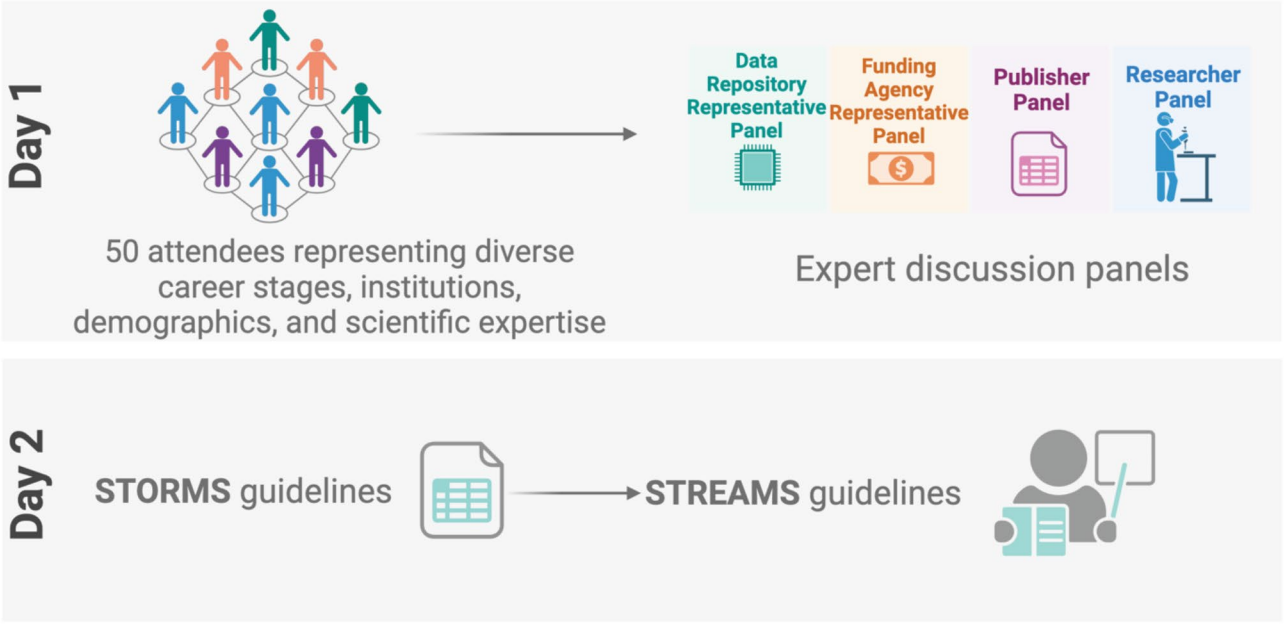

Fig. 1 Overview of the workshop

Key questions

The workshop aimed to address the following key questions:

1. Where are we now? What is the current ecosystem for microbiome multi-omics data management, standards, and data re-use? What have we learned from past efforts?
2. What actions are needed from researchers, funders, data repositories, and publishers to advance data management beyond current practices?
3. What developments, technical and otherwise, are required to achieve the standards, best practices, community adoption, and data stewardship needed?

Attendees

To capture the interdisciplinary nature of microbiome research, the 50 attendees were selected to represent diverse experiences and perspectives and with an understanding that participation of groups typically underrepresented in science and engineering is critical to advancing microbiome research. For example, the attendees ranged in their career stage (student to senior scientist), science expertise (e.g., bioinformatics, ecology, multi-omics), studied environment (e.g., aquatic, terrestrial, and animal-associated microbiomes), gender, sexual orientation, ethnicity, background, sector (e.g., government, academia), institution (included representatives from several Minority Serving Institutions (MSIs)), and geographic location (included several international representatives). Further, follow-on meetings and active recruitment and engagement with the broader microbiome research community is planned to ensure community consensus in building a collaborative data management roadmap.

