



nmdc

National Microbiome
Data Collaborative

Urobiome Metadata Standards Workshop

Lisa Karstens, PhD

Assistant Professor

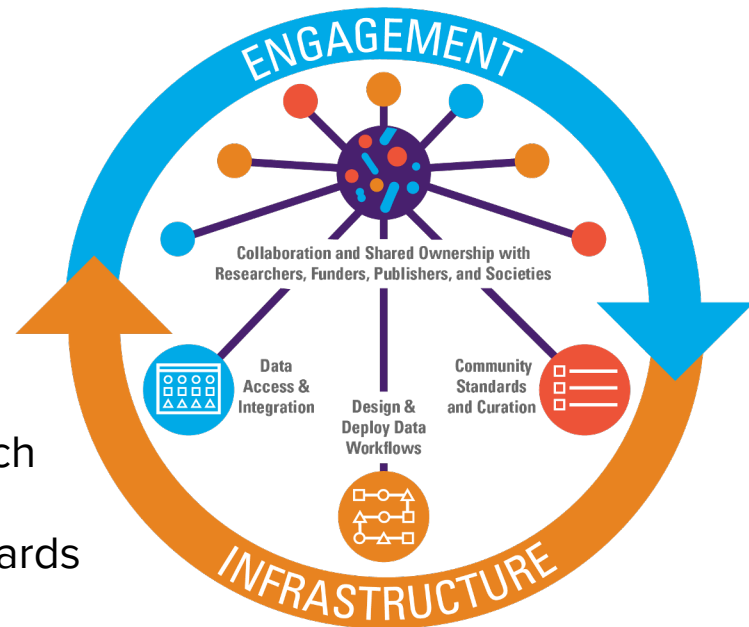
Department of Medical Informatics and Clinical Epidemiology, OHSU

NMDC Ambassador

June 22, 2022

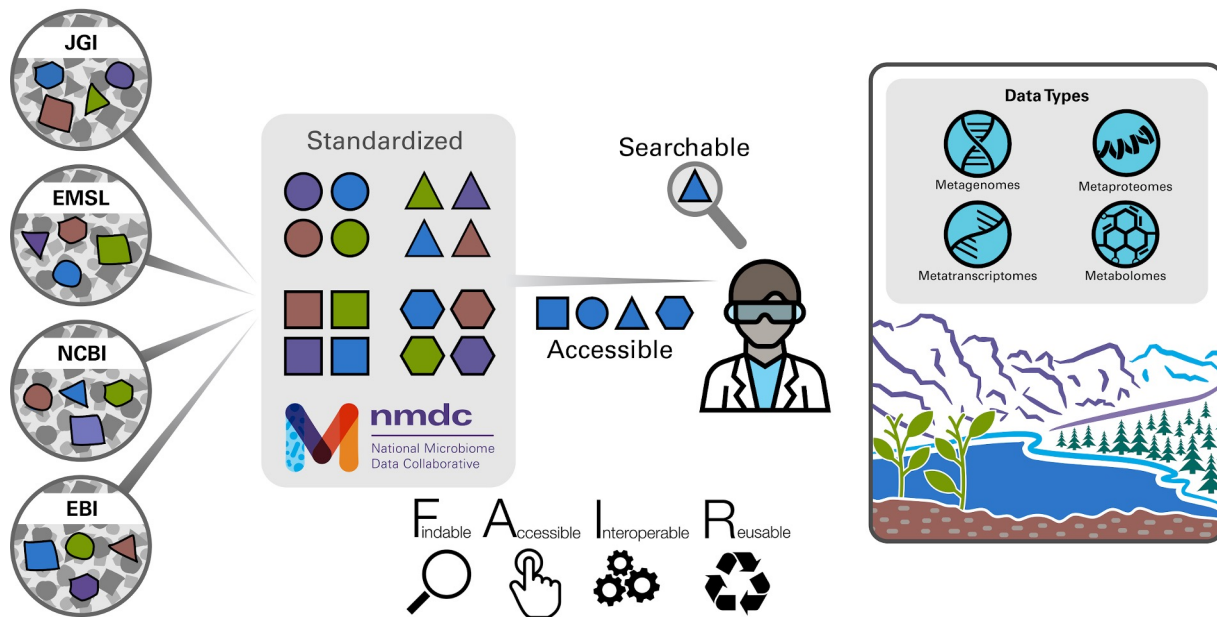
Overview

- Overview of NMDC Mission
- **Icebreaker:** get to know your groups!
- Overview of Metadata Standards
- **Activity:** Designing a metadata table
- Current Metadata Standards for Urobiome Research
- **Activity:** Completing a metadata table using standards
- Group discussion



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

Found 638 results.



