



nmdc

National Microbiome
Data Collaborative

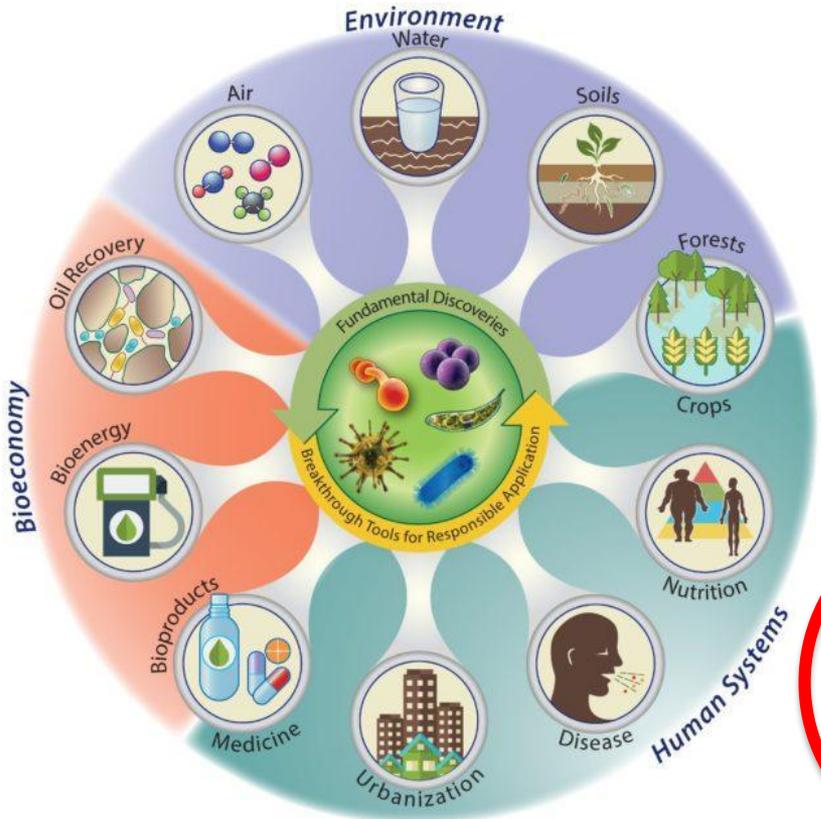
MAGIC meeting

5 Feb 2020

Kjiersten Fagnan

CIO Joint Genome Institute, Berkeley Lab

Cross-cutting nature of microbiome research



NSF
Interdisciplinary
microbiome
research



DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA
Environmental
microbiomes



NIH
National Institutes
of Health
Human
microbiomes

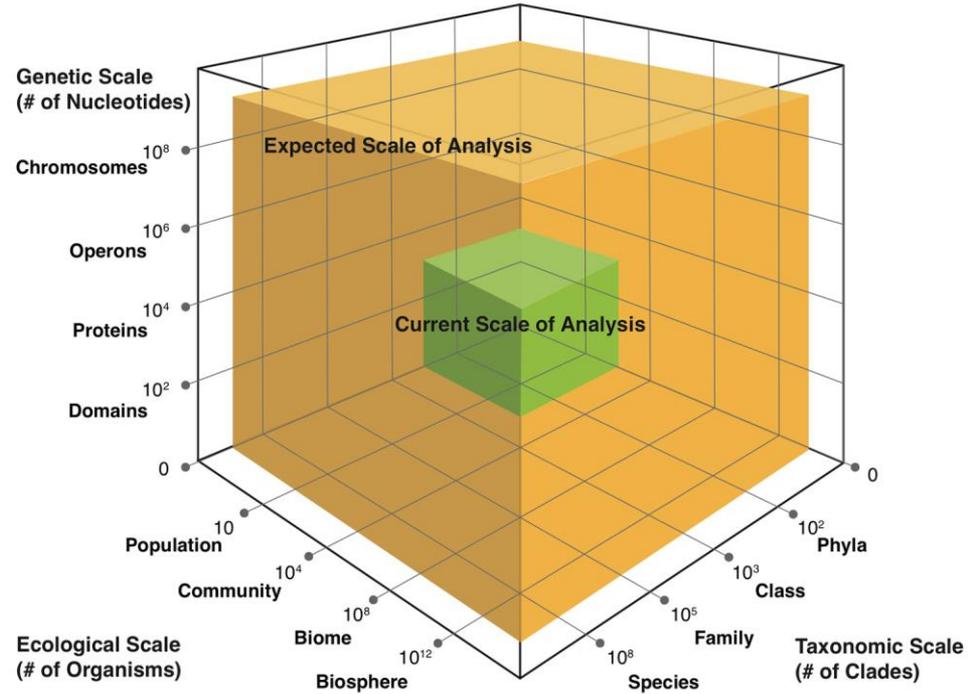
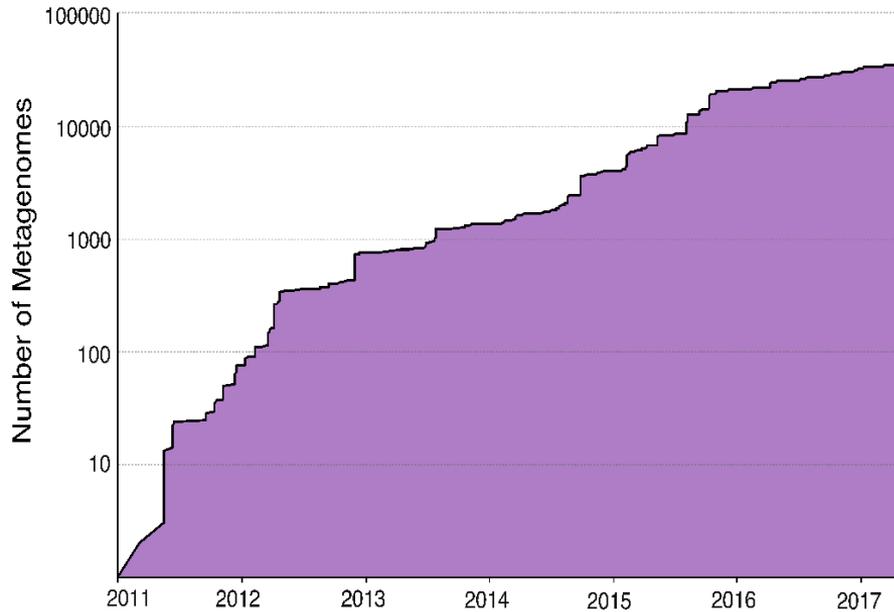


NASA
Microbiomes on
Earth and in space



USDA
Plant and livestock
microbiomes

The immense scale of data



Advances in sequencing and omics technologies have **far outpaced** infrastructure for collecting, processing, and distributing data in an effective and uniform way.

NMDC Pilot (through Sept 2021)

DOE Task Letter to Berkeley Lab

- *Enable data access, advanced analyses, and tool development
- *Promote reproducibility and enhance cross-study comparison
- *Take advantage of our HPC resources
- *Facilitate incorporation of new or updated information
- *Maintain contact with the community





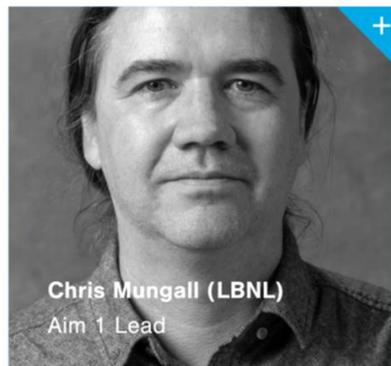
Emiley Eloe-Fadrosch (LBNL)
NMDC Lead



Nigel Mouncey (LBNL)
NMDC Deputy Lead



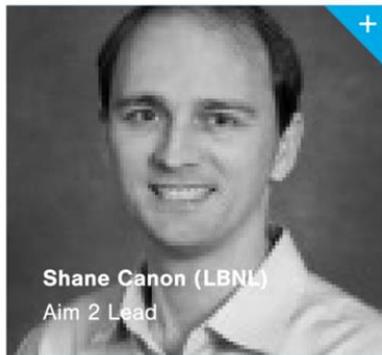
Lee Ann McCue (PNNL)
Aim 1 Lead



Chris Mungall (LBNL)
Aim 1 Lead



Patrick Chain (LANL)
Aim 2 Lead



Shane Canon (LBNL)
Aim 2 Lead



Kijersten Fagnan (LBNL)
Aim 3 Lead



Elisha Wood-Charlson (LBNL)
Aim 4 Lead



Stanton Martin (ORNL)
FAIR Strategic Team Lead

Guiding principles for data

Empower the research community to harness microbiome data exploration and discovery through a collaborative integrative data science system.



Standards

Community-driven and accepted

Continued development to address future needs



Quality

Curation and quality control to ensure data adheres to those standards



Integration

Standardized, reproducible analytical pipelines across heterogeneous data sets



Access

Discovery based on scientific inquiry

Search using existing data

NMDC Pilot - 4 Aims



Design metadata standards

Ontology mapping, data curation, automated annotation

Design and deploy workflows

Meta-genome, -transcriptome, -proteome, -metabolome

Data facilitation and integration

Registration, indexing, search, APIs

NMDC community engagement

Assess current and future needs, develop plan for sustainability

Community-driven
standards

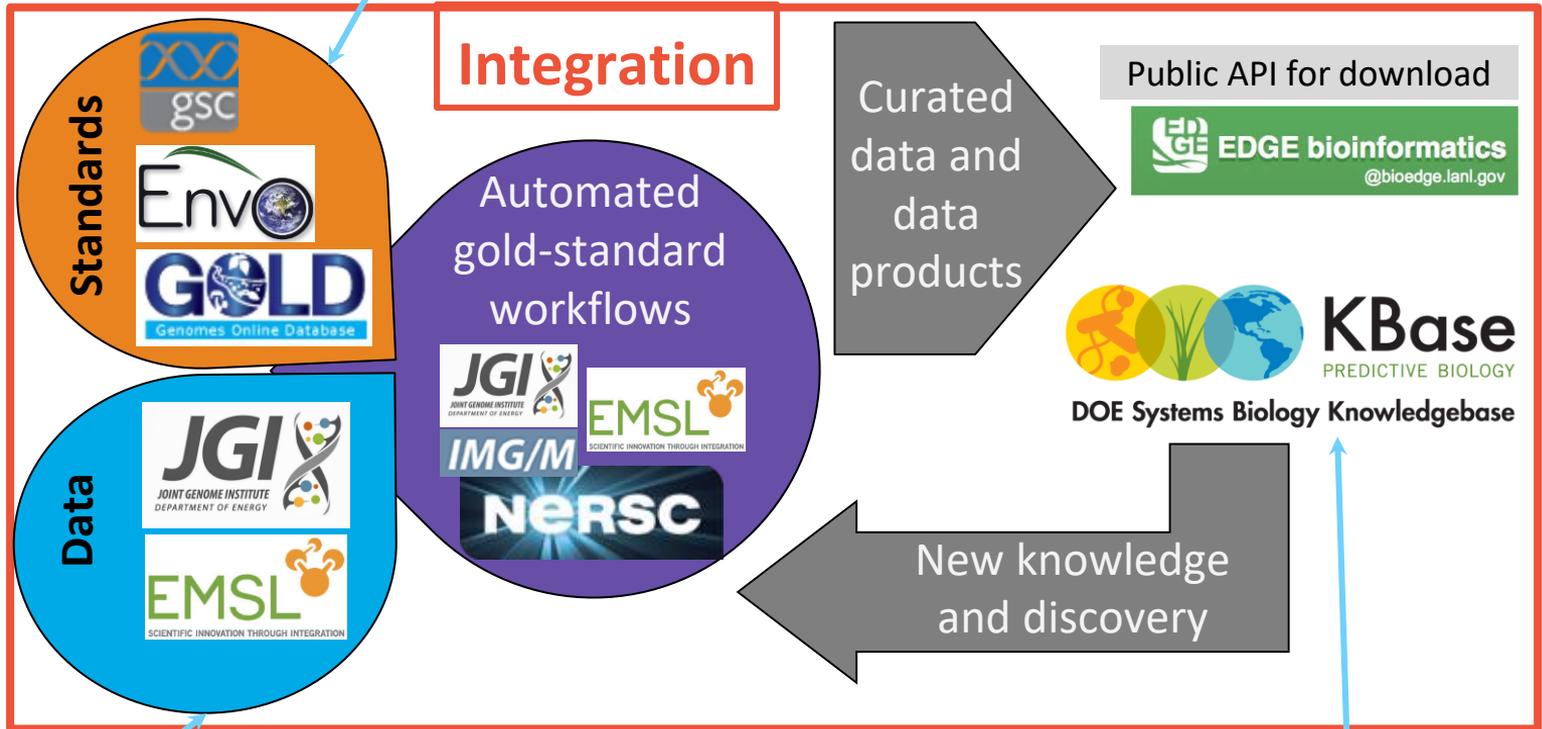
NMDC Pilot Phase I

FAIR

Findable
Accessible
Interoperable
Reusable

DATA

Curation
Linkages
QA/QC
Management



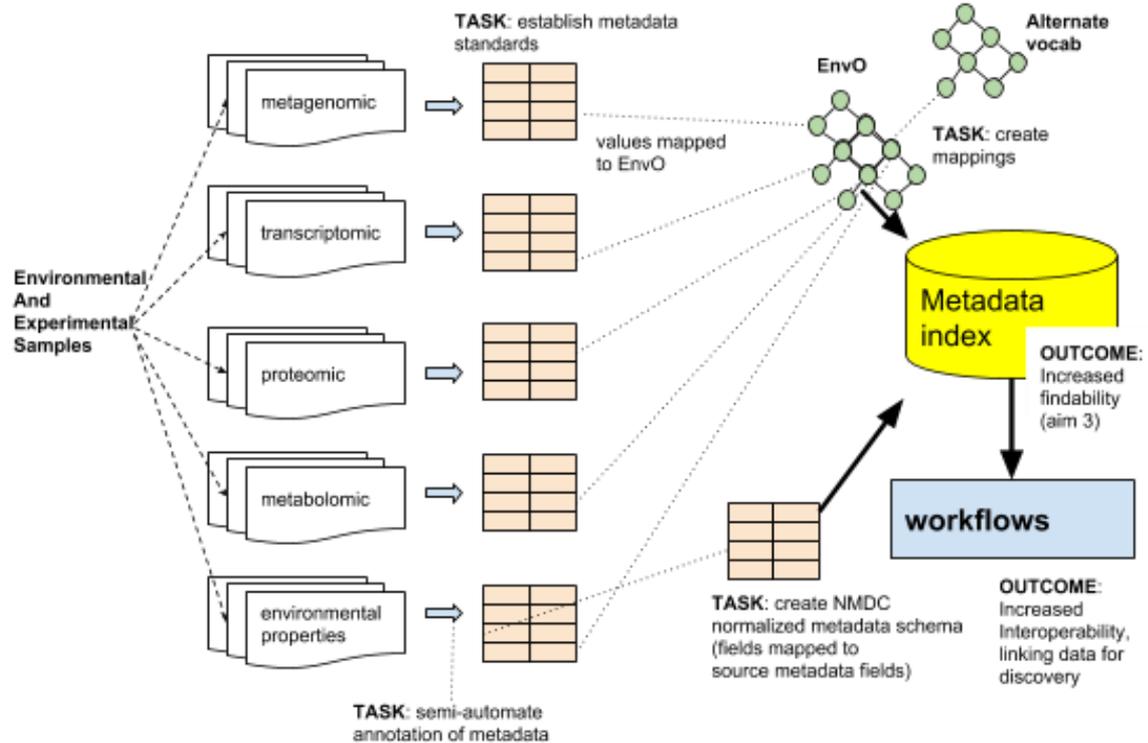
Community collected samples
(expanded to user submission in Phase II)

**Community
Engagement**

Public & private user data
Collaborative analysis
Knowledge engine
Publication

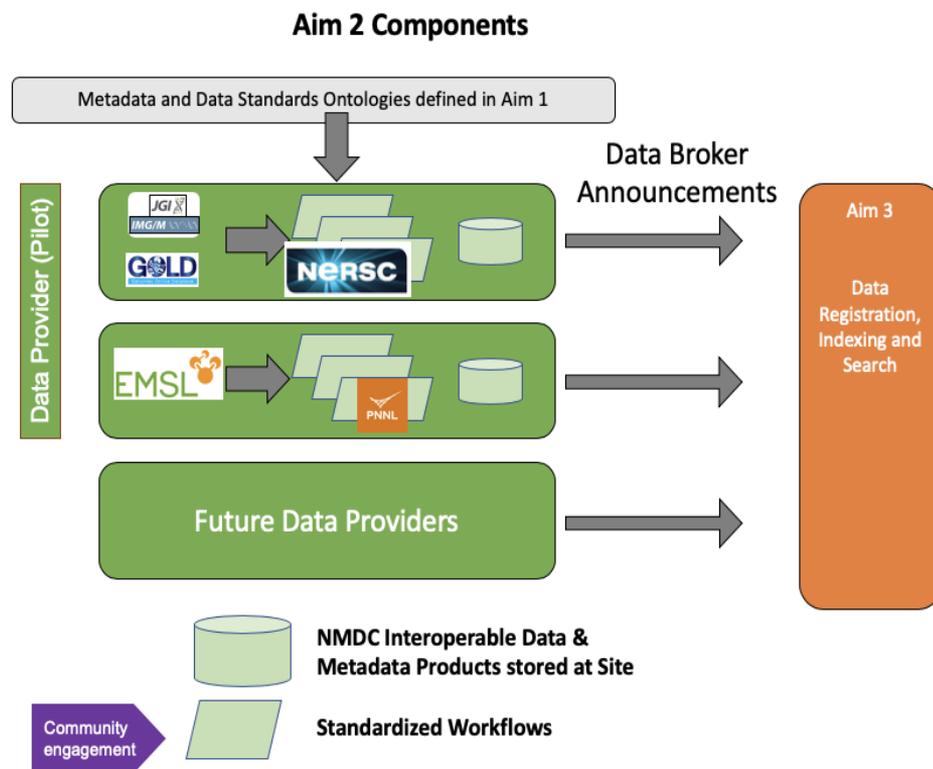
Aim 1: Design metadata standards

Existing ontology mapping
software and curation
resources -> automated
annotation



Aim 2: Design and deploy workflows

- Metagenome, metatranscriptome, metaproteome, and metabolomics
- DOE HPC systems
- Integrate execution to produce “NMDC-compliant” data products



Phase 2 (and beyond)



Expand data providers

NCBI, PATIRC, iMicrobe, HMP, DACC, Qiita, EBI, MG-RAST, ESS-DIVE

Expand data types

Environment/host measurements, hyperspectral imaging, CryoEM, structural biology

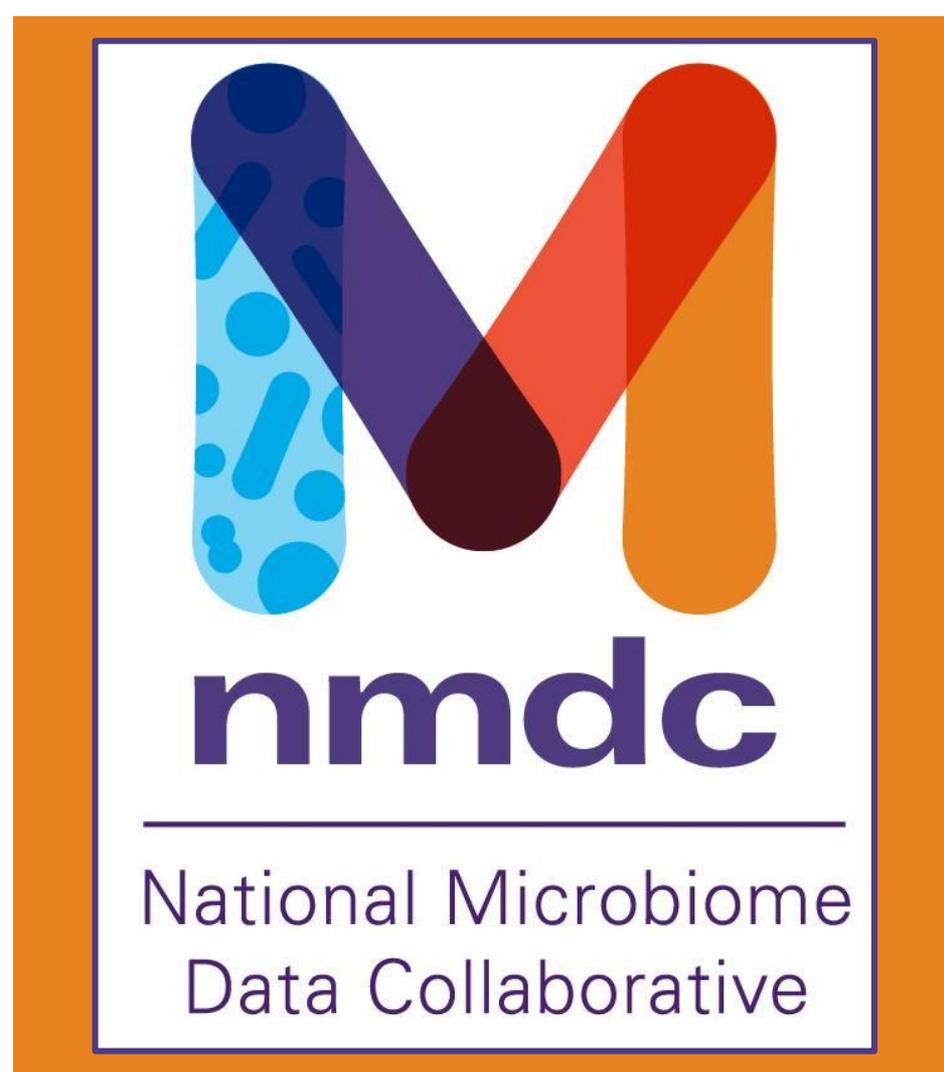
Expand computer, analysis, & vis

Exabiome !

Expand engagement

Biomedical, agricultural, built environment, atmospheric science, national defense

www.microbiomedata.org
[@microbiomedata](https://twitter.com/microbiomedata)



"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development
(NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674,
Fax: 202-459-9673, Email: nco@nitrd.gov, Website: <https://www.nitrd.gov>

