



nmdc

National Microbiome
Data Collaborative

Introduction to Urobiome Metadata Standards

Lisa Karstens, PhD

Assistant Professor

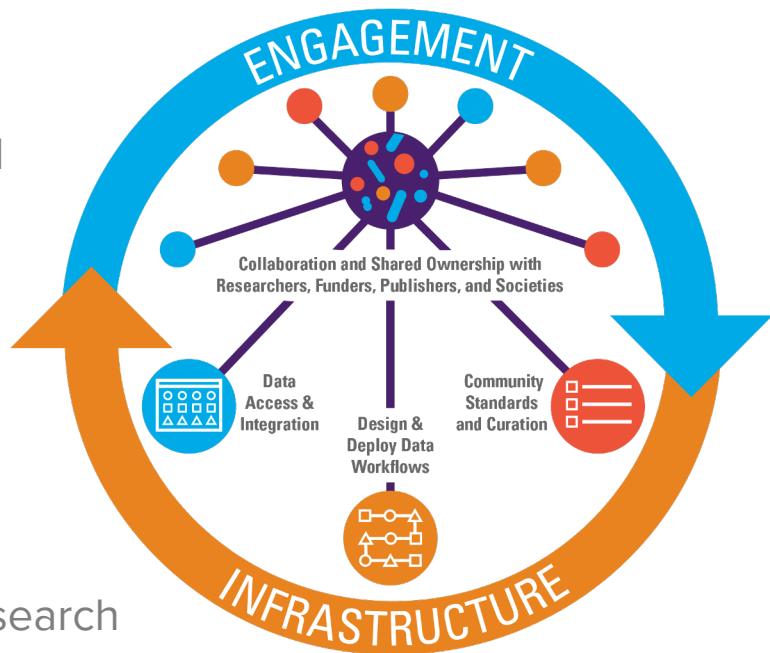
Department of Medical Informatics and Clinical Epidemiology, OHSU

NMDC Ambassador

April 28, 2022

Overview

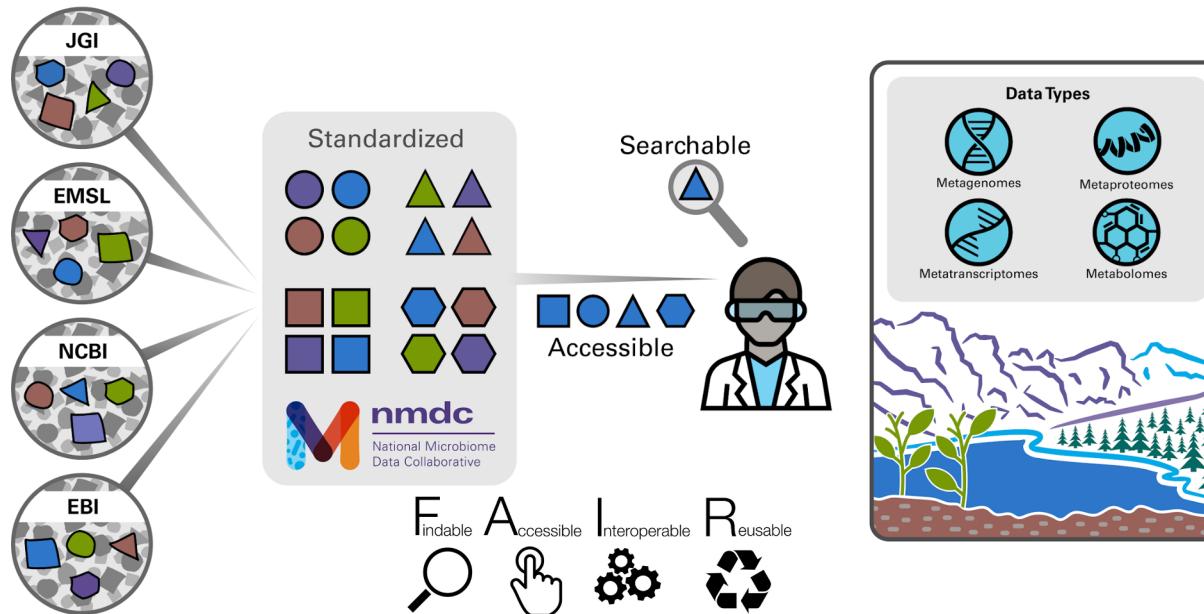
- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



The NMDC Mission



Provide a gateway to FAIR multi-omics microbiome data by leveraging best practices for data curation and processing



The NMDC data portal

[NMDC HOME](#)[METADATA STANDARDS](#)[BIOINFORMATICS WORKFLOWS](#) [ORCID LOGIN](#)

Found 638 results.

Study

PI Name

Sample

Depth

Latitude

Longitude

Geographic Location Name

GOLD Ecosystems

GOLD classification

ENVO

Environmental biome

Environmental feature

Environmental material

[OMICS](#) [ENVIRONMENT](#)

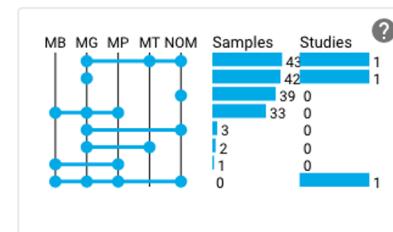
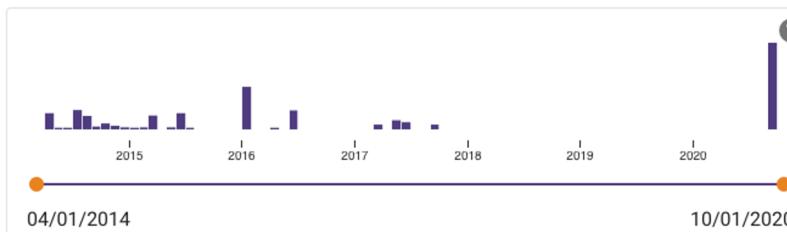
Organic Matter ?

Metagenome ?

Metatranscriptome ?

Proteomics ?

Metabolomics ?



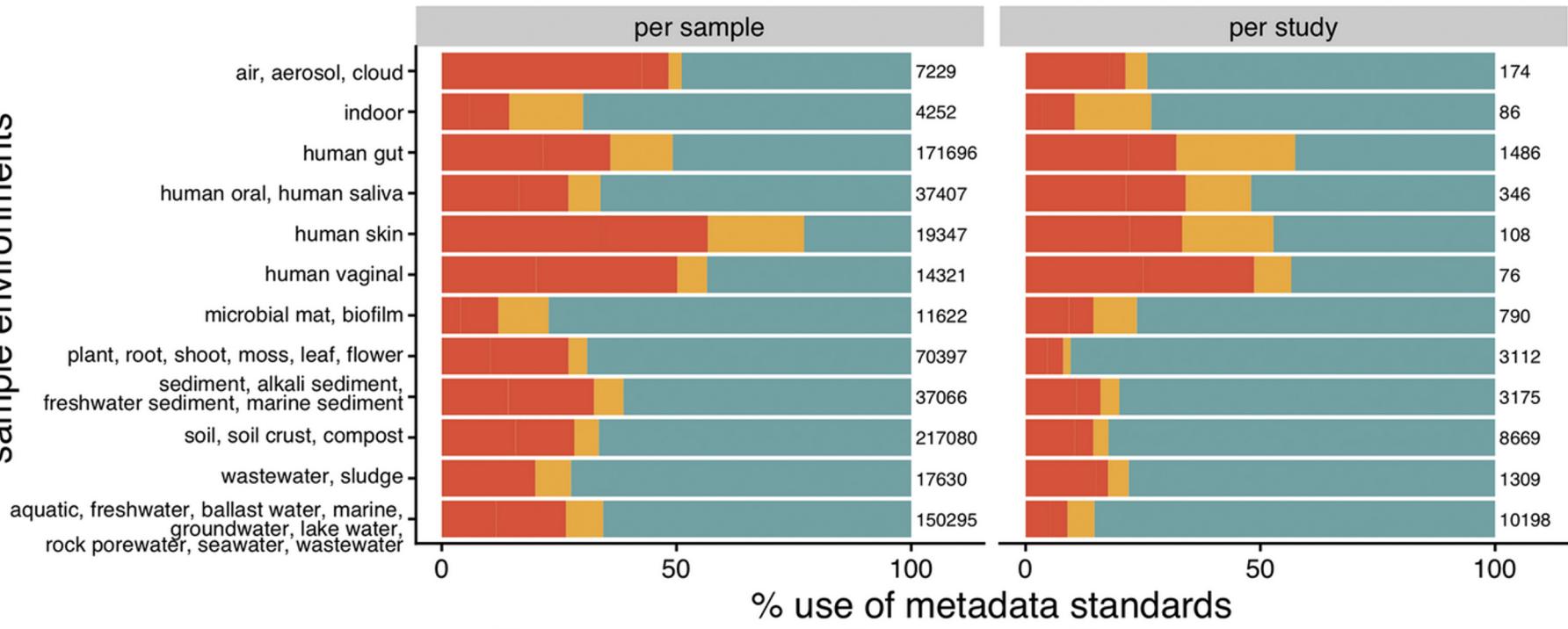
We could do better...