Study

PI Name

Sample

Depth

Latitude

Longitude

Geographic Location Name

GOLD Ecosystems

GOLD classification

ENVO

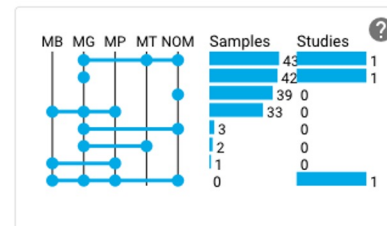
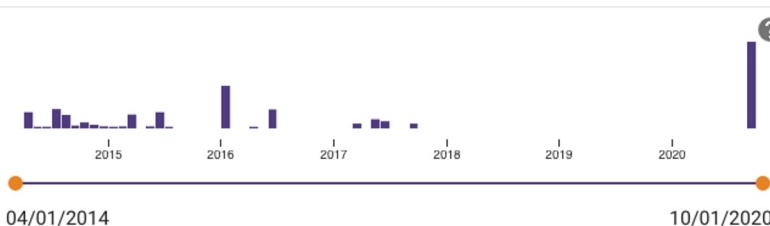
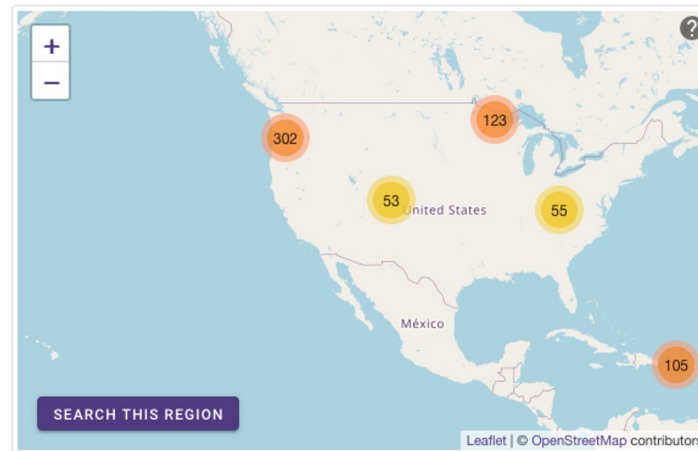
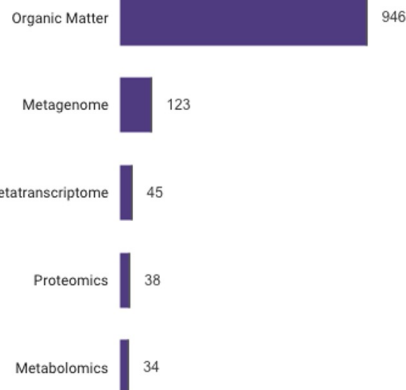
Environmental biome

Environmental feature

Environmental material

OMICS

ENVIRONMENT

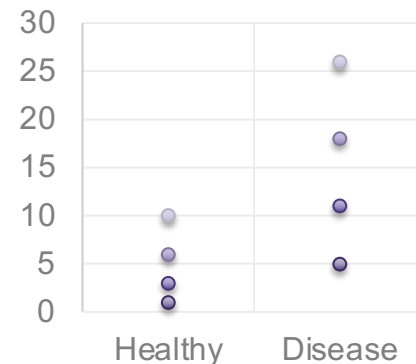


Ice breaker! Join your breakout group.

Go around and let everyone know:

- 1. Your name**
- 2. Your institution**
- 3. Your position**
- 4. Why you are here**