Meeting format & logistics

The workshop was organized and run by a steering committee of five microbiome researchers spanning academia, national laboratories, government agencies, and international consortia. The workshop took place at the Georgia World Congress Center in Atlanta, Georgia, USA on June 12–13, prior to the *American Society for Microbiology (ASM): Microbe* conference. The workshop included a virtual participation option for invitees who could not attend in person. In-person attendees were offered reimbursement for their travel expenses to increase inclusion and participation. Name tags and pronoun stickers were provided to in-person attendees. Guidelines for discussions and a code of conduct were provided to ensure respectful and safe participation. Steering committee members monitored the virtual chat for questions, and a live notes document captured participant notes, questions, links, and comments.

The workshop was divided into four main sections, described in detail below: (1) Welcome & Workshop Overview; (2) Current State of Microbiome Data Management & A Path Towards Data Stewardship Across Stakeholders; (3) Expert Panels; and (4) Breakout Sessions (Fig. 1, https://doi.org/10.5281/zenodo.13760811).

Welcome & workshop overview

In the first presentation of the event, Principal Investigator Kelliher from the New Mexico Consortium presented

on the meeting objectives and deliverables, clarifying that the workshop was intended to lead to outputs and actions (https://doi.org/10.5281/zenodo.13760905). She emphasized the interdisciplinary nature of microbiome research and why inclusion was prioritized for this workshop. While the workshop was designed to be inclusive, it was also important to the organizers that the power imbalances in the room be recognized, and the organizers offered suggestions on how to mitigate the effects of these imbalances. The opening presentation set a context of maintaining a positive, collaborative, non-judgmental environment, and emphasized interest in hearing from everyone to facilitate open communication and effective collaboration. This presentation also highlighted the follow-on ASM mini-conference session: *EEB-MC-001: Microbial Data and Tools without Borders: Advancing an Open Science Ecosystem*, on June 13th, which included a presentation to the ASM community on the workshop takeaways to obtain preliminary feedback from the microbiome research community regarding these efforts. The overview presentation ended with participant introductions of each attendee's name and affiliation.

Current state of microbiome data management

Steering committee member and first author on the STORMS reporting guidelines publication Dr. Mirzayi from CUNY Graduate School of Public Health and Health Policy presented on the process of creating the STORMS guidelines, including timelines, lessons learned, and recommendations for how to adapt these guidelines to non-human research. Questions from attendees during this presentation led to discussions around five main topics: (1) the future of STORMS updates; (2) how to engage the larger microbiome research community when gathering feedback; (3) how to incorporate conflicting feedback into a consensus statement; (4) how to increase awareness and adoption of standards; and (5) how guidelines can be effectively enforced by funding agencies or publishers.

Expert panels

This workshop convened four unique expert panels that allowed for targeted questions regarding each group's roles in data management to be asked by the panel moderator and the audience of participants. All panelist answers represented their personal views and did not necessarily represent the views of their associated institutions.

Data repository representative panel

Five representatives from four data repositories and community resources - the European Bioinformatics Institute (EBI), the National Center for Biotechnology Information (NCBI), the National Ecological Observatory Network (NEON), and Lawrence Berkeley National Laboratory (LBNL) - shared their experiences on how researchers currently deposit data into these repositories, what metadata are required, and how the datasets are linked with other contextual information or publications.

The panelists shared information about how well researchers are adhering to their repositories' standards, and how these existing standards may need to change. The panelists also discussed what they think next steps should be for data repository handling of data and how the repositories can work with the other groups in attendance, especially publishers, to promote standardization and proper data stewardship across the field of microbiome research.

Key takeaways

- Researchers feel like it can be extremely difficult, if not impossible, to properly deposit their data into repositories without making mistakes.
- Researchers would like more trainings, user guides, videos, etc. to walk them through the process of data deposition; this task often falls on students and postdocs.
- By the time researchers go to deposit their data, it is often too late to enforce standards, especially for metadata, since they will have already collected their data.
- Data creators are often excited about making their data open and FAIR, but don't always know how best to accomplish this.
- More obvious metrics for data reuse may be a good incentive; to enable metrics this requires development of a robust culture of data citation.
- Data repositories generally do not have an abundance of personnel time and resources to enforce standards or support community requests to use standards.
- Long-term sustainability is always a concern for every stakeholder group; it can be difficult to know which repositories will be maintained long-term, and how the landscape will change over time in terms of where people are depositing their multi-omics or other data types.

