

making

/ˈmeɪkɪŋ/

noun

the process of making or producing something.

metametrics

/ˈmɛtrɪk/

adjective

a system or standard of measurement.

more

/mɔː/

determiner & pronoun

a greater or additional amount or degree.

meaningful

/ˈmiːnɪŋfʊl,-f(ə)l/

adjective

having meaning.

serious, important, or worthwhile.

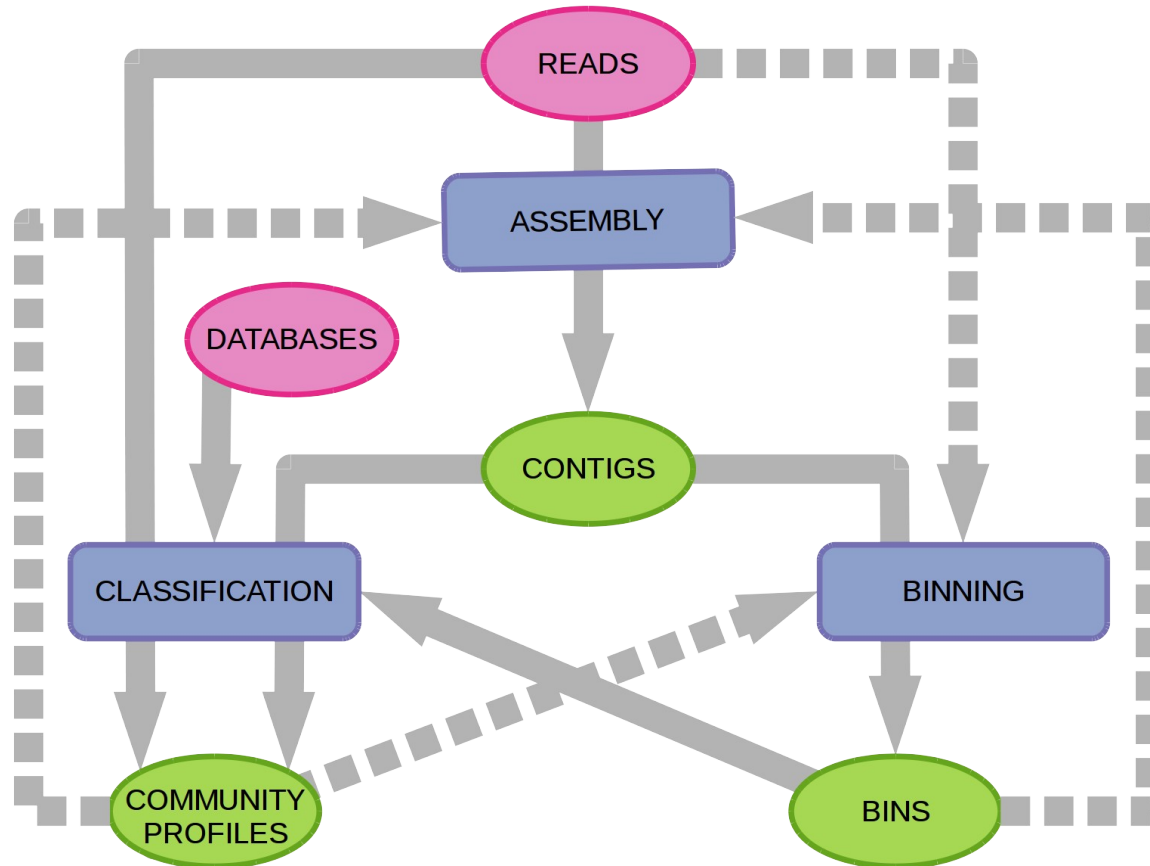
Dr. Michael Imelfort
ISME15 - Aug 25th 2014