nmdc
National Microbiome
Data Collaborative

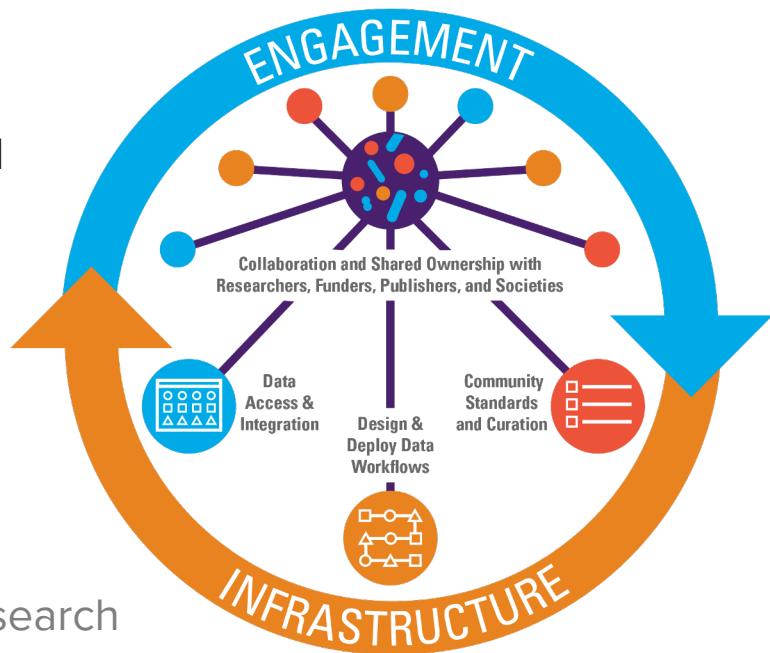
Metadata standards usage in SRA

sample environments

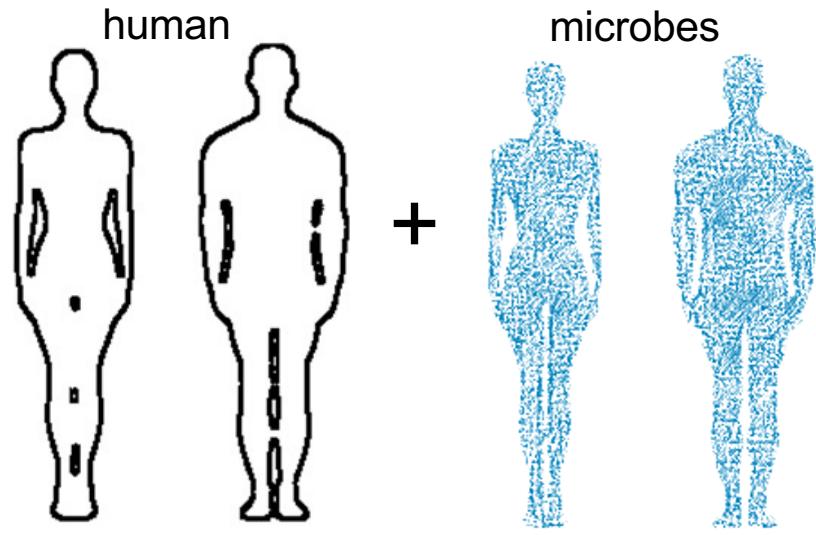


Overview

- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



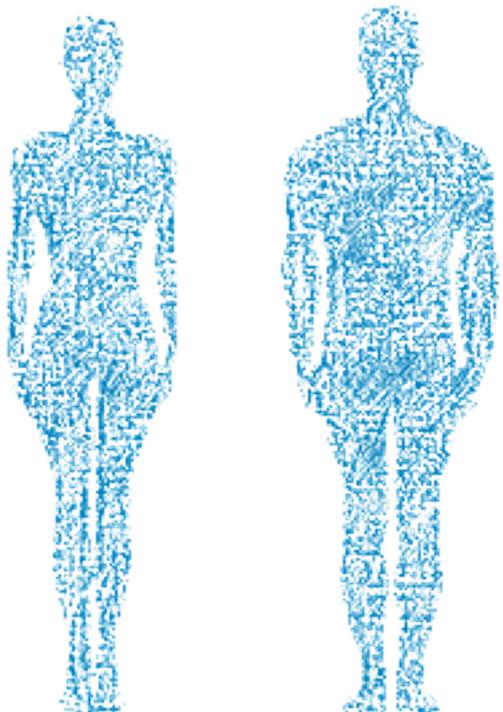
Part human, part microbes



10^{14} human cells
23,000 human genes

$> 10^{14}$ microbial cells
2 – 10 million microbial genes

→ Without understanding the interactions between microbes and humans, it is impossible to obtain a complete picture of our biology



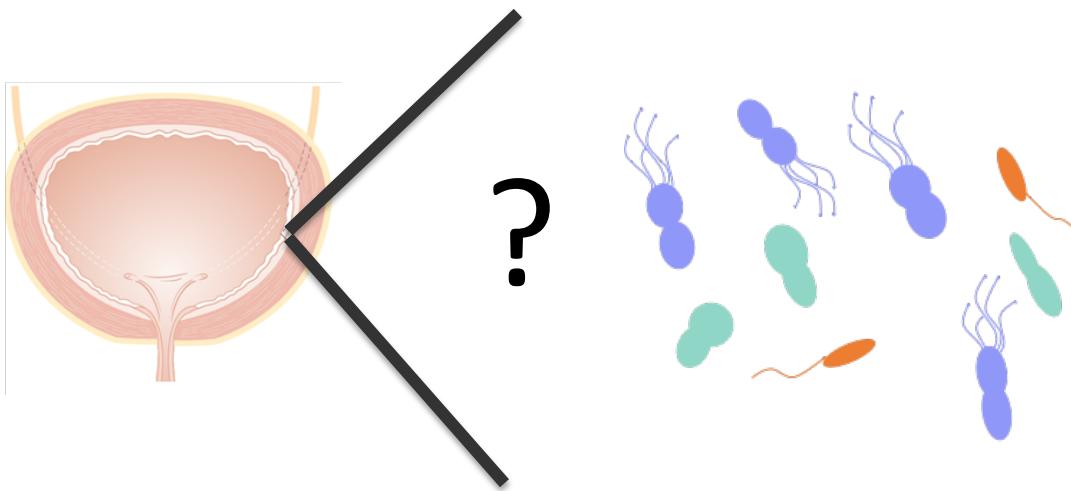
Associations with disease:

- Inflammatory bowel disease
- Obesity
- Diabetes
- Heart disease
- Depression

Association \neq causation

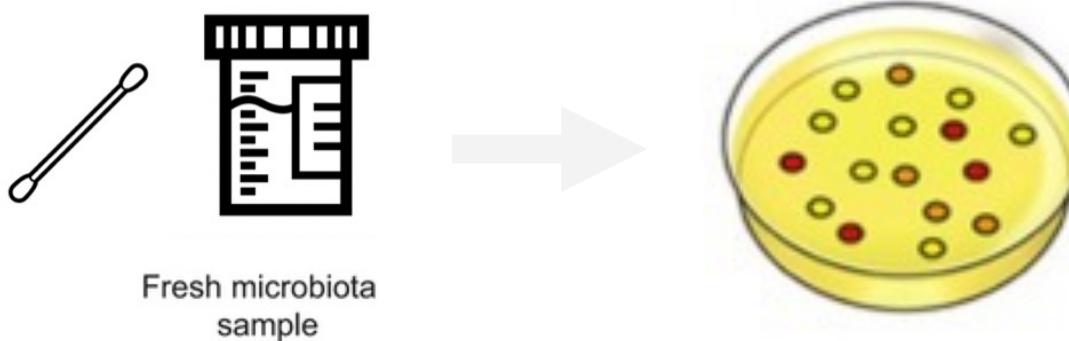
Potential for diagnostics and
prognostics

Rethinking bacteria in the bladder



Rethinking bacteria in the bladder

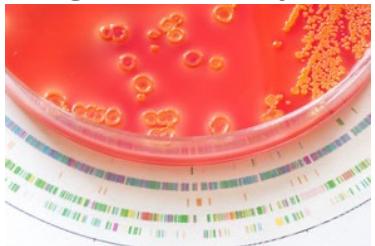
Early methods of microbial investigation relied on culturing techniques



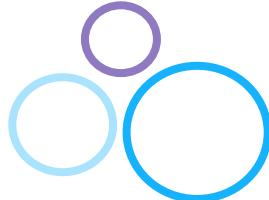
>Most microbes don't grow under **standard** culturing conditions.

Rethinking bacteria in the bladder

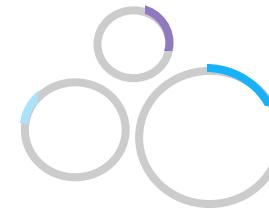
Whole genome sequencing



Shotgun metagenomic sequencing



Marker gene sequencing



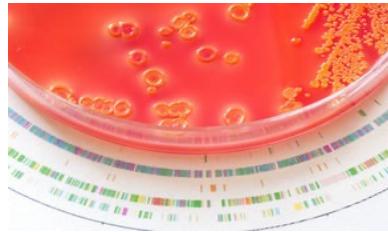
Other 'omics

metaproteomics

metabolomics

metatranscriptomics

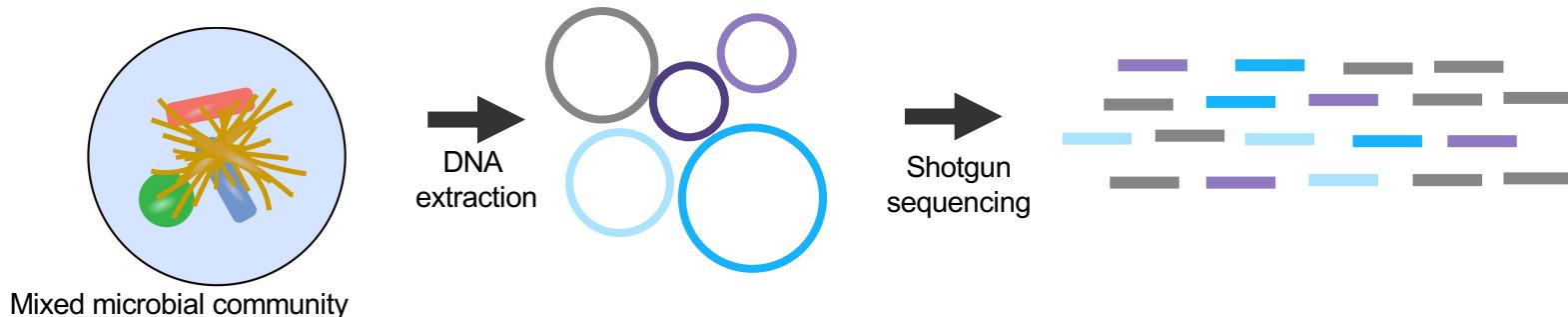
Whole genome sequencing



What it detects	What it reveals	Advantage	Disadvantage
Bacteria, fungi, bacteriophage	Taxonomy All genes	Comprehensive, function	Labor intensive, some microbes are difficult to grow

Rethinking bacteria in the bladder

Shotgun metagenomics sequencing

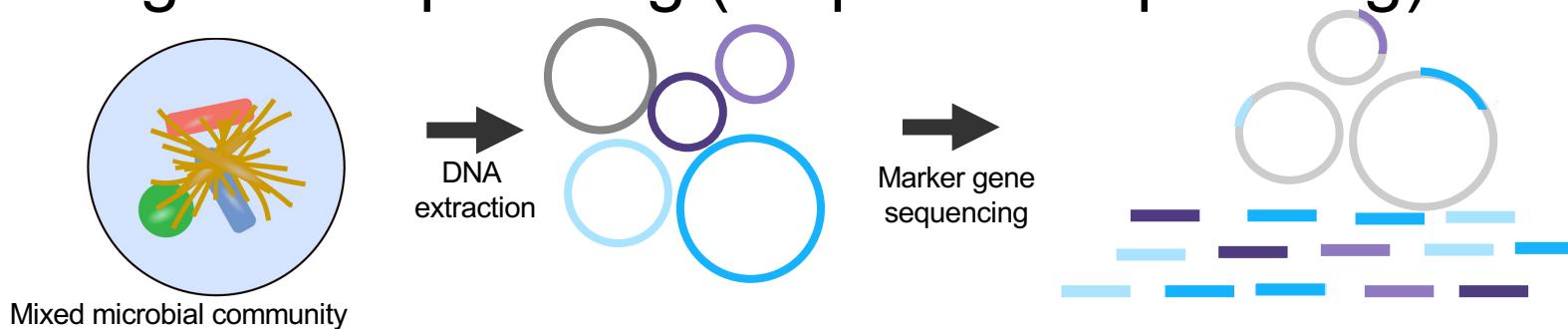


What it detects	What it reveals	Advantage	Disadvantage
Bacteria, fungi, viruses, human	Taxonomy All genes	Breadth, function	\$ \$\$, host depletion, complex bioinformatics

Not frequently used on urobiome samples (yet)

Rethinking bacteria in the bladder

Marker gene sequencing (amplicon sequencing)



What it detects	What it reveals	Advantage	Disadvantage
Bacteria only (16S rRNA gene)	Taxonomy	Low cost, increased depth	Limited view of the microbiome, compositionality
Fungi only (ITS or 18S rRNA gene)			

Rethinking bacteria in the bladder



How do we translate sequences to bacterial information?

CTCTTG TCCGAATACCTTA ATGGCACTCTTG TCCGAATACCTTA ATGGGATTACCGGA
ACTAACCTTAAT ATCCTTACTAGCGCT ACTAACCTTAAT ATCCTTACTAGCGCT TGGA
ATGGGATTACCGGA ATCCGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
ACCTTTTC ATCCTTACTAGCGCT AGCTTTC ATCCTTACTAGCGCT AGCTCGCTG
TGGTGCACACTAC TCGAAGGAATTAGC ATGGGATTACCGGA ATGGGATTACCGGA
CTAGCTTAAT ATCCTTACTAGCGCT ACTAACCTTAAT ATCCTTACTAGCGCT AGCTCGCTT
CTCTTG TCCGAATACCTTA ATGGCACTCTTG TCCGAATACCTTA ATGGGATTACCGGA
ACTAACCTTAAT ATCCTTACTAGCGCT ACTAACCTTAAT ATCCTTACTAGCGCT TGGA
ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
ACCTTTTC ATCCTTACTAGCGCT AGCTTTC ATCCTTACTAGCGCT AGCTCGCTG
TGGTGCACACTAC TCGAAGGAATTAGC ATGGGATTACCGGA ATGGGATTACCGGA
ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
CTCTTG TCCGAATACCTTA ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
ACTAACCTTAAT ATCCTTACTAGCGCT ACTAACCTTAAT ATCCTTACTAGCGCT AGCTCGCTT
ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
ACCTTTTC ATCCTTACTAGCGCT AGCTTTC ATCCTTACTAGCGCT AGCTCGCTG
TGGTGCACACTAC TCGAAGGAATTAGC ATGGGATTACCGGA ATGGGATTACCGGA
ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA ATGGGATTACCGGA
CTAGCTTAAT ATCCTTACTAGCGCT ACTAACCTTAAT ATCCTTACTAGCGCT AGCTCGCTT



	S1	S2	S...
Bacteria 1	4	0	2
Bacteria 2	43	49	24
Bacteria 3	56	65	43
...			