Funding agency representative panel

Five panelists represented four funding agencies: the National Science Foundation (NSF), the Department of Energy (DOE), the National Institutes of Health (NIH), and the National Oceanic and Atmospheric Administration (NOAA).

The panel included discussions on the funding agencies' current data management requirements for researchers, and how these could change to adopt new guidelines in specific research fields. The discussions also led to how interagency collaborations can be leveraged to generate more broadly impactful data management

guidelines and how data are and will be shared between groups and agencies. Discussions about specific funding for data management plan (required documentation by grant funding agencies) implementation and the potential for having designated data managers as funded positions within grant proposals also occurred.

Key takeaways

- It would be beneficial to have evaluation guidelines for proposal reviewers or agency staff members, especially for when they are evaluating proposal data management plans or the relevant text within a proposal outlining data management practices.
- Funding agencies have been providing more guidance for proposers writing data management plans, but it is difficult, and not perceived as within the funders' purview to provide detailed guidance across environment types, sample types, or other study-specific aspects.
- Several funders have started enforcing more frequent reporting on the progress of the data management plan throughout the project's lifecycle (e.g., in annual reports).
- Some funding agencies are allowing researchers to request money specifically for data management plan implementation.
- A uniform data management plan structure would be very difficult to implement across funding agencies, but creation of something more modular and not too prescriptive may work better.
- Having a minimum standard for a data management plan and making that broadly available and achievable is key.
- Many funding agency data management policies are based on the needs, asks, and best practices of the research community; however, this process cannot be solely 'bottom-up' (from the research community), more guidance is needed from funding agency perspective as well.

Publisher panel

This panel included four members representing *Springer Nature (Nature Microbiology)*, *Cell Press/Elsevier*, *Springer Nature (BioMed Central)*, and *PLOS*.

This panel included discussions about how publishers can promote microbiome data management best practices and how they can ensure compliance when researchers have submitted their manuscripts for review and inclusion in their respective journals. The panel brought up that journals cannot easily enforce standards, tools, or external checklists but can recommend options for authors and reviewers. The panelists also discussed the interconnectedness with standards for data repositories, as it is already established that many journals publishing microbiome research mandate open access of the data in data repositories. Discussions also took place about data reuse and what journals can do to ensure data is being properly cited and credited when reused in publications.

Key takeaways

- Journals want to make it as easy as possible for authors to submit; if a journal enforces too much formatting or too many standards (e.g., a checklist), authors may choose to submit elsewhere.
- Compliance checking is necessary, as without it, author behaviors do not change. A lot of compliance checking is still performed manually, and this requires personnel time and money that not all journals have. For example, reviewers often do not go through the data in excruciating detail (e.g., don't check all 500 accession numbers of data); artificial intelligence is projected to help with this, but still in earliest stages.
- It has been difficult to get authors to abide by data availability rules; many authors submit without their data being publicly available.
- Journals often want to see buy-in from the community first regarding a recommendation or standard (e.g., STORMS) before adding it to their lists of recommendations, and even more so to make a standard a requirement of publication.
- Researchers often think the journals hold more power in terms of enforcement of best practices than they actually do.

Researcher panel

The Researcher Panel consisted of four scientists representing a range of career stages, from student to senior scientist, all from academic institutions across the United States. This panel covered the responsibilities of researchers to adhere to data management best practices, and how a standard set of guidelines can be created to be reasonably achievable, without excessive burden on researcher's time and effort. Panelists were also tasked with suggesting advances they foresee in their research if data management barriers are eliminated, and potential pitfalls if best practices for data management are not generally followed.