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

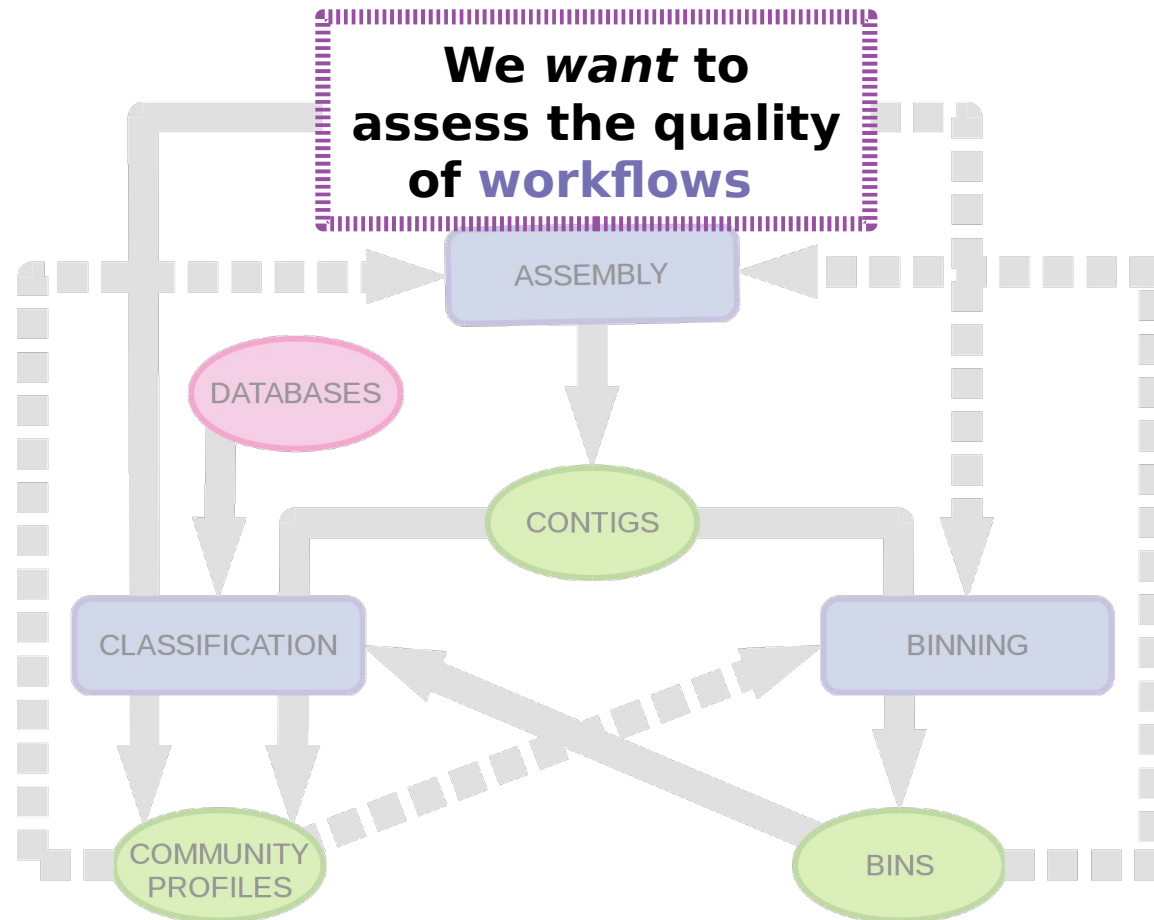