CTCTTG TCGGAATACCTTA ATGCGATTACCGGATCGGTAATTCGCGAATACCTTA ATCGGATTACCGGA
ACTACGCTTAAT ATCCTTACTAGCGCTACGAGCTTAAT ATCCTTACTAGCGCT TGGACTA
ATCGCATTACCGGA ATGCGATTACCGGATCGGATTACCGGA ATCGGATTACCGGA
AGCTTTTC ATCCTTACTAGCGCT AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG
TCGTTGCAACTAGC TCGAAGGAATTAG TCGTTGCAACTAGC TCGAAGGAATTAG
ATCGCATTACCGGA ATGCGATTACCGGATCGGATTACCGGA ATCGGATTACCGGA
CTAGCTTAAT ATCCTTACTAGCGCT AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG
CTCTTG TCGGAATACCTTA ATGCGATTACCGGATCGGTAATTCGCGAATACCTTA ATCGGATTACCGGA
ACTACGCTTAAT ATCCTTACTAGCGCTACGAGCTTAAT ATCCTTACTAGCGCT TGGACTA
ATCGCATTACCGGA ATGCGATTACCGGATCGGATTACCGGA ATCGGATTACCGGA
AGCTTTTC ATCCTTACTAGCGCT AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG
TCGTTGCAACTAGC TCGAAGGAATTAG TCGTTGCAACTAGC TCGAAGGAATTAG
TCGTTGCAACTAGC TCGAAGGAATTAG TCGTTGCAACTAGC TCGAAGGAATTAG
CTCTTG TCGGAATACCTTA ATGCGATTACCGGATCGGTAATTCGCGAATACCTTA ATCGGATTACCGGA
ACTACGCTTAAT ATCCTTACTAGCGCTACGAGCTTAAT ATCCTTACTAGCGCT TGGACTA
ATCGCATTACCGGA ATGCGATTACCGGATCGGATTACCGGA ATCGGATTACCGGA
AGCTTTTC ATCCTTACTAGCGCT AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG
TCGTTGCAACTAGC TCGAAGGAATTAG TCGTTGCAACTAGC TCGAAGGAATTAG
ATCGCATTACCGGA ATGCGATTACCGGATCGGATTACCGGA ATCGGATTACCGGA
CTAGCTTAAT ATCCTTACTAGCGCT AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG



How do we translate microbiome data into biological and clinical insights?

Sequencing data

```
CTCTTG TCCGAAATACCTTA ATGCGATTACCGGA  
ACTAGCTTAAT ATCCTTACTAGCGCT TGGACTA  
ATGCGATTACCGGA ATGCGATTACCGGA  
AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG  
TCGTTGCACTAGC TCGAAGGAATTAGG  
ATGCGATTACCGGA ATGCGATTACCGGA  
CTAGCTTAAT ATCCTTACTAGCGCT AGCTCGCTT  
CTCTTG TCCGAAATACCTTA ATGCGATTACCGGA  
ACTAGCTTAAT ATCCTTACTAGCGCT TGGACTA  
ATGCGATTACCGGA ATGCGATTACCGGA  
AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG  
TCGTTGCACTAGC TCGAAGGAATTAGG  
ATGCGATTACCGGA ATGCGATTACCGGA  
CTCTTG TCCGAAATACCTTA ATGCGATTACCGGA  
ACTAGCTTAAT ATCCTTACTAGCGCT TGGACTA  
ATGCGATTACCGGA ATGCGATTACCGGA  
AGCTTTTC ATCCTTACTAGCGCT AGCTCGCTTG  
TCGTTGCACTAGC TCGAAGGAATTAGG  
ATGCGATTACCGGA ATGCGATTACCGGA  
CTAGCTTAAT ATCCTTACTAGCGCT AGCTCGCTT
```



Data table

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

Metadata contains the context we need to understand data.