Key takeaways

- Machine learning (ML) and artificial intelligence (AI) hold the potential to significantly help researchers in many aspects of data management.
- Education about the importance of data management and metadata standards needs to start earlier, as part

of undergraduate or graduate coursework, or at least before a research project starts; many researchers do not know what they should be doing until it is too late; there is currently too much redundancy in researchers having the same issues throughout the data management process.
- Researchers may also not be aware of current best practices; older methods may be repeatedly passed down without improvement.
- There has been a culture change that needs to continue regarding how researchers view FAIR data and data management; it was seen as a burden and necessary for publishing; it needs to be continuously reiterated that this benefits the field now and in the future.
- Collaboration across the field can be extremely helpful, but can be difficult with resource imbalances across institutions; collaborations and sharing of best practices, standards, and data should not be dependent on what another institution can offer you; inclusion of historically black colleges and universities (HBCUs), minority serving institutions (MSIs), and collaborations outside of typical networks can be beneficial for everyone involved; it is important to develop standards that can work across a variety of groups and institutions.
- Standards need to be considered for every step of the data lifecycle including sample collection, wet laboratory work, bioinformatics analyses, and data deposition.

Day 2: breakout groups

On the second day, workshop attendees were divided into six groups, with each containing six to eight individuals. The virtual attendees formed one of these groups. In-person attendees were asked to sit with different people than the previous day and self-organize to evenly distribute career stage and affiliation. The breakout groups were tasked with adapting the STORMS guidelines to environmental, synthetic community, and non-human host-associated microbiome studies. It was explained to participants that these recommendations will be collated, and the new guidelines will be circulated throughout the microbiome research community for feedback before a consensus statement is created. Once the reporting checklist is developed, it is envisioned to be integrated into data management plan guidance, reviewer guidelines, an author checklist, and overall guidelines for the field.

Attendees were also tasked with discussing how to increase community adoption, how to communicate incentives, and how to educate others regarding standards and data management, with an emphasis on tangible actions and outputs. The breakout groups also discussed what the next steps for this group should be and how they would like to be involved moving forward. Each group reported back to the larger group and kept their notes in a shared notes document so that all groups could see the discussion items and outputs from other groups.

Lessons learned

The structure and logistics described here are intended to assist other groups in organizing a workshop of similar size and scope. As has been noted throughout the COVID-19 pandemic, a hybrid format for a workshop presents its own unique set of challenges. While it was ideal to be able to offer virtual attendance for those who were unable to attend in person, having more cameras and more microphones in the room would have made it easier for virtual participants to follow the discussions and feel more included as a part of the larger group. Additionally, a more effective audio-visual set-up for virtual attendees would have allowed us to invite more virtual participants to the workshop.

The breakout groups were extremely effective for moving towards actionable insights rather than discussions, which was a goal for this workshop. Additional time on the second day for these breakout groups would have been preferable, as more questions could have been asked of the groups. The engagement during these breakout groups was outstanding, and there were even some breakout groups within breakout groups working on specific terms or a smaller subset of action items. For the breakout groups, we found it beneficial to have attendees sit with other participants they did not know very well, different people from the previous day, and according to the sector they identify with (funding agency representative, researcher, publisher, etc.). This allowed for productive conversations across various aspects of the microbiome research field, although it is recognized that assigning a set of different tasks for each breakout would have minimized duplicate work across the groups.

Next steps

A primary focus of this workshop was to start conversations with the intention of generating high-impact outputs and sustained collaborations for the future. Immediately following the workshop, the insights, outcomes, and the idea of a consensus roadmap was presented to the larger microbiology research community as part of an open session *EEB-MC-001: Microbial Data and Tools without Borders: Advancing an Open Science Ecosystem* at the ASM Microbe conference as a means to gain additional diverse perspectives, keep researchers informed of standardization efforts, and invite community feedback. There were approximately 150 attendees for this open session.

Further, the workshop steering committee plans to recruit and engage with the broader research community.

Many of the workshop participants expressed eagerness to continue to meet regularly as a working group to advance the reporting guidelines and data management best practices. The proposed guidelines will be distributed throughout the microbiome research community to gather ideas, feedback, and comments that will be implemented into the consensus checklist. This will ensure that we are capturing the diverse needs of the community and generating standards that are reasonably achievable. This workshop was also intended to be a launch point for a voluntary interagency working group of attendees and any other interested parties that will meet virtually to discuss the finalization of the guidelines, and discuss next steps and follow-on activities.