How do we translate data into biological insights?

	S1	S2	S...
Bacteria 1	4	0	2
Bacteria 2	43	49	24
Bacteria 3	56	65	43
...			



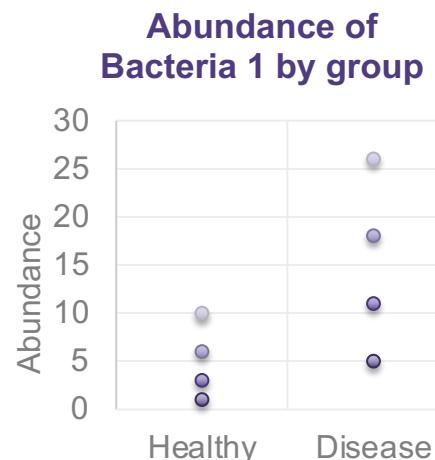
Rethinking bacteria in the bladder

Metadata contains the context we need to understand data.

	group	age	...
S1	healthy	23	...
S2	disease	72	...
S3	healthy	54	...
...			

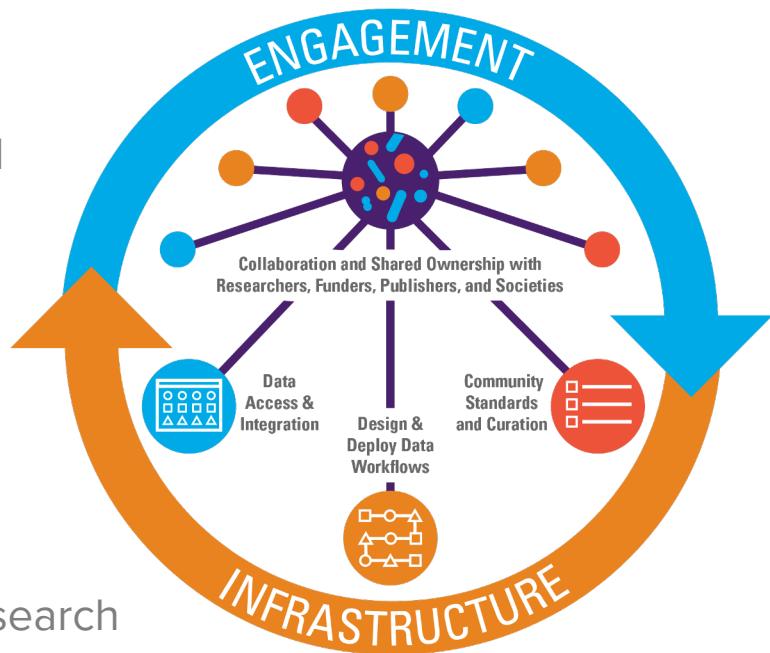


	S1	S2	S...
Bacteria 1	4	15	28
Bacteria 2	43	49	24
Bacteria 3	56	65	43
...			

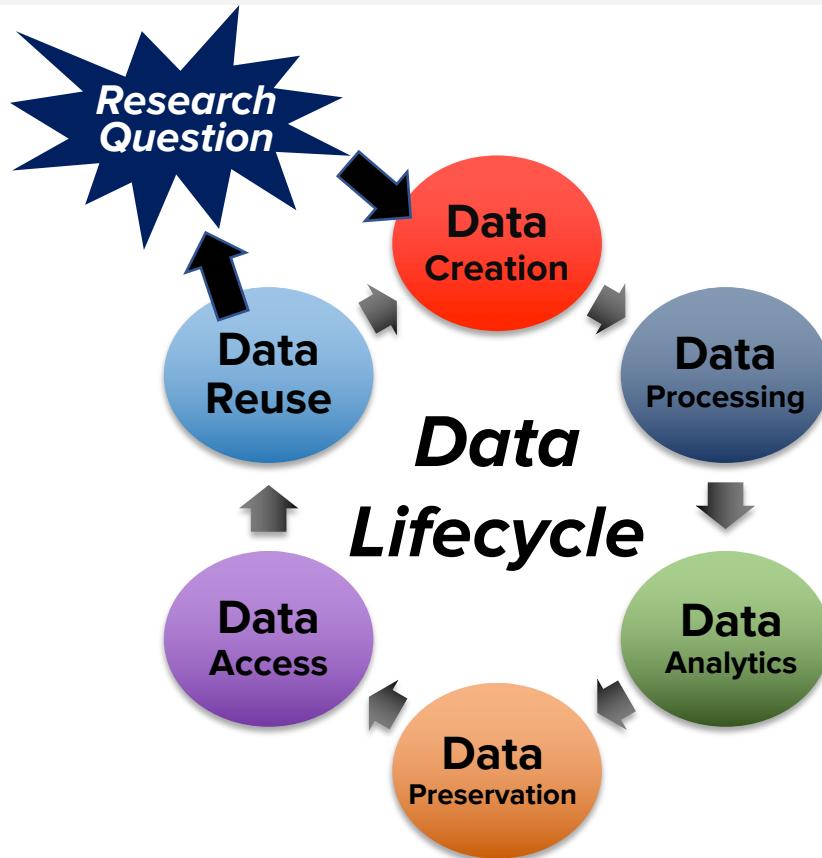


Overview

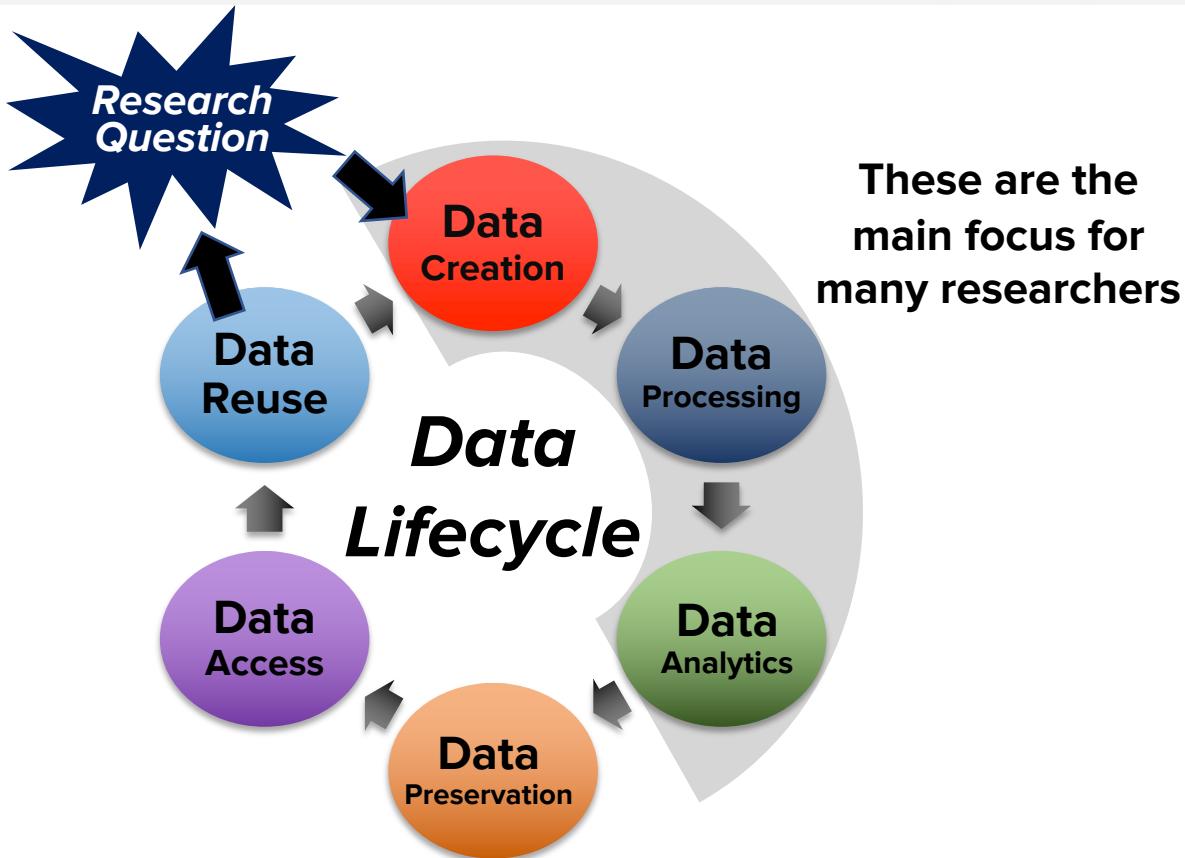
- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