Data table

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

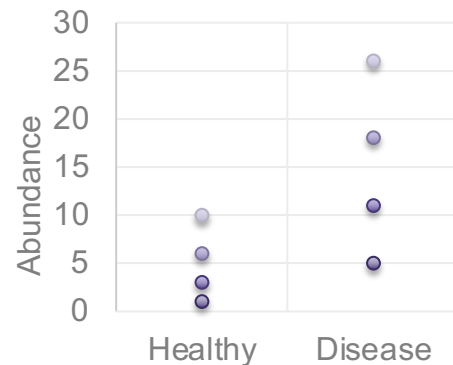


Metadata table

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

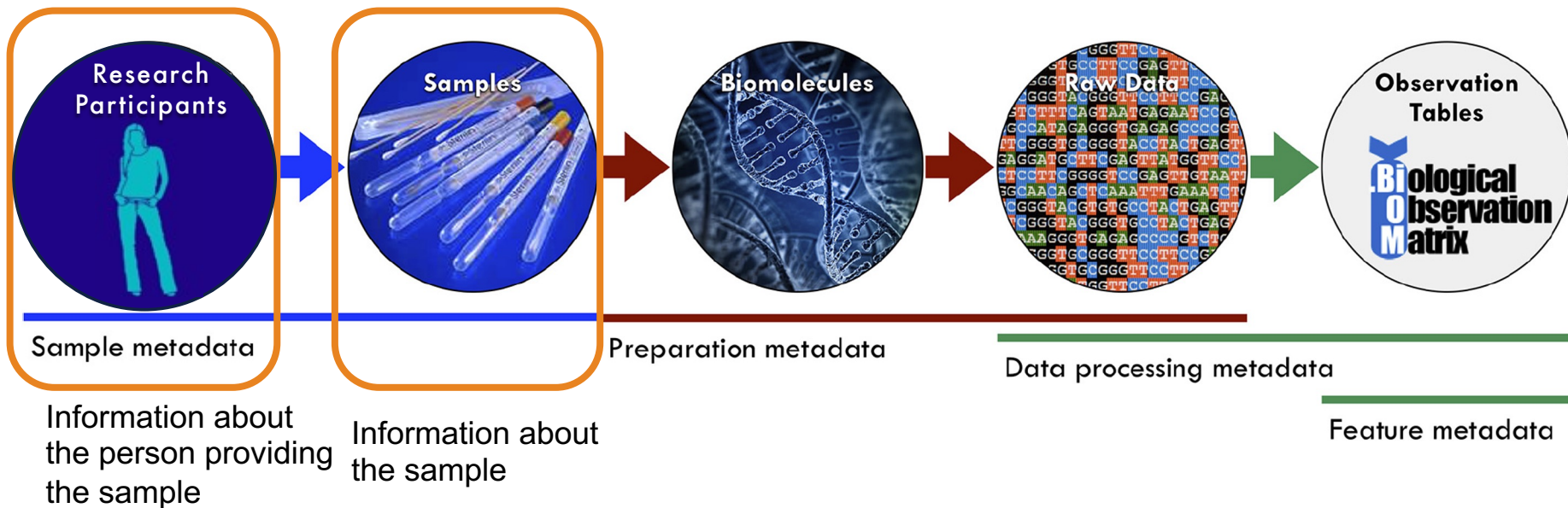


Abundance of
Bacteria 1 by group



What are Metadata ?