For scientific outputs, the workshop participants are working towards a consensus reporting checklist for environmental, synthetic community, and non-human host microbiome data modified from the STORMS checklist. This new reporting checklist, termed Standards for Technical Reporting in Environmental and host-Associated Microbiome Studies (STREAMS, https://streamsmicrobiome.org), will be broadly shared with the research community over the next year to solicit feedback and build consensus. Application of these guidelines to data management plan development, scientific manuscript submission, data submission to repositories, and proposals and publications review, will be developed to help integrate the checklist into key activities surrounding microbiome data streams.

Educational materials, outreach efforts, and strategies for lowering the barrier to adoption of microbiome reporting standards will also be created, shared, and implemented, as existing standards and guidelines in microbiology can be difficult to find, interpret, and adhere to. Several workshop attendees indicated their interest in developing this educational and instructional content, and a sub-group may be formed.

Conclusions

The *Microbiome Data Management in Action* Workshop assembled fifty participants across the microbiome research field to begin creating a consensus set of guidelines for environmental, synthetic community, and non-human host-associated microbiome data management and reporting. The participants emphasized that these goals require immediate action from every facet of microbiome research, and everyone has a responsibility to do their part in improving microbiome data management. The outputs from this workshop will assist in moving the field towards more FAIR microbiome data across agency borders, working to make the microbiome research field more equitable and inclusive. Data management best practices will benefit scientific innovation now and in years to come, as these data continue to be used in large-scale models and machine learning efforts. The outcomes of this workshop are intended to directly lead to an increased ability to more consistently and comprehensively manage microbiome data to enable reuse, thus unlocking research opportunities for all researchers, and improving the microbiome data ecosystem. This will positively impact environmental, synthetic community, and non-human host-associated microbiome research, which have demonstrated their importance to ecosystem health, food security, and environmental change.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Award No. 2422717. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We would like to extend our deepest gratitude to the American Society for Microbiology for support during the Microbiome Data Management in Action workshop, and especially Amalia Corby and Nguyen Nguyen for their participation.The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. This material is based in part upon work supported by the National Science Foundation through the NEON Program. Work conducted at LLNL was conducted under the auspices of the US Department of Energy under Award SCW1632 and Contract DE-AC52-07NA27344. This work was supported in part by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health. Any opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect the views of the U.S. Government or the affiliated institutions.

Author contributions
J.K. and E.A.E-F. wrote the workshop report with input from all authors. M.A. created Figure 1. All authors contributed, read, and approved the final version of the manuscript.

Funding
This material is based upon work supported by the National Science Foundation under Award No. 2422717.

Data availability
The workshop agenda (https://doi.org/10.5281/zenodo.13760811) and introductory slides (https://doi.org/10.5281/zenodo.13760905) have been made publicly available.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]New Mexico Consortium, Los Alamos, NM, USA
[2]Center for Gastrointestinal Biology and Disease (CGIBD), School of Medicine, UNC Microbiome Core, University of North Carolina, Chapel Hill, NC, USA
[3]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia
[4]Departments of Biology & Entomology, Pennsylvania State University, University Park, PA, USA
[5]One Health Microbiome Center, Pennsylvania State University, University Park, PA, USA