Metagenomic workflows, inputs and outputs

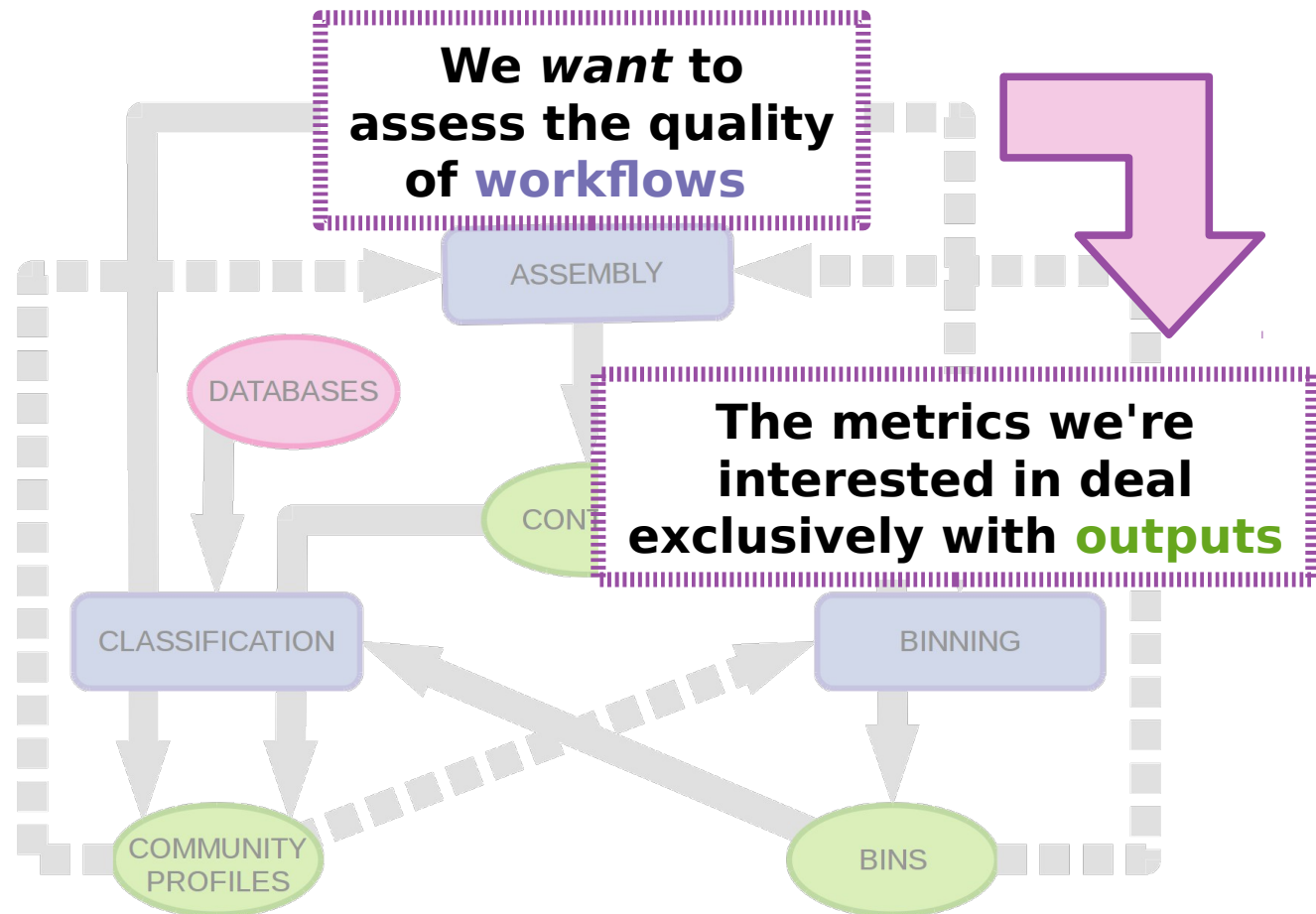


Initial CAMI targets: Assembly, taxonomic binning and profiling

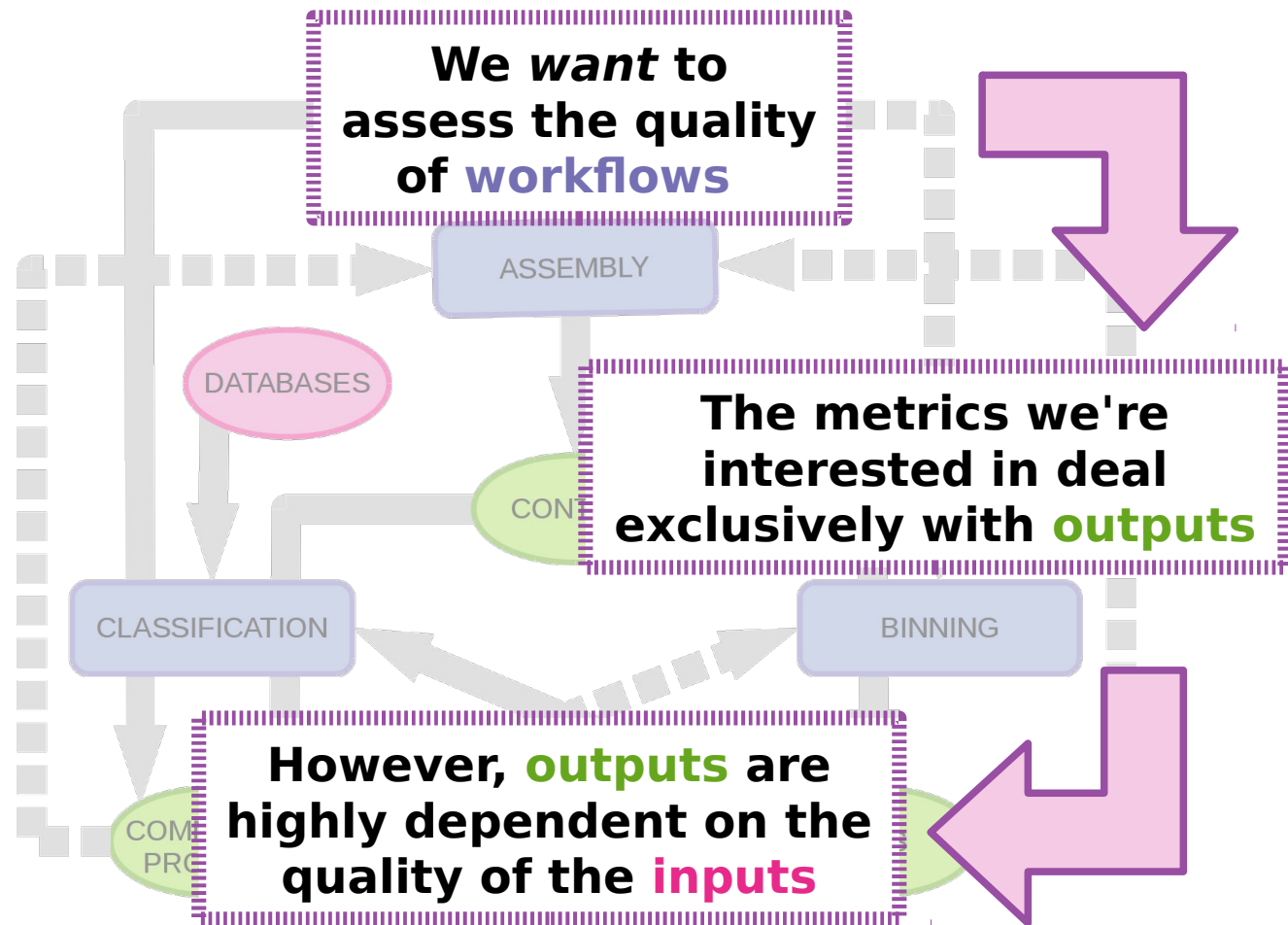
Metagenomic workflows, inputs and outputs