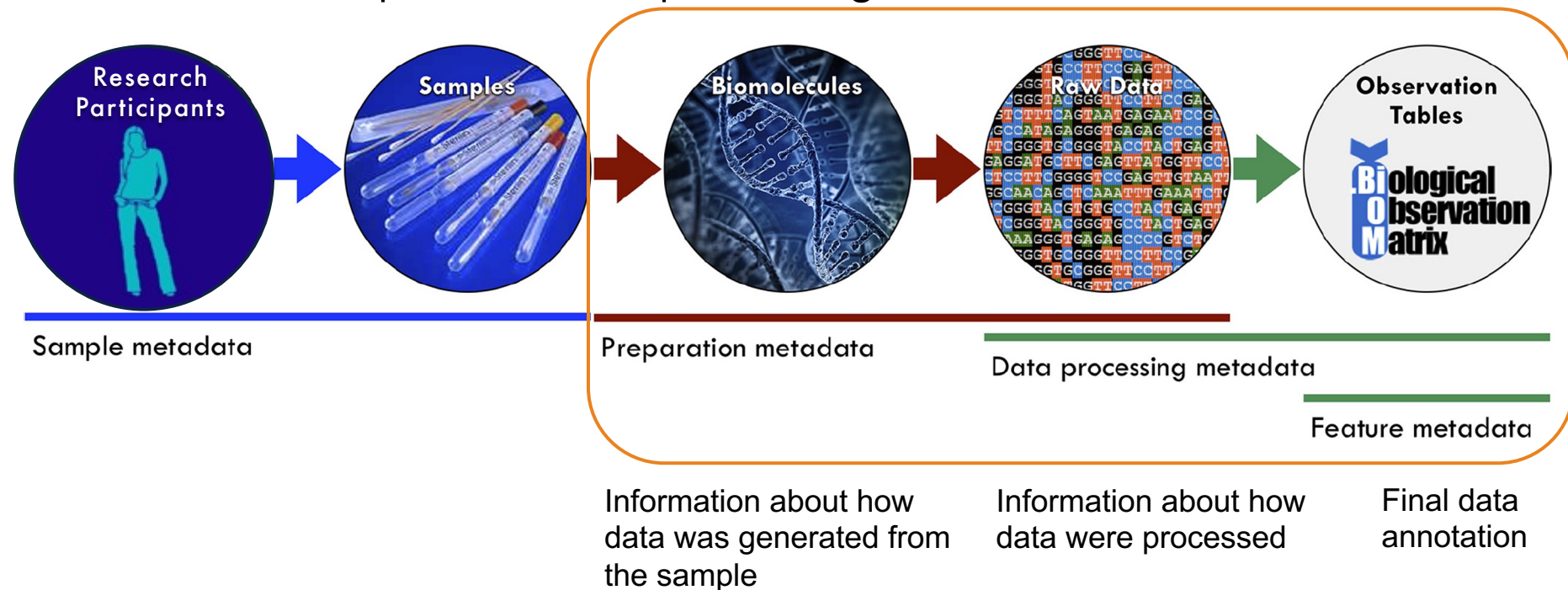
Microbiome Metadata - Sample metadata



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 ?

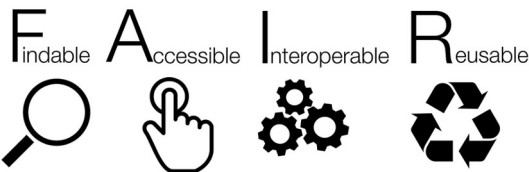
Not covered – Preparation, data processing, and feature metadata



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 from



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?

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

Example urobiome research study

You are a researcher interested in if the urobiome is different in women that have urgency urinary incontinence compared to women with no known lower urinary tract disorders.

To address this question, you design a cross-sectional study. You will collect urine from women who enroll. The samples will be processed and subject to 16S rRNA amplicon sequencing to determine the bacterial composition of the samples. Statistical analysis will be performed to identify group differences between the bacterial composition of the urobiome from women with UUI compared to women without UUI.

TASK: Identify the columns you will want for your metadata table

Breakout group activity

Individually create the column headers for the example urobiome research study's metadata table.

What information do you want to collect **about the participant** providing the samples?

What information do you want to collect **about the samples**?

→ **Discuss your metadata table in breakout groups and come up with a consensus**

Breakout group activity

Review metadata tables created by each group

Proposed Urobiome Metadata Standards

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
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	Pubertal stage ^a
		Pregnant/postpartum
		Menopausal status: perimenopausal, postmenopausal
		Also specify if taking supplemental hormones (estrogen) and route (oral, transdermal, or vaginal, etc.)
		Last menstrual period (if menstruating)

More variables in **Table 1** → <https://journals.asm.org/doi/10.1128/mSystems.01371-20>

Proposed Urobiome Metadata Standards



Translating proposed elements into structured metadata

variable (column name)	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: <ul style="list-style-type: none">- 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.

More variables in **Supplemental Material** → <https://journals.asm.org/doi/10.1128/mSystems.01371-20>

Creating a metadata table

Translating proposed elements into structured metadata

host_age	host_sex	recent_abx	menstrual_status	suppl_hormone	...
67 years	female	6 months	postmenopausal	vaginal estrogen cream	

Steps for creating a metadata table for a research study

1. Create column headers for REQUIRED variables
2. Create column headers for DESIRED UROBIOME variables that are **relevant for the study**.
3. Create column headers for **any additional variables** that are relevant for the study
4. Complete the table using **structured terminology and formats** when possible

Creating a metadata table

Step 1. Create column headers for REQUIRED variables

sample_name	host_age	host_sex	samp_collect_method	geo_loc_name	lat_lon	collection_date	samp_freeze_interval	samp_store_temp	samp_preservative	dna_extraction	seq_method

Creating a metadata table

Step 1. Create column headers for REQUIRED variables

Are there any special considerations for human subject research?

Look for information that may be considered protected health information (PHI)

sample_name	host_age	host_sex	samp_collect_method	geo_loc_name	lat_lon	collection_date	samp_freeze_interval	samp_store_temp	samp_preservative	dna_extraction	seq_method

Creating a metadata table

Step 1. Create column headers for REQUIRED variables

Are there any special considerations for human subject research?

DATES and LOCATION can be considered PHI!

sample_name	host_age	host_sex	samp_collection_method	geo_location_name	lat_lon	collection_date	samp_freeze_interval	samp_store_temp	samp_preservative	dna_extraction	seq_method

Recommendation: use broad or offset dates and location

Creating a metadata table

Step 1. Create column headers for REQUIRED variables

Are there any special considerations for human subject research?