[6]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
[7]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA
[8]Department of Biology and Chemistry, California State University, Monterey Bay, Seaside, CA, USA
[9]Department of Plant Pathology, University of California Davis, Davis, CA, USA
[10]Department of Biological Sciences, Florida Atlantic University, Boca Raton, FL, USA
[11]National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
[12]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA
[13]Department of Plant Pathology, University of Minnesota, Saint Paul, MN, USA
[14]National Institute of Standards and Technology, Gaithersburg, MD, USA
[15]U.S. Geological Survey, St. Petersburg Coastal and Marine Science Center, St. Petersburg, FL, USA
[16]Biological Systems Science Division, Department of Energy, Office of Biological & Environmental Research, Germantown, MD, USA
[17]Department of Biology, University of New Mexico, Albuquerque, NM, USA
[18]CUNY Graduate School of Public Health and Health Policy, Institute for Implementation Science in Public Health, New York, NY, USA
[19]Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[20]Center for Microbial Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[21]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[22]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[23]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
[24]Public Library of Science, San Francisco, CA, USA
[25]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA
[26]Life & Environmental Sciences Department, University of California Merced, Merced, CA, USA
[27]Innovative Genomics Institute, University of California Berkeley, Berkeley, CA, USA
[28]Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, ME, USA
[29]Department of Geophysical Sciences, University of Chicago, Chicago, IL, USA
[30]Center for Life in Extreme Environments, Portland State University, Portland, OR, USA
[31]Center of Microbiome Science, The Ohio State University, Columbus, OH, USA
[32]EMBL-European Bioinformatics Institute, Hinxton, UK
[33]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[34]Office of Data Science and Emerging Technologies, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA
[35]Tri-Institutional Computational Biology and Medicine Program, Weill Cornell Medicine, New York, NY, USA
[36]Benioff Center for Microbiome Medicine, University of California San Francisco, San Francisco, CA, USA
[37]National Ecological Observatory Network, Battelle, Boulder, CO, USA
[38]Northern Gulf Institute, Mississippi State University, Starkville, MS, USA
[39]Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL, USA
[40]Vermont Integrative Genomics, University of Vermont, Burlington, VT, USA
[41]Science, Technology, Engineering and Mathematics Division, Morehouse College, Atlanta, GA, USA

Received: 17 January 2025 / Accepted: 8 April 2025
Published online: 19 April 2025

References

1. Alivisatos AP, Blaser MJ, Brodie EL, Chun M, Dangl JL, Donohue TJ, et al. A unified initiative to Harness Earth's microbiomes. Science. 2015. https://doi.org/10.1126/science.aac8480.
2. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551:457–63.
3. 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.
4. Stulberg E, Fravel D, Proctor LM, Murray DM, LoTempio J, Chrisey L, et al. An assessment of US Microbiome research. Nat Microbiol. 2016;1:15015.
5. Integrative HMP (iHMP) Research Network Consortium. Integr Hum Microbiome Project Nat. 2019;569:641–8.
6. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13:260–70.
7. Anthony WE, Allison SD, Broderick CM, Chavez Rodriguez L, Clum A, Cross H, et al. From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology. Environ Microbiome. 2024;19:56.
8. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, et al. Qiita: rapid, web-enabled Microbiome meta-analysis. Nat Methods. 2018;15:796–8.
9. Oliveira FS, Brestelli J, Cade S, Zheng J, Iodice J, Fischer S, et al. MicrobiomeDB: a systems biology platform for integrating, mining and analyzing Microbiome experiments. Nucleic Acids Res. 2018;46:D684–91.
10. Eloe-Fadrosh EA, Ahmed F, Anubhav A, Babinski M, Baumes J, Borkum M, et al. Nucleic Acids Res. 2021;50:D828–36. The National Microbiome Data Collaborative Data Portal: an integrated multi-omics microbiome data resource.
11. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, et al. MGnify: the Microbiome sequence data analysis resource in 2023. Nucleic Acids Res. 2023;51:D753–9.
12. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system V.7: content updates and new features. Nucleic Acids Res. 2023;51:D723–32.
13. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
14. Mirzayi C, Renson A, Genomic Standards Consortium, Massive Analysis and Quality Control Society, Zohra F, Elsafoury S et al. Reporting guidelines for human microbiome research: the STORMS checklist. Nat Med. 2021;27:1885–1892.
15. 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;29:415–20.
16. Vangay Pajau B, Josephine J, Anjanette, Beck Kristen L, Berrios Daniel C, Kai B, et al. Microbiome metadata standards: report of the National Microbiome data collaborative's workshop and Follow-On activities. mSystems. 2021;6. https://doi.org/10.1128/mSystems.01194-20
17. Interagency strategic plan for microbiome research. FY 2018–2022. USDOE Office of Science (SC) (United States); 2018 Apr. https://doi.org/10.2172/1471707

Publisher's note