

Manifesto of the FAIR Microbiome Implementation Network

Introduction

The cross-cutting nature of microbiome research in environmental sciences, health, agriculture, energy, and natural and built environments requires the development of new solutions and community coordination to tackle grand challenges that will accelerate basic discovery and lead to transformative advances. The National Microbiome Data Collaborative (NMDC) aims to tackle infrastructure challenges in microbiome data science through developing a community-centric framework based on large-scale, collaborative partnerships leveraging unique capabilities, expertise, and resources available across four Department of Energy (DOE) National Laboratories (Lawrence Berkeley National Laboratory (LBNL), Los Alamos National Laboratory (LANL), Pacific Northwest National Laboratory (PNNL) and Oak Ridge National Laboratory (ORNL)). The vision of the NMDC is to **empower the research community to harness microbiome data exploration and discovery** through a collaborative and integrative data science system. The NMDC will address fundamental roadblocks in microbiome data science through implementation of **guiding principles to make data and services more automatically findable, accessible, interoperable, and reusable (FAIR)**. To realize this vision, our multi-Lab collaborative partnership will pilot an integrated, community-centric framework within to fully leveraged existing microbiome data science resources and high-performance computing systems available within the DOE complex for data access, integration, and advanced analyses.

Phase II of the NMDC will further build upon these foundational integrative capabilities to reach additional microbiome data providers and existing resources outside of the DOE complex, including but not limited to the biomedical research community, the agricultural sector and phytobiome initiative, and the built microbiome community, to name a few. The Collaborative will strengthen engagement with the international community, as well as extending into the private sector for exploring new technologies for microbiome analysis and characterization and partnering to develop microbiome applications.

Purpose of the Implementation Network

The FAIR Microbiome Implementation Network seeks to align and synergize the efforts of the NMDC with that of the GO FAIR community, leveraging the GO FAIR open and inclusive ecosystem of Implementation Networks for the microbiome research community (for example, the Chemistry IN, Metabolomics IN, Biodiversities INs). The NMDC believes that by fostering community support for FAIR data stewardship practices, we will greatly accelerate the discovery process and enable effective communication within the community on issues of critical importance. The NMDC seeks to leverage community-driven standards and ontologies, and develop linked data models that allow for discovery and access based on community-supplied scientific use cases. The purpose of this Implementation Network is to work with microbiome research communities to (1) formalize core and domain-specific microbiome ontologies that promote discovery and reuse, and (2) establish training on the NMDC data models that allow for broader dissemination of knowledge and compliance for both humans and machines.

Overarching Principle of Operation

We commit to comply with the Rules of Engagement of GO FAIR Implementation Networks.

Targeted Objectives for the Internet of FAIR Data and Services (IFDS):

1. Foster and strategically oversee the adoption and expansion of FAIR data principles by the microbiome research community.
2. Consolidate and improve existing microbiome ontologies and standard vocabularies through broad engagement with the wider research community.
3. Make existing microbiome ontologies and standard vocabularies compliant with FAIR principles through broad engagement with the wider research community.
4. Communicate the NMDC implementation framework through the [FAIR convergence matrix](#) platform currently under development within the GO-FAIR community.
5. Develop NMDC data sharing infrastructure to enable distributed analysis accessible to the wider research community.

The Primary Tasks:

Task 1: Design the NMDC FAIR roadmap to align with the process of becoming a GO FAIR Implementation Network.

Task 2: Review current usage of standards across microbiome communities at the core data repositories. Identify target areas for curation, especially through automated validation, and improved compliance through training.

Task 2.5: Actively seek synergies between the FAIR Microbiome and other existing GO FAIR INs.

Task 3: Working with members of the GO TRAIN Pillar, develop communication tools and tutorials around FAIR data management best practices that enable accessibility and dissemination across the broader research community.

Task 4: Implement a framework for automated validation and metadata prediction by providing well-curated training data sets, ensuring ontologies and data models are consistent, and conducting extensive user testing across microbiome research domains to confirm edge cases are supported.

Membership list:

We consider this Manifesto to be one way by which the undersigned stakeholders can **speak with one voice** on a number of critical issues that are of generic importance to the objectives of FAIR, and on which we feel we have reached consensus.

National Microbiome Data Collaborative (NMDC)

Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas (CBGP) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Madrid, Spain

Timeline:

27 months

Date: November 11, 2019