AGE over 90 years can be considered PHI!

sample_name	host_age	host_sex	samp_collect_method	geo_loc_name	lat_lon	collection_date	samp_freeze_interval	samp_store_temp	samp_preservative	dna_extraction	seq_method

Recommendation: use broad group for those over 90 years of age (>89)

Creating a metadata table

Step 1. Create column headers for REQUIRED variables

Are there any special considerations for human subject research?

Make sure the sample_name does NOT contain PHI!

sample_name	host_age	host_sex	samp_collect_method	geo_loc_name	lat_lon	collection_date	samp_freeze_interval	samp_store_temp	samp_preservative	dna_extraction	seq_method

Recommendation: use generic unique identifier

Creating a metadata table

Step 2. Create column headers for DESIRED UROBIOME variables that are **relevant for the study**.

menstrual_status	suppl_hormone	recent_abx	samp_dna_conc

Creating a metadata table

Step 3. Create column headers for additional variables that are **relevant for the study.**

menstrual_status	suppl_hormone	recent_abx	samp_dna_conc	host_disease_stat

Creating a metadata table

Step 4. Fill in the table.

Use **structured terminology/ ontologies** when available

sample_name	host_age	samp_collect_method	collection_date
S05_101_ur			

Environmental pack	Structured comment name	Package item	Definition	Expected value	Value syntax	Example
human-associated	host_age	host age	scale depends on species and study, e.g. Could be seconds for amoebae or centuries	value	{float} {unit}	30 years

Extensive list of available terms:

https://github.com/GenomicsStandardsConsortium/mixs/blob/main/release/excel/mixs_v6.xlsx

Creating a metadata table

Step 4. Fill in the table.

Use **structured terminology/ ontologies** when available

sample_name	host_age	samp_collect_method	collection_date
S05_101_ur			

Samp_collect_method - urobiome recommended terms:

- voided,
- catheter (specify type),
- urine collection device (Peezy),
- suprapubic aspirate

Creating a metadata table

Step 4. Fill in the table.

Use **structured terminology/ ontologies** when available

sample_name	host_age	samp_collect_method	collection_date
S05_101_ur			

The date on which the sample was collected; date/time ranges are supported by providing two dates from among the supported value formats, delimited by a forward-slash character; Recommended formats **Mmm-YYYY or YYYY-mm** ; e.g. **Jun-2022 or 2022-06**

Creating a metadata table

Step 3. Complete the table

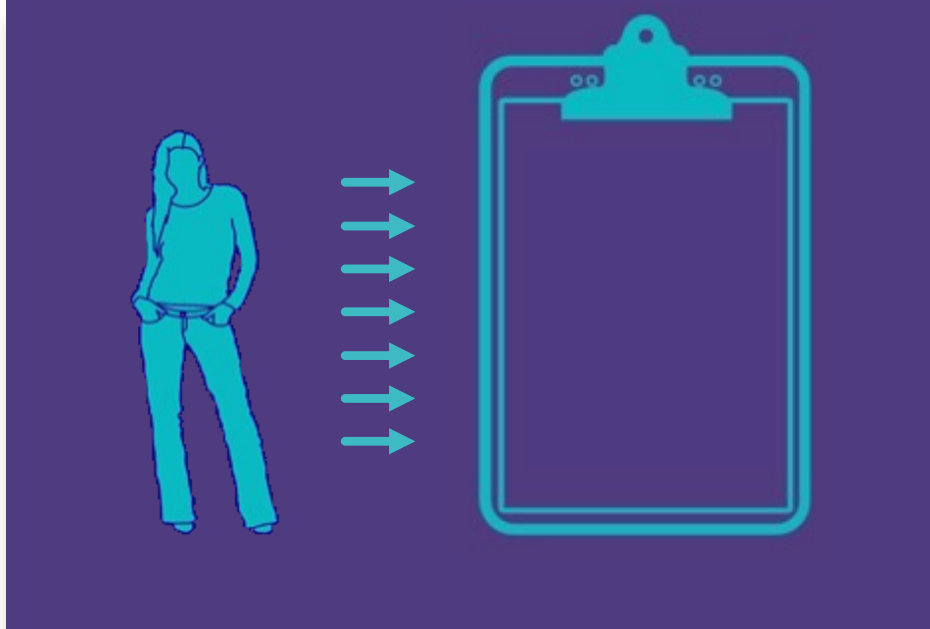
sample_name	host_age	samp_collect_method	collection_date
S05_101_ur			