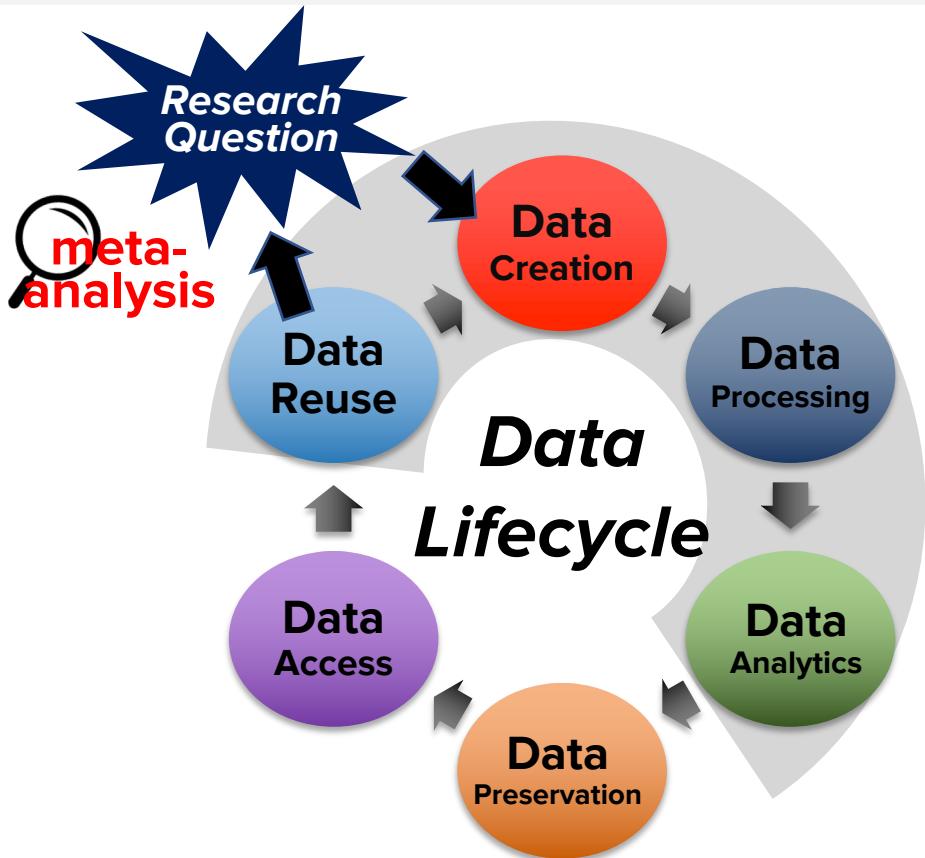
Data Lifecycle and Metadata



Data Lifecycle and Metadata



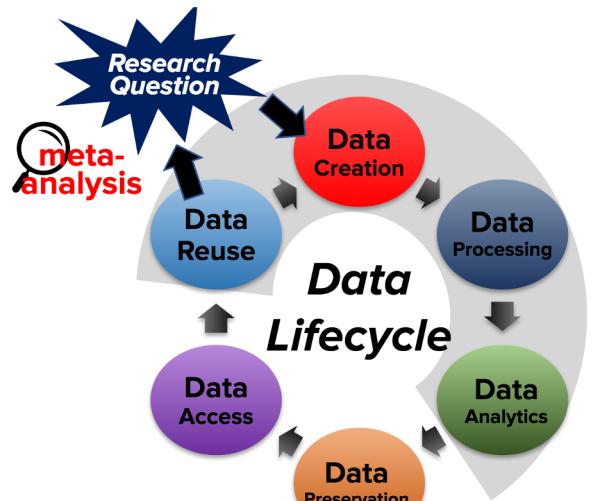
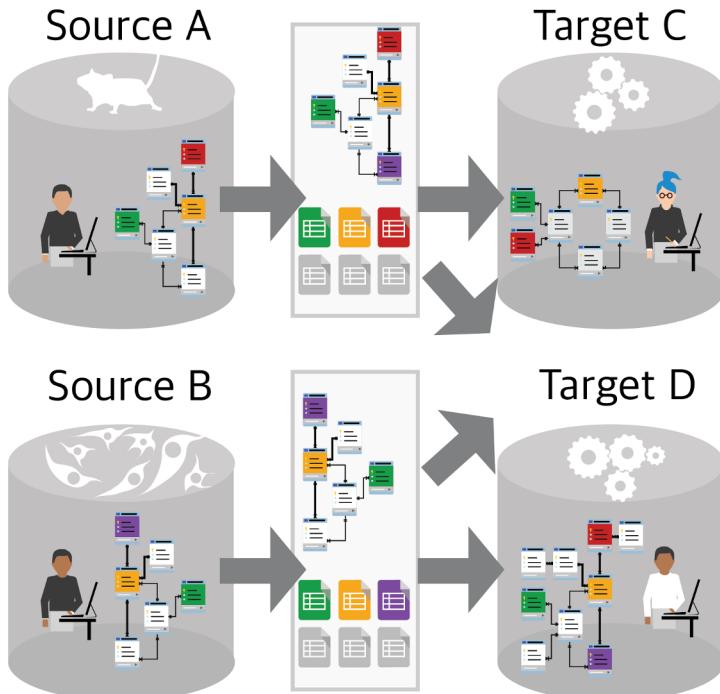
Data Lifecycle and Metadata



Data Lifecycle and Metadata



Data access and reuse - *Ideally*

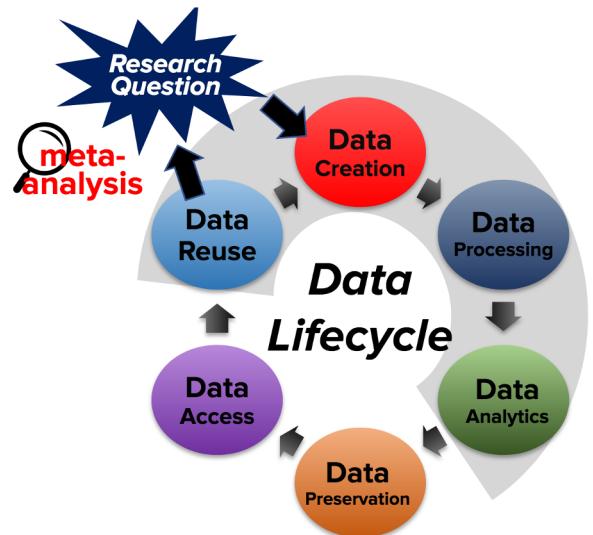


Data Lifecycle and Metadata



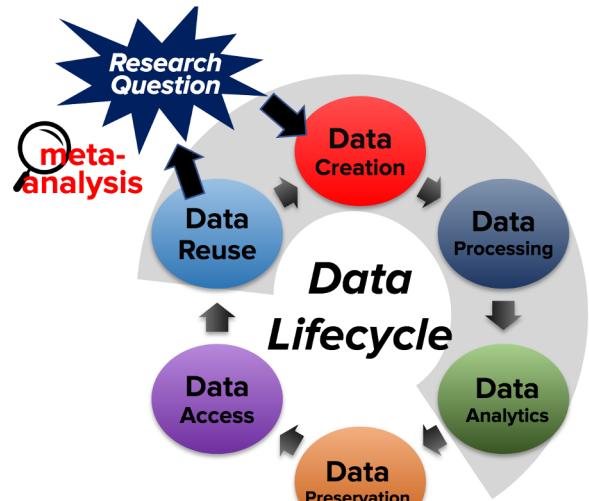
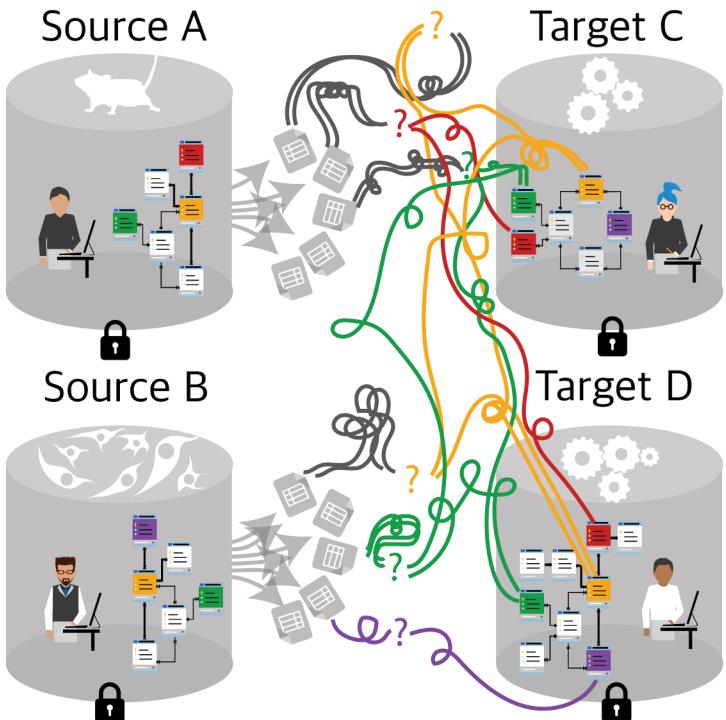
POLL! What's your data reuse experience?