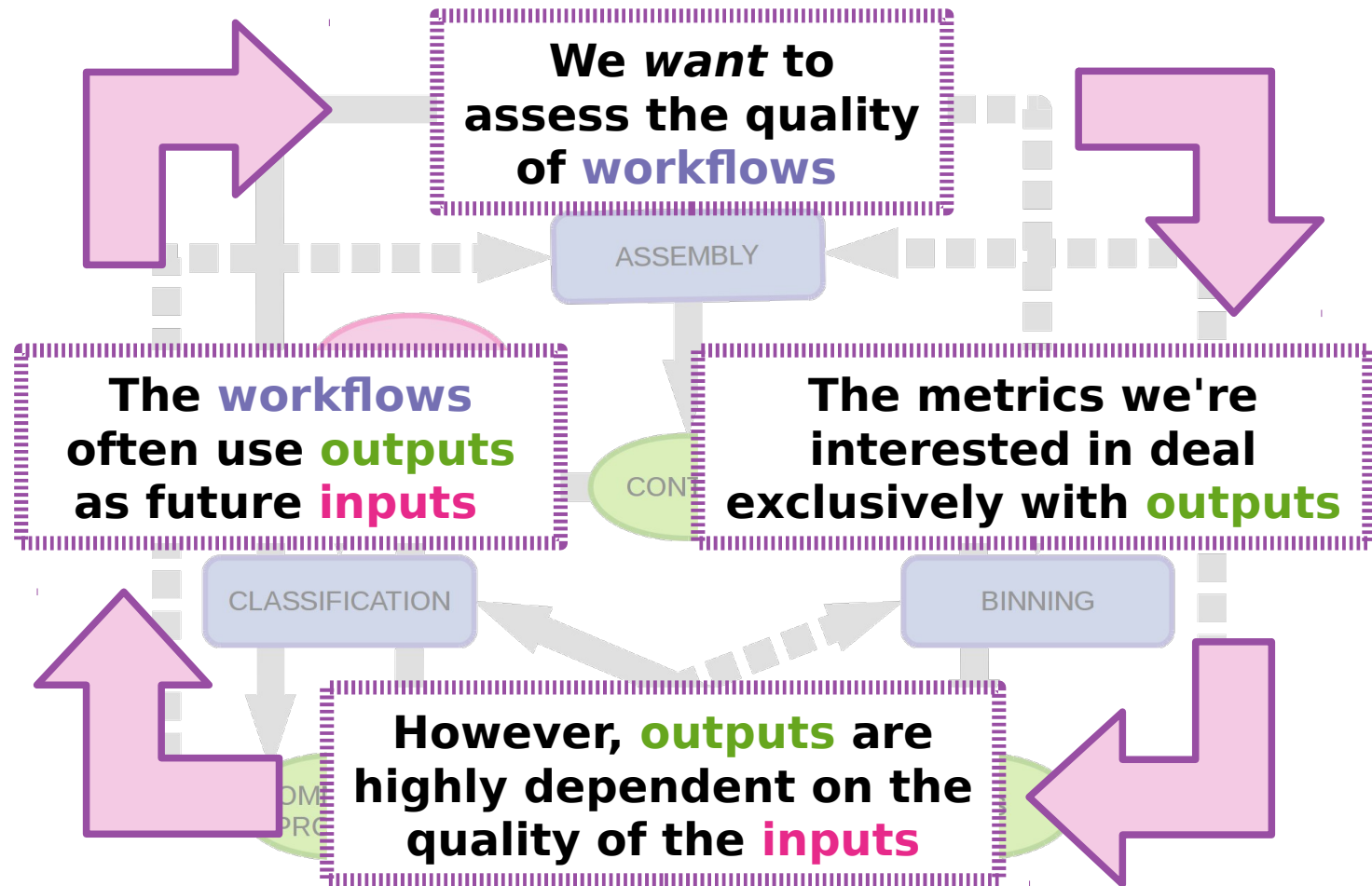
Metagenomic workflows, **inputs** and **outputs**



Metagenomic workflows, **inputs** and **outputs**



Metagenomic workflows, inputs and outputs



Metric mixing must be managed!

We must render unto the assembly algorithm the things that are caused by the assembly algorithm, and unto the binning algorithm the things that are caused by the binning algorithm

Good metrics should have clear **domains**

Some amount of mixing is unavoidable

We must be very careful

Must metametrics have domains?

How much of the final result can be attributed to community structure?

Good outcomes:	Rare / novel species recovered
Bad outcomes:	Strain nightmares

How much of the final result can be attributed to the structure of / bias in the reference databases?

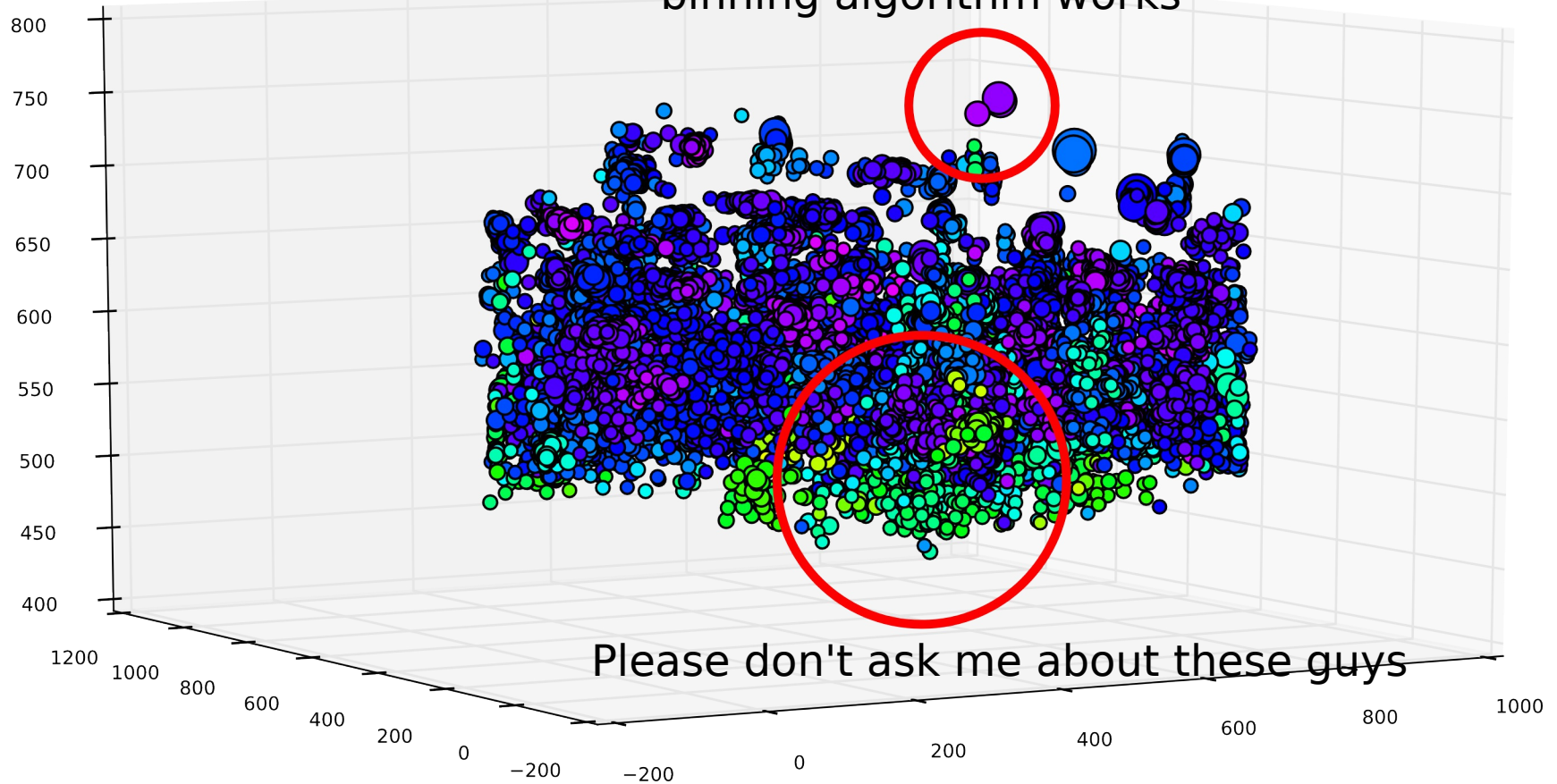
Who's to blame in the strain game?

Do we care about “rare” or “novel” species?

(from a metrics perspective)

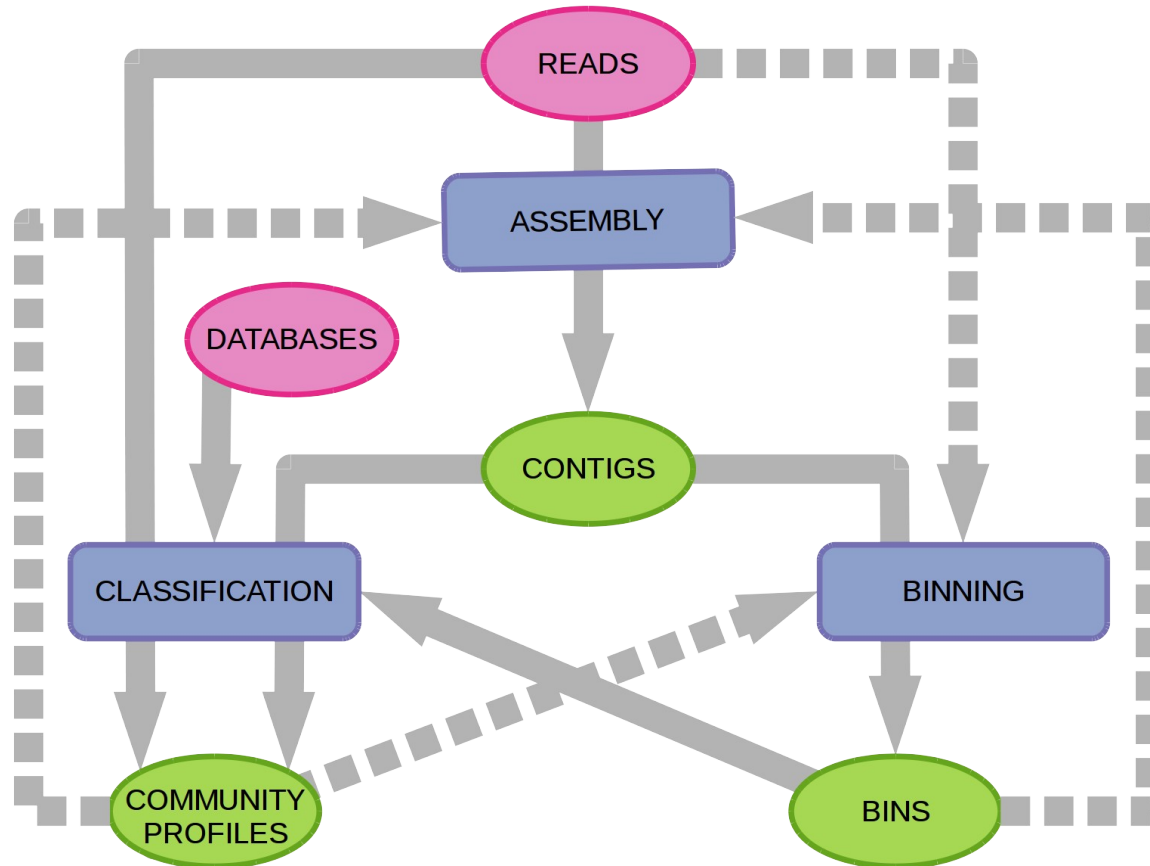
The strain game is THE game

WOW! Look at how well my
binning algorithm works