Creating a metadata table

Step 3. Complete the table

sample_name	host_age	samp_collect_method	collection_date
S05_101_ur	42 years	voided urine sample	2022-06

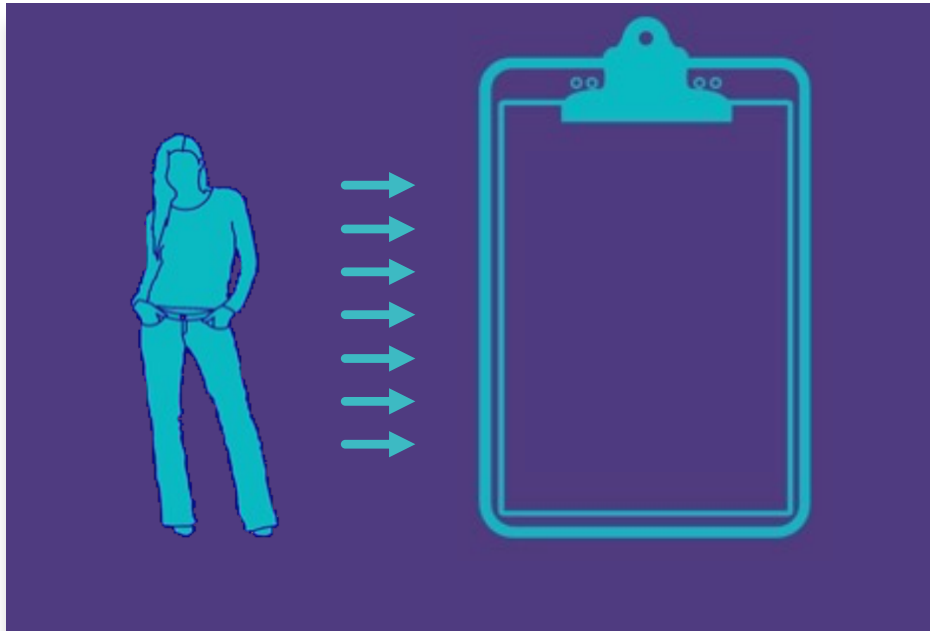
Breakout group activity



1. As a group, review the urobiome required and desired terms
2. How do the terms from the spreadsheet compare to the terms your group came up with? Are there extra terms? Missing terms?
3. Complete the proposed urobiome template for the example case study

Breakout group activity

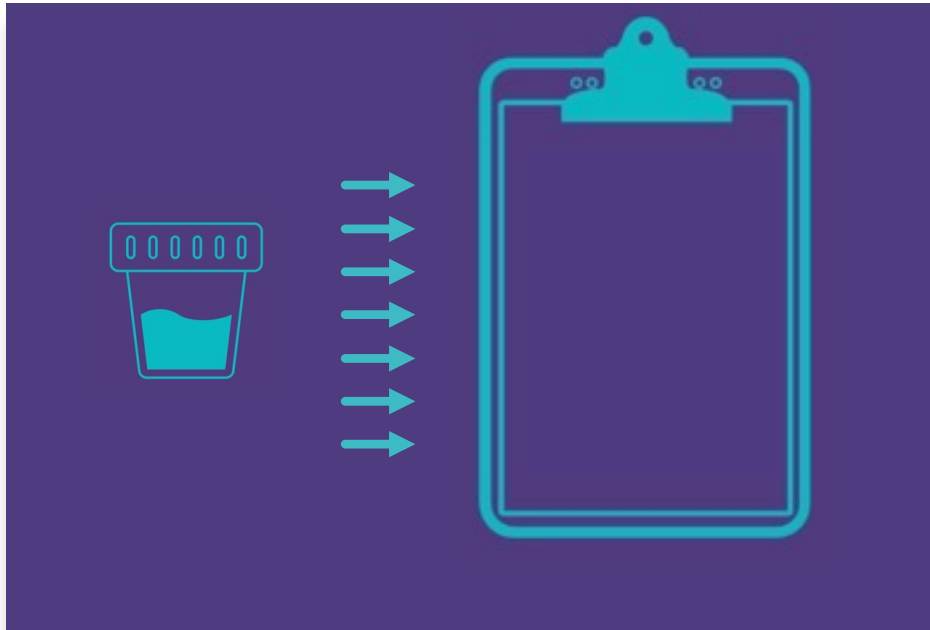
Complete the metadata table for the following example



A 75 year old woman is enrolled in the urobiome research study in Portland, OR. She has daily urgency urinary leakage. She provides a urine specimen today, which is collected by a nurse with a catheter. The participant has been on vaginal estrogen for the past 2 years. She also used antibiotics for a sinus infection 1 year ago.