Slido

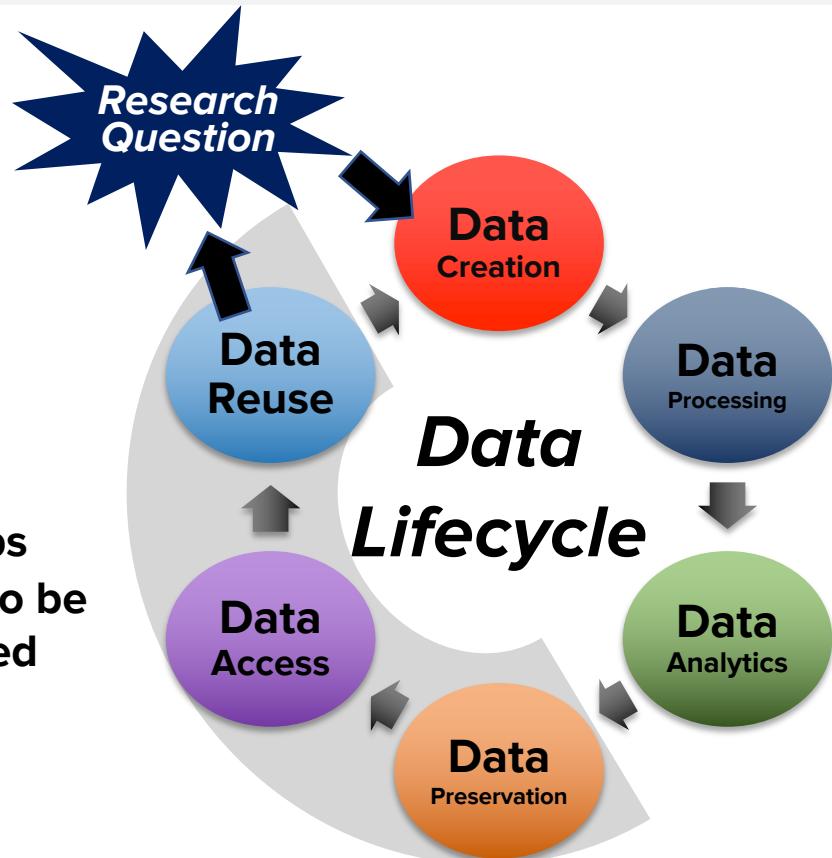


Data Lifecycle and Metadata

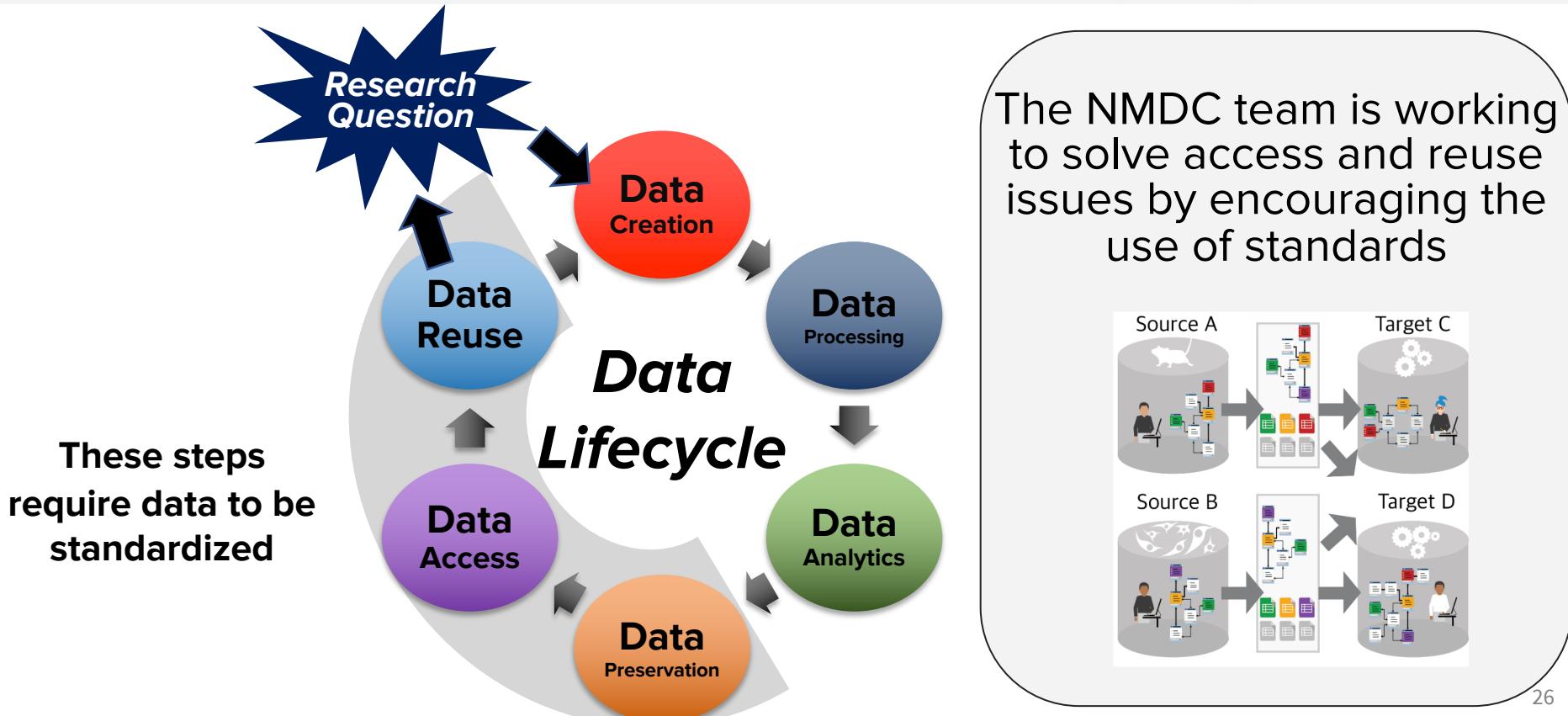
Data access and reuse - *Reality*



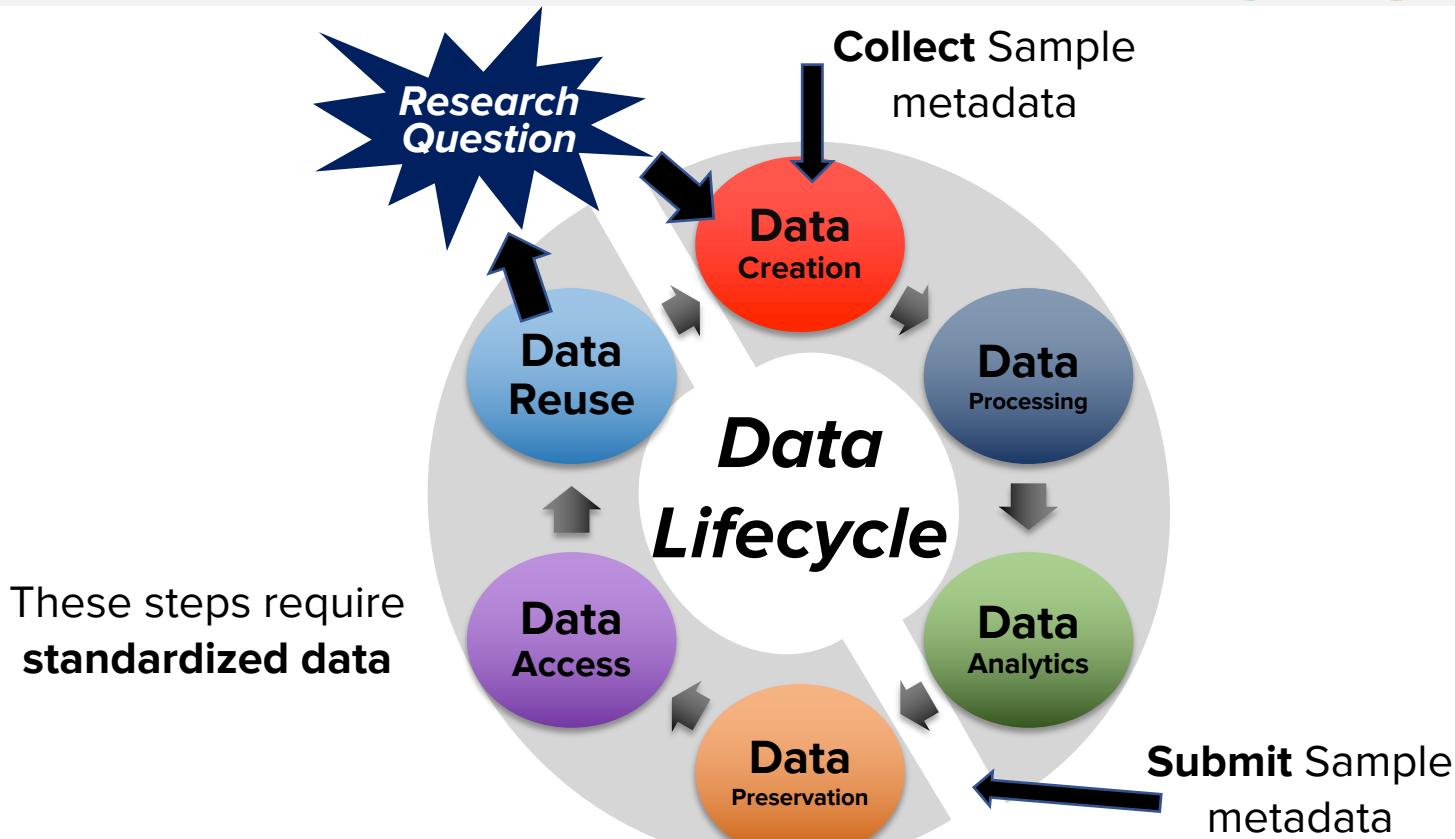
Data Lifecycle and Metadata



Data Lifecycle and Metadata



Data Lifecycle and Metadata



Best Practices in Data Management

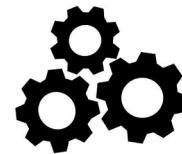


For research: the *intentional* process of collecting, storing, processing, and protecting data

Best Practices in Data Management

For data preservation: data are findable, accessible, interoperable, and reusable (FAIR) now and into the future. Good data management follows the **FAIR** data principles.

F_{indable} A_{ccessible} I_{nteroperable} R_{Reusable}



Wilkinson. et al. *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

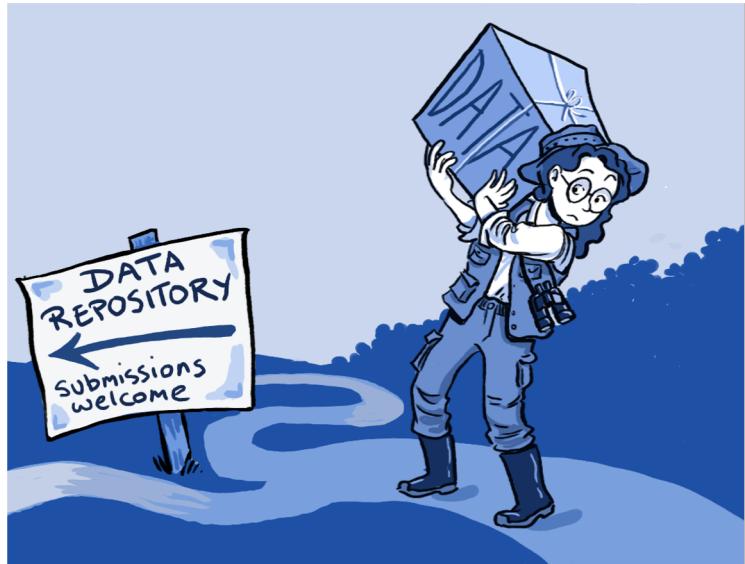
Best Practices in Data Management

Benefits:

- Data are discoverable
- Data are accessible
- Data are comparable
- Data are persistent (e.g., DOI)

Impact:

- Increased scientific exposure
- Work is appropriately credited
- Impact is measurable / quantifiable



From: Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, et al. (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol 12(1): e1001779.
<https://doi.org/10.1371/journal.pbio.1001779>

Best Practices in Data Management



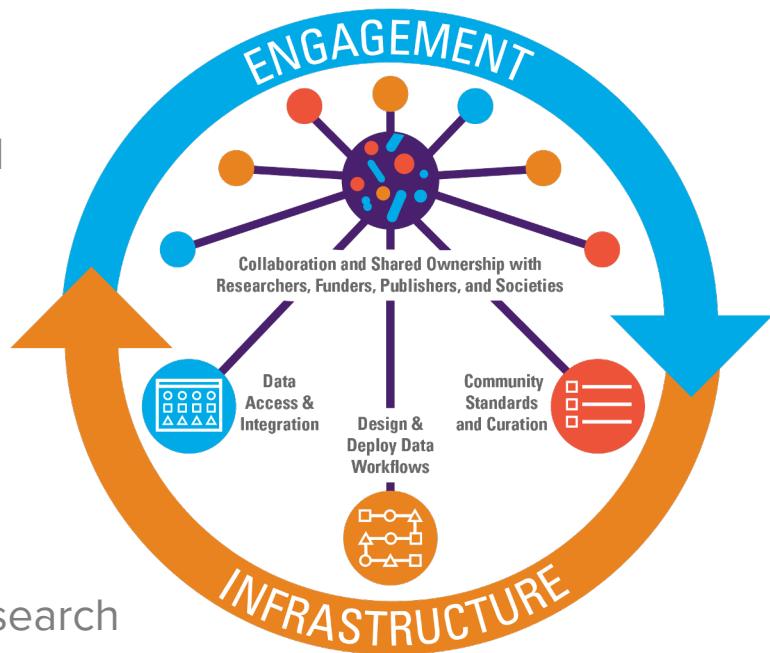
Benefits: Good data management allows for better science.

The effective reuse of data and integration of knowledge by the community after publication/release is *the key conduit leading to knowledge discovery and innovation.*

Goal: Enable great science through data sharing and reuse.

Overview

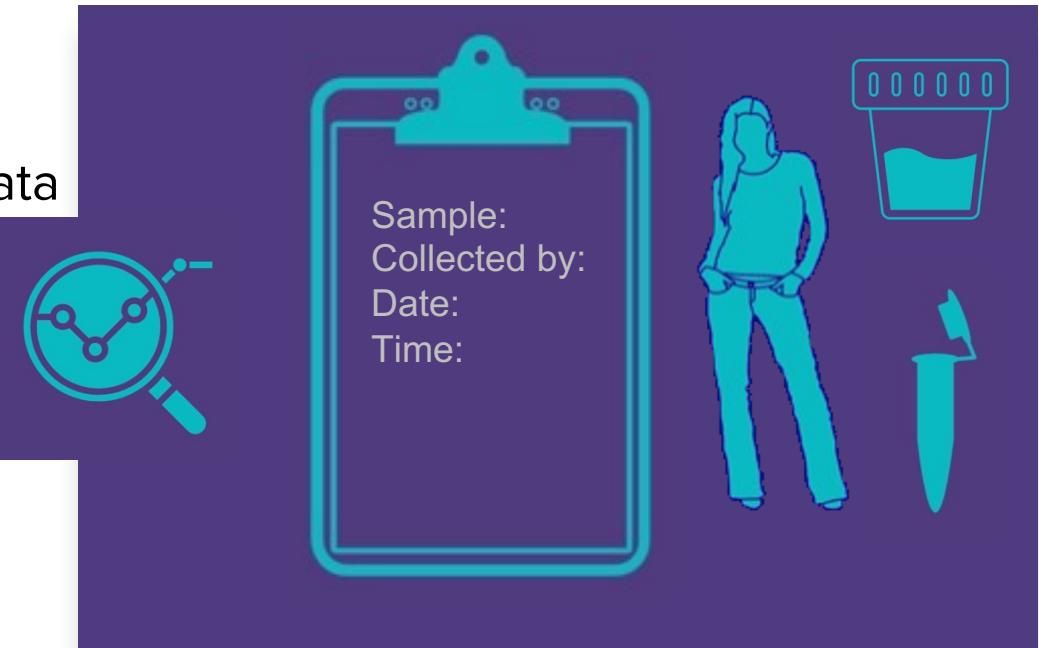
- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



What are Metadata ?

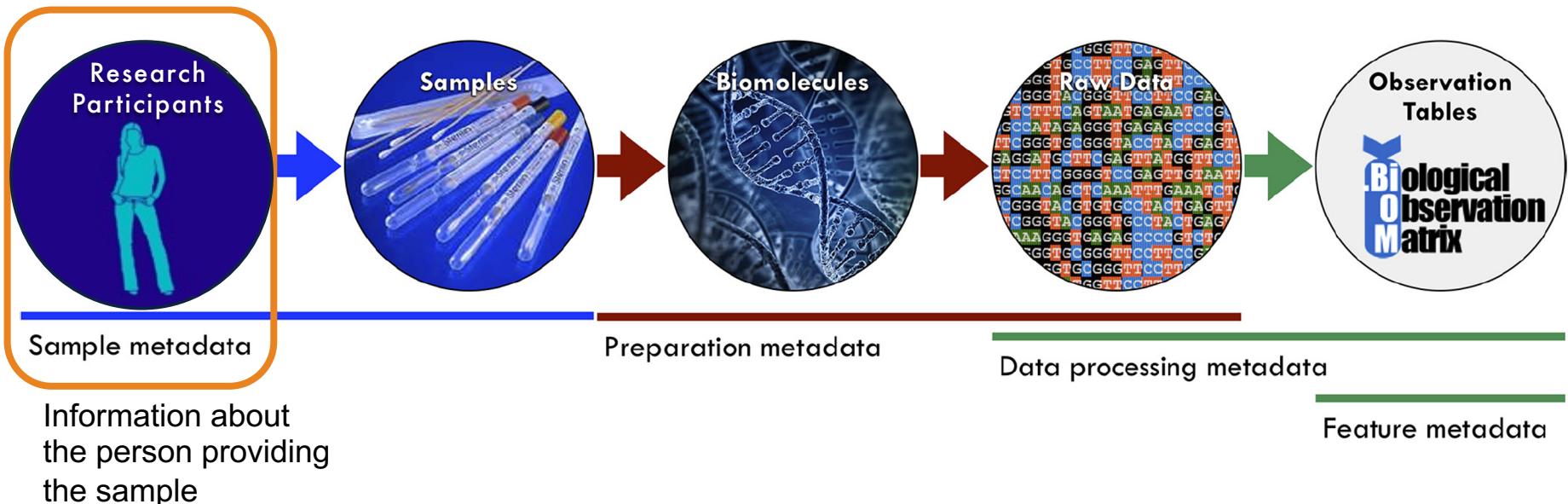
Metadata are ...

- Contextual data about your data
- Vital for data
 - Preservation
 - Discovery
 - Access
 - Reuse



What are Metadata ?

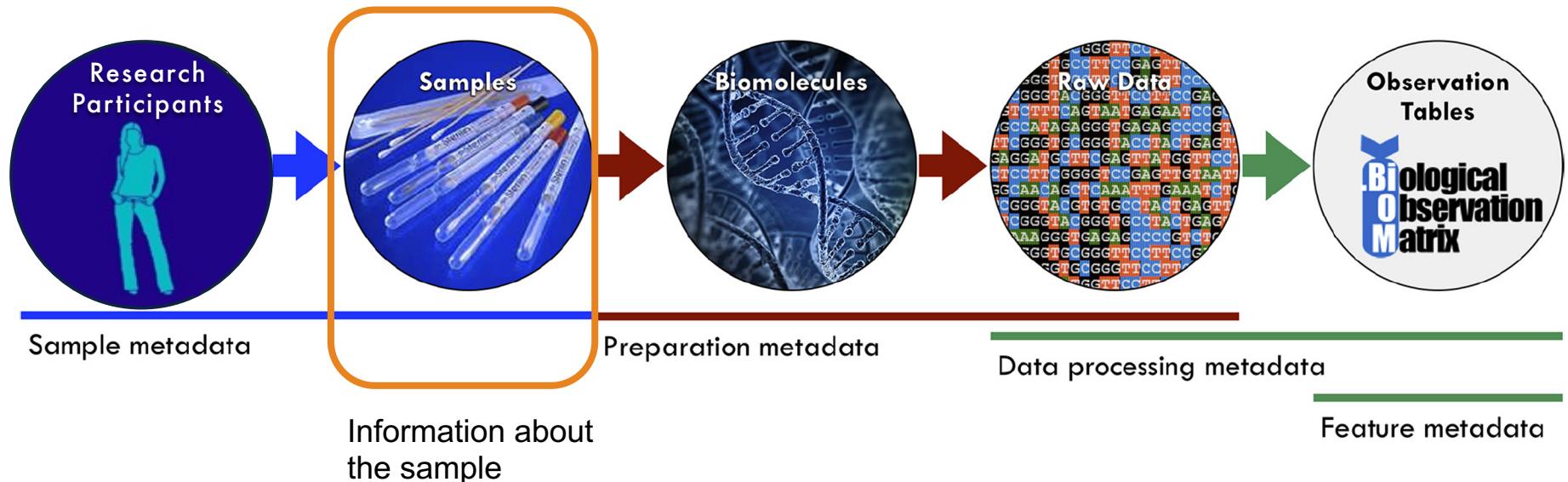
Microbiome Metadata - Sample metadata



Adapted from: **Introduction to Metadata and Ontologies: Everything You Always Wanted to Know About Metadata and Ontologies (But Were Afraid to Ask)** DOI: [10.25979/1607365](https://doi.org/10.25979/1607365)

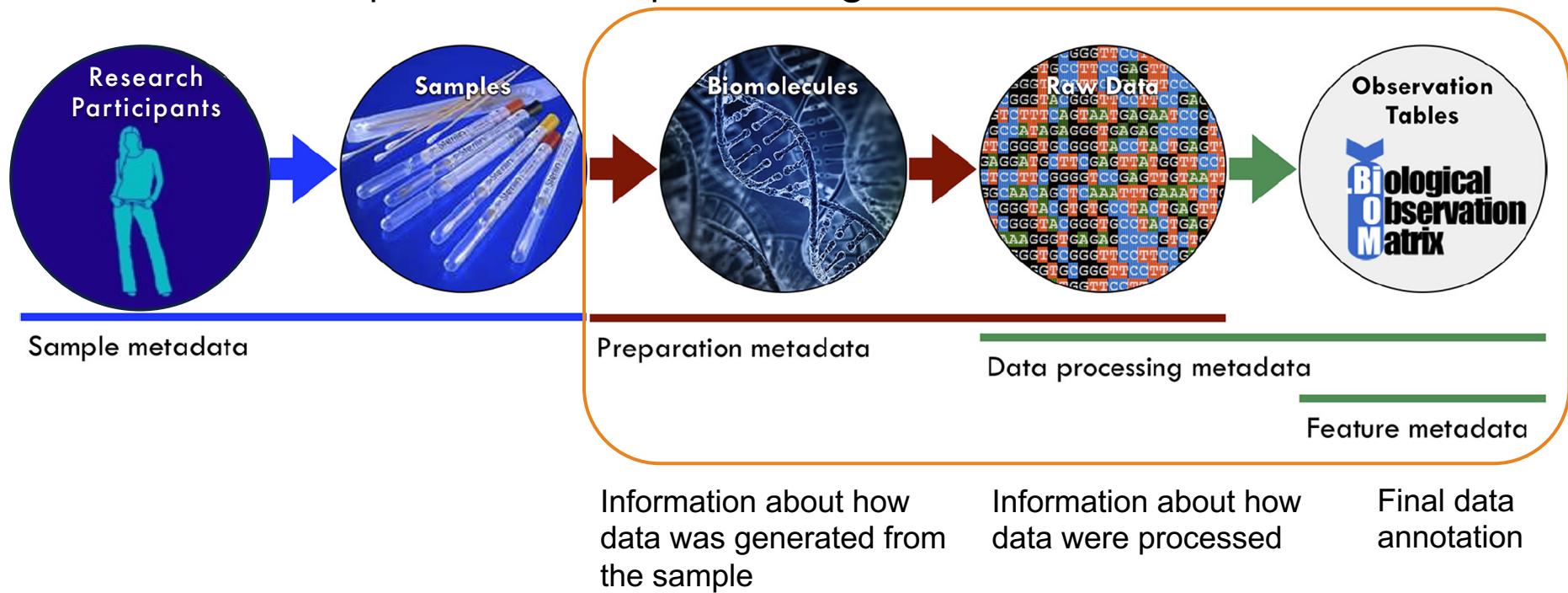
What are Metadata ?

Microbiome Metadata - Sample metadata



What are Metadata ?

Not covered – Preparation, data processing, and feature metadata

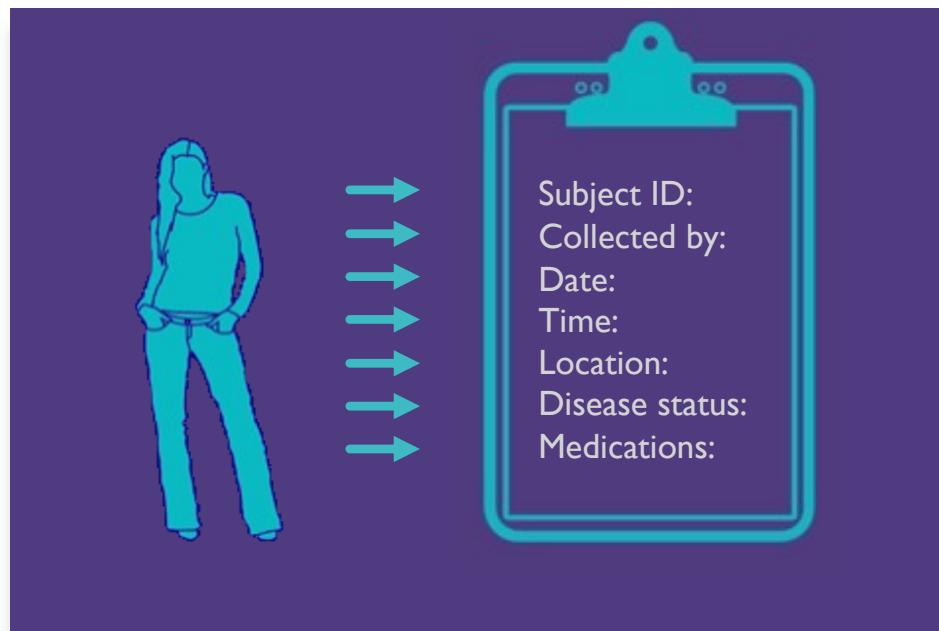
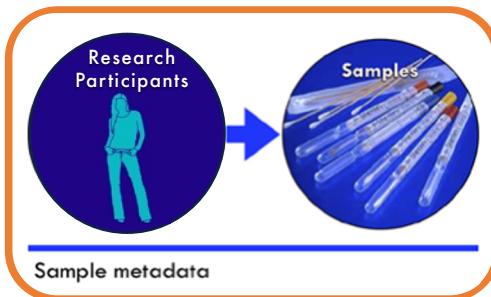


What are Metadata ?

Sample metadata includes

information about:

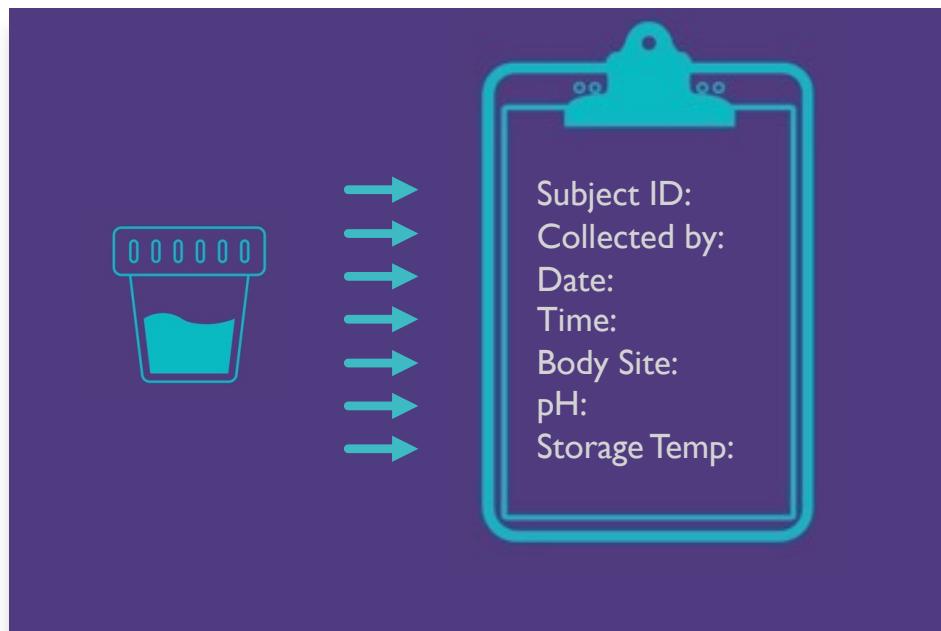
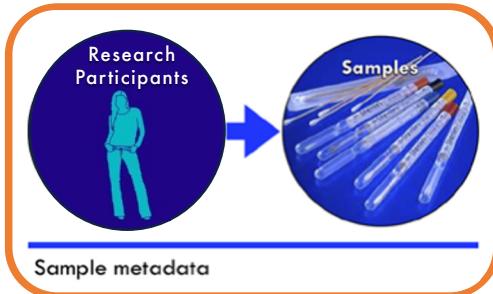
- ***Phenotypic Properties*** about the individual the sample was collected from



What are Metadata ?

Sample metadata includes information about:

- **When** it was collected
- **Where** it was collected
- **What** kind of sample is it



Why metadata standards?

Similar types of data can be difficult to reuse if they lack consistent language and formatting





Why metadata standards?

Similar types of data can be difficult to reuse if they lack consistent language and formatting



sample_id	sample_type	subject_age
c5643	clean catch	23
d5632	catheter	35
c5672	catheter	50
v8972	vaginal	50

ID number	Sample origin	age
C5643	voided urine	23 y
D5632	urine	35 y 3 mo
C5672	urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?

Standardized column headers



sample_id	sample_type	subject_age
c5643	clean catch	23
d5632	catheter	35
c5672	catheter	50
v8972	vaginal	50



sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	urine	35 y 3 mo
C5672	urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?



Similar types of data can be difficult to reuse if they lack consistent language and formatting



sample_id	sample_type	subject_age
c5643	clean catch	23
d5632	catheter	35
c5672	catheter	50
v8972	vaginal	50

sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	urine	35 y 3 mo
C5672	urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?

Consistent capitalization



sample_id	sample_type	subject_age
C5643	clean catch	23
D5632	catheter	35
C5672	catheter	50
V8972	vaginal	50



sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	urine	35 y 3 mo
C5672	urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?

Similar types of data can be difficult to reuse if they lack consistent language and formatting



sample_id	sample_type	subject_age
C5643	clean catch	23
D5632	catheter	35
C5672	catheter	50
V8972	vaginal	50

sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	urine	35 y 3 mo
C5672	urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?

Consistent terminology



sample_id	sample_type	subject_age
C5643	voided urine	23
D5632	catheter urine	35
C5672	catheter urine	50
V8972	vaginal swab	50

sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	catheter urine	35 y 3 mo
C5672	catheter urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?



Similar types of data can be difficult to reuse if they lack consistent language and formatting



sample_id	sample_type	subject_age
C5643	voided urine	23
D5632	catheter urine	35
C5672	catheter urine	50
V8972	vaginal swab	50

sample_id	sample_type	subject_age
C5643	voided urine	23 y
D5632	catheter urine	35 y 3 mo
C5672	catheter urine	50 y
V8972	vaginal swab	50 y 1 mo

Why metadata standards?

Consistent formatting



sample_id	sample_type	subject_age
C5643	voided urine	23 years
D5632	catheter urine	35 years
C5672	catheter urine	50 years
V8972	vaginal swab	50 years



sample_id	sample_type	subject_age
C5643	voided urine	23 years
D5632	catheter urine	35 years
C5672	catheter urine	50 years
V8972	vaginal swab	50 years

Why metadata standards?

Similar types of data can be **easy** to reuse if they have consistent language and formatting



sample_id	sample_type	subject_age
C5643	voided urine	23 years
D5632	catheter urine	35 years
C5672	catheter urine	50 years
V8972	vaginal swab	50 years

sample_id	sample_type	subject_age
C5643	voided urine	23 years
D5632	catheter urine	35 years
C5672	catheter urine	50 years
V8972	vaginal	50 years

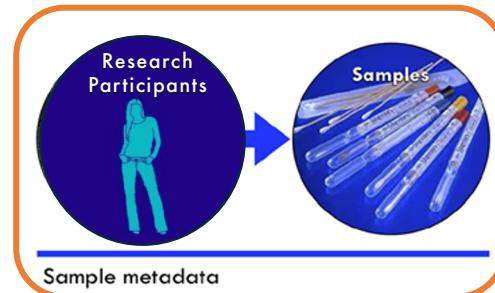
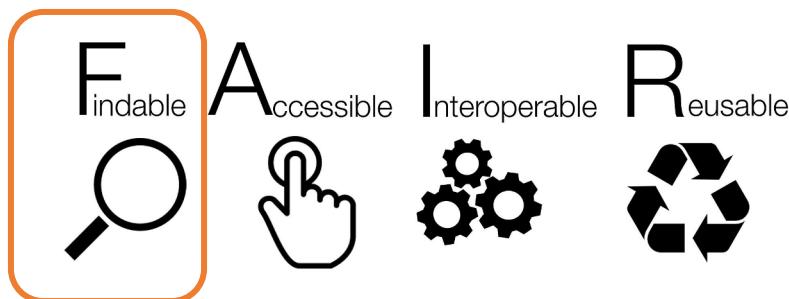
What are Metadata ?

Standardized sample metadata

support **search and discovery**

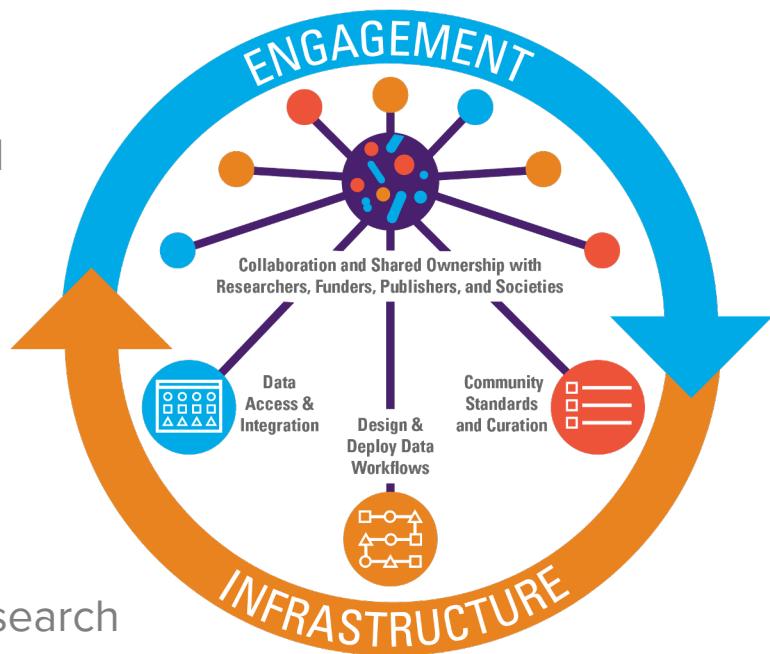
based on:

- **When** it was collected
- **Where** it was collected
- **What** kind of sample is it
- **Phenotypic Properties** about the individual the sample was collected from



Overview

- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



Community Metadata Standards



**The NMDC metadata standards utilize and enhance
existing community-driven standards**



1. MIxS: Minimum Information about any (x) Sequence
Genomic Standards Consortium (GSC)



1. [GOLD: Genomes OnLine Database](#)
Joint Genome Institute (JGI)



3. [EnvO: Environment Ontology](#)
Open Biological and Biomedical Ontology (OBO) Foundry

The NMDC + Community Standards



NMDC – derive mandatory metadata from community standards

MixS

- Sample Identifiers : *sample_name*
- Geographic Location (latitude longitude) : *lat_lon*
- Geographic Region (country and/or sea, region) : *geo_loc_name*
- Collection Date : *collection_date*
- Storage Conditions : *store_cond*
- Storage Temperature : *samp_store_temp*
- Broad-scale Environmental Context : *env_broad_scale*
- Local Environmental Context : *env_local_scale*
- Environmental Medium : *env_medium*

EnvO



Genomic Standards Consortium

- MIxS: Environmental Packages + Checklists



MIxS About Introduction ▾

Human-associated
Combinations
Slots
Usages
Identifier and Mapping
Information

Human-associated

human-associated

URI: [mixs.vocab:Human-associated](#)

Combinations

- Human-associated
 - [Human-associatedMIMS](#) [MIMS]
 - [Human-associatedMIMARKSSpecimen](#) [MIMARKS specimen]

<https://genomicsstandardsconsortium.github.io/mixs/Human-associated/>

MIxS Checklists



GSC supports 5 data types as
“Minimum information about [checklist] sequence”



MIxS Checklists	
	Genome (MIGS)
	Metagenome (MIMS)
	MARKer gene (MIMARKS)
	Single Amplified Genome (MISAG)
	Metagenome Amplified Genome (MIMAG)



The NMDC supports
Metagenome
checklists (MIMS)

MIxS Environmental Packages



17 packages (*currently available in NMDC)

New packages added with community input



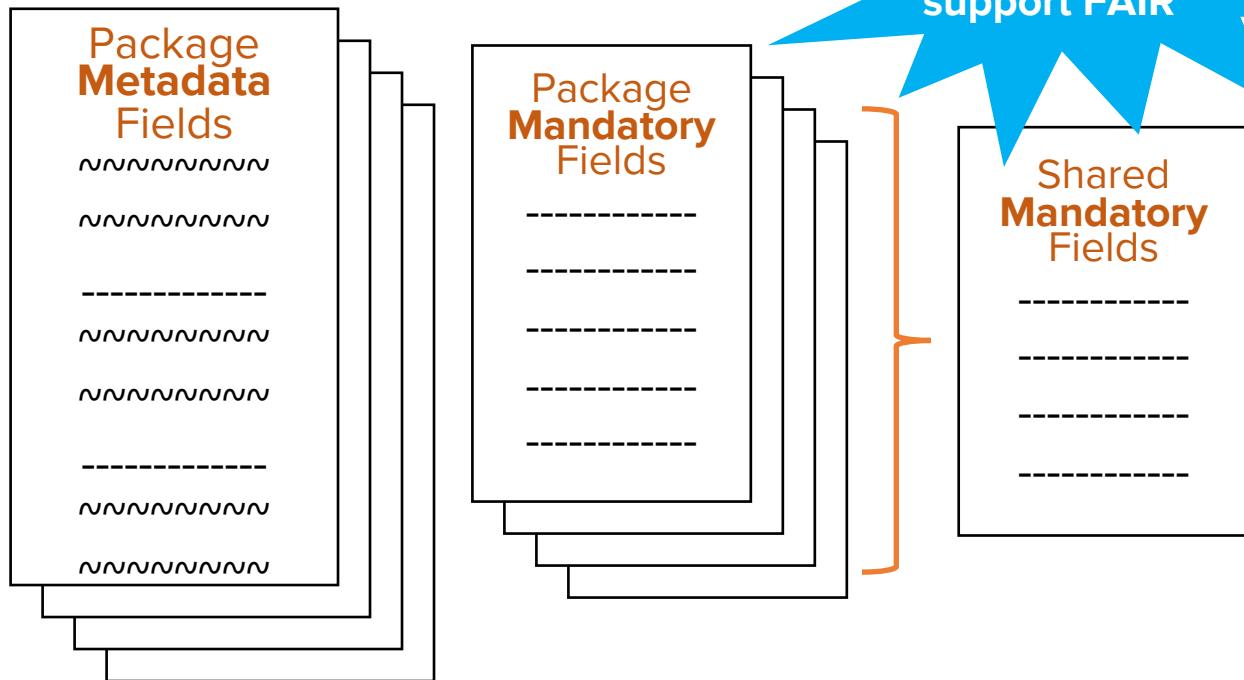
MIxS Environmental Packages	
air	built environment
host-associated	human-associated
human-gut	human-oral
human-skin	human-vaginal
hydrocarbon resources-cores	hydrocarbon resources-fluids/swabs
microbial mat/biofilm	miscellaneous natural or artificial environment
*plant associated	*sediment
*soil	wastewater/sludge
*water	<i>Pending: human-uropiome.</i>

MIxS Mandatory Fields

Mandatory metadata fields (columns) are shared
across all MIxS environmental packages and checklists



e.g., *human-associated*, *human-gut*, *human-vaginal*, ...



MIxS

EnvO

OBO EnvO (Environment Ontology)

- Mandated by MIxS for environment fields
- Dynamic, community resource
- Hierarchical classification of samples by environment
- **Community:**
 - Biosciences, Earth, environment



+ [biome](#)
+ [terrestrial biome](#)
+ [anthropogenic terrestrial biome](#)
+ [dense settlement biome](#)
- [urban biome](#)

MIxS Requires Environment Terms



MIxS descriptors specify the ***sample environment*** with
Environment Ontology (EnvO)



broad-scale environmental context

- + biome
 - + terrestrial biome
 - + anthropogenic terrestrial biome
 - + dense settlement biome
 - urban biome

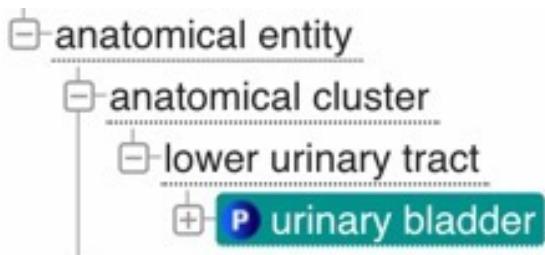
MIxS Requires Environment Terms



MIxS descriptors specify the ***sample environment*** with
Uber-anatomy ontology (UBERON)



local-scale environmental context



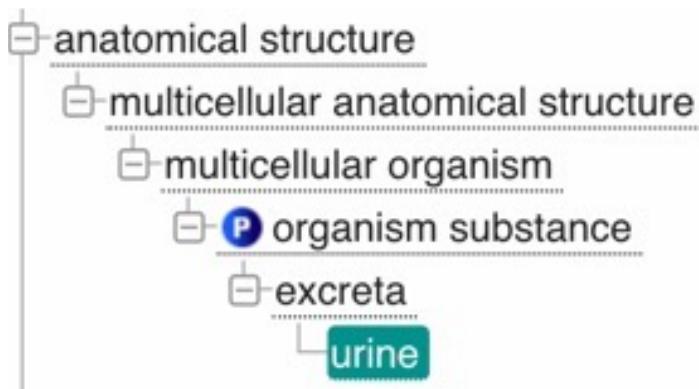
MIxS Requires Environment Terms



MIxS descriptors specify the ***sample environment*** with
Uber-anatomy ontology (UBERON)

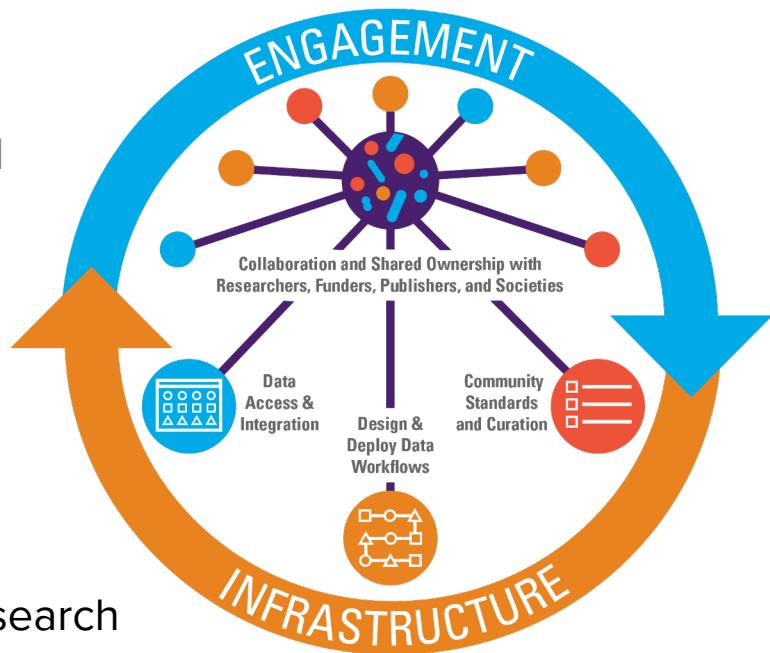


environmental medium



Overview

- NMDC Mission
- Brief introduction to the human microbiome and urinary microbiome (urobiome)
- Data Lifecycle and Metadata
- What are Metadata
- Microbiome Metadata Community Standards
- Adapting Metadata Standards for Urobiome Research



Urobiome Consensus



- Urobiome research is in its infancy
→ Ideal time to form standards!
- Effective collaboration
→ more rapid, impactful science
- Ultimate goal
→ improve healthcare for affected patients



Forming Consensus To Advance Urobiome Research

Brubaker et al. *mSystems*. 2021 Aug 31;6(4):e0137120

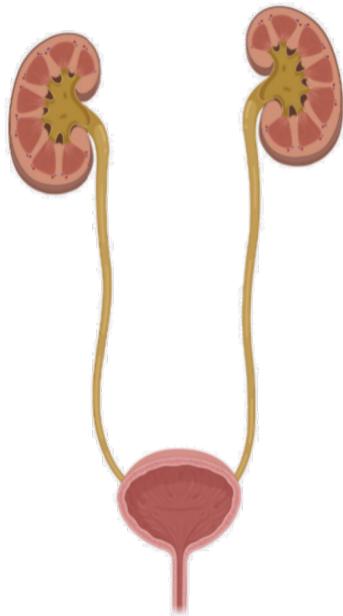
Human Subjects Research

- De-identified data only!
- Balance between being comprehensive and protecting privacy
 - Dates
 - Location
 - Age over 90

Several others!



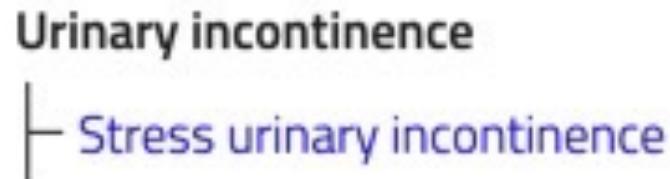
Urobiome Terminology



Are we studying the
urinary microbiome?
urine microbiome?
kidney microbiome?
bladder microbiome?
urethral microbiome?

→ *How is this captured in the metadata?*

Are the needed terms in the recommended ontologies?



Urinary incontinence HP:0000020

Loss of the ability to control the urinary bladder leading to involuntary urination.

Synonyms: *Loss of bladder control, Bladder incontinence*

Comment: Urinary incontinence can be defined as the complaint of any involuntary leakage of urine.

Are the needed terms in the recommended ontologies?

Urinary urgency HP:0000012

Urge incontinence is the strong, sudden need to urinate.

Synonyms: *Overactive bladder, Urgency frequency syndrome, Overactive bladder syndrome, Urinary urgency*

Comment: Urinary urgency is the strong, sudden need to urinate and is usually due to bladder spasms or contractions. This symptom is suggestive of, but not necessarily conclusive for urodynamically demonstrable detrusor hyperactivity.

Urobiome Consensus

Forming Consensus To Advance Urobiome Research

Element(s)	Required/desired	Description ^a
Biological elements		
Age	Required	Age in years or months/days if appropriate for infant/young child population ^b
Sex	Required	Biological sex; gender if relevant for the study
Antibiotic usage	Desired	There is a lack of knowledge about postantibiotic microbiome recovery; when possible, we recommend recording of use in the prior 3 months or length of time between last antibiotic exposure and sample collection
Hormone status	Desired	<p>Pubertal stage^a</p> <p>Pregnant/postpartum</p> <p>Menopausal status: perimenopausal, postmenopausal</p> <p>Also specify if taking supplemental hormones (estrogen) and route (oral, transdermal, or vaginal, etc.)</p> <p>Last menstrual period (if menstruating)</p>
Contraception	Desired	Use of oral contraceptives, other hormonal or nonhormonal/barrier, or none
Body mass index	Desired	Body mass index at the time of the study visit, calculated from height and weight



Forming Consensus To Advance Urobiome Research

Brubaker et al. *mSystems*. 2021 Aug 31;6(4):e0137120

Proposed Urobiome Metadata Standards



variable	description
host_age	Age of host at the time of sampling; relevant scale depends on species and study.
host_sex	Biological sex of the host.
recent_abx	Antibiotic usage in the prior 3 months. If yes, indicate if possible the date of the last dose
menstrual_status	If applicable (studies on female urobiome), indicate either: - Pre-menarche (optional, specify Tanner stage 1-5) - Menstruating - Pregnant - Postpartum (up to 1 year after delivery) - Post-menopausal
suppl_hormone	Any supplemental hormone delivered in any routes. If possible, specify the route e.g. oral, transdermal, vaginal, etc.
contraception	Use of any oral contraceptives, other hormonal, non-hormonal/barrier or none. Suggested ontology: systemized nomenclature of medicine (SNOMED) http://bioportal.bioontology.org/ontologies/SNOMEDCT
host_body_mass_index	Body mass index, calculated as weight/(height)squared

Using Proposed Urobiome Metadata Standards



Forming Consensus To Advance Urobiome Research

Brubaker et al. *mSystems*. 2021 Aug 31;6(4):e0137120

Supplemental Material

A screenshot of a Microsoft Excel spreadsheet. The top row contains column headers: 'sample_name' (in green), 'host_age' (in yellow), 'host_sex' (in yellow), and 'host_gender' (in yellow). The first row of data below the header (row 1) also has these four columns. Row 2 is empty. Rows 3 through 7 are also empty. The top of the Excel window shows standard ribbon icons for file, home, insert, etc., and an 'AutoSave' button set to 'OFF'. The background of the slide features a light gray grid pattern.

	A	B	C	D
1	sample_name	host_age	host_sex	host_gender
2				
3				
4				
5				
6				
7				

Demographics

Host Age

Age of host at the time of sampling

Example: If the host is 20 years old, enter 20

* must provide value

Numerical value only

Host Age (Units)

Units for age of host at the time of sampling

Example: If the host is 20 years old, enter years

* must provide value



Metadata standards to make urobiome research FAIR



Forming Consensus To Advance Urobiome Research

TABLE 1 Proposed elements to be included in the minimum metadata standards for reporting of urobiome research

Element(s)	Required/desired	Description ^d
Biological elements		
Age	Required	Age in years or months/days if appropriate for infant/young child population ^b
Environmental variables		
Method of collection	Required	Void, collection device (Peezy) Catheter (use of Mitrofanoff ^a)
Technical variables		
Date and time of collection ^c with conditions	Required	Used to ensure that samples stored at room temp for long periods are highlighted as such, potentially impacting the validity of results

Hands-on Introduction to Urobiome Metadata Standards workshop

May 11th 8 am PT / 11 am ET

Virtual - Register!

bit.ly/urobiome-metadata-hands-on-2022

Urobiome 101 with Drs. Naz Siddiqui, Alan Wolfe, and Linda Brubaker (in person, AUGS/IUGA June 14th 2022)

Urobiome 2023 – TBD!

Contributors:

The Karstens Lab:

Lisa Karstens, PhD

Erin Dahl

bit.ly/KarstensLab

The NMDC team:

Pajau Vangay

microbiomedata.org/team

Resources and references:

bit.ly/urobiome-metadata-pres-2022

Register for hands-on metadata workshop!

bit.ly/urobiome-metadata-hands-on-2022



Thank you!

Questions?



karstens@ohsu.edu



bit.ly/KarstensLab



[@lisakarstens](https://twitter.com/lisakarstens)