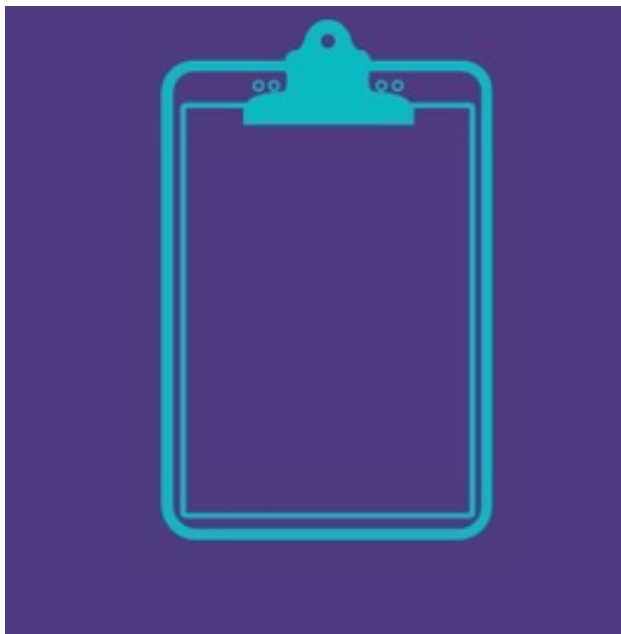
Breakout group activity

Complete the metadata table for the participant's samples below



After the study visit, the research coordinator takes the 35 mL urine sample and mixes it with AssayAssure (a preservative). They place the sample in the lab refrigerator. At the end of the day, the urine sample is centrifuged with all the other samples collected that day to form a pellet, which is frozen at -80. Once all the samples are collected for the study in 6 months, the sample undergoes DNA extraction with the urobiomeDNAextract kit. The sample has 10 ng/ul of DNA and is sent for sequencing with Illumina miseq.

Feedback for metadata



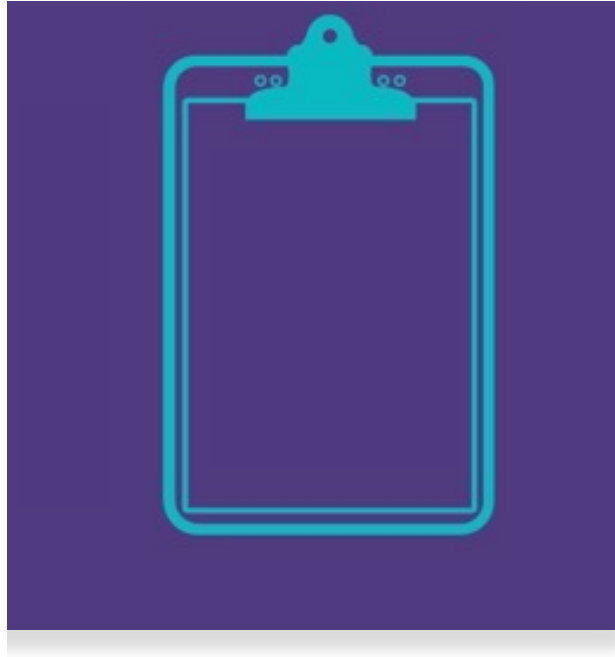
- How different were your metadata table compared to the urobiome standard?
- How was completing the urobiome metadata table?
- Did you all agree on how information should be captured?
- What information was not captured?

Resources for metadata



- Urobiome consensus manuscript
 - Discussion, definitions, and tables
 - Brubaker L, et al. **PMID: 34282932**
- Urology /urogynecology terms
 - International Continence Society (ICS) glossary
<https://www.ics.org/glossary#A>
- General terms
 - Genomic Standards Consortium
MiXS 6, start with human-associated packages
- Technical issues
 - Sequence Read Archive helpdesk

POST EVENT SURVEY



Contributors:

The Karstens Lab:

Lisa Karstens, PhD

Erin Dahl

bit.ly/KarstensLab

The NMDC team:

Pajau Vangay

Montana Smith

Setareh Sarrafan

microbiomedata.org/team

Resources and references:

https://karstenslab.github.io/post/metadata_resources/



Thank you!

Questions? Comments?



karstens@ohsu.edu



bit.ly/KarstensLab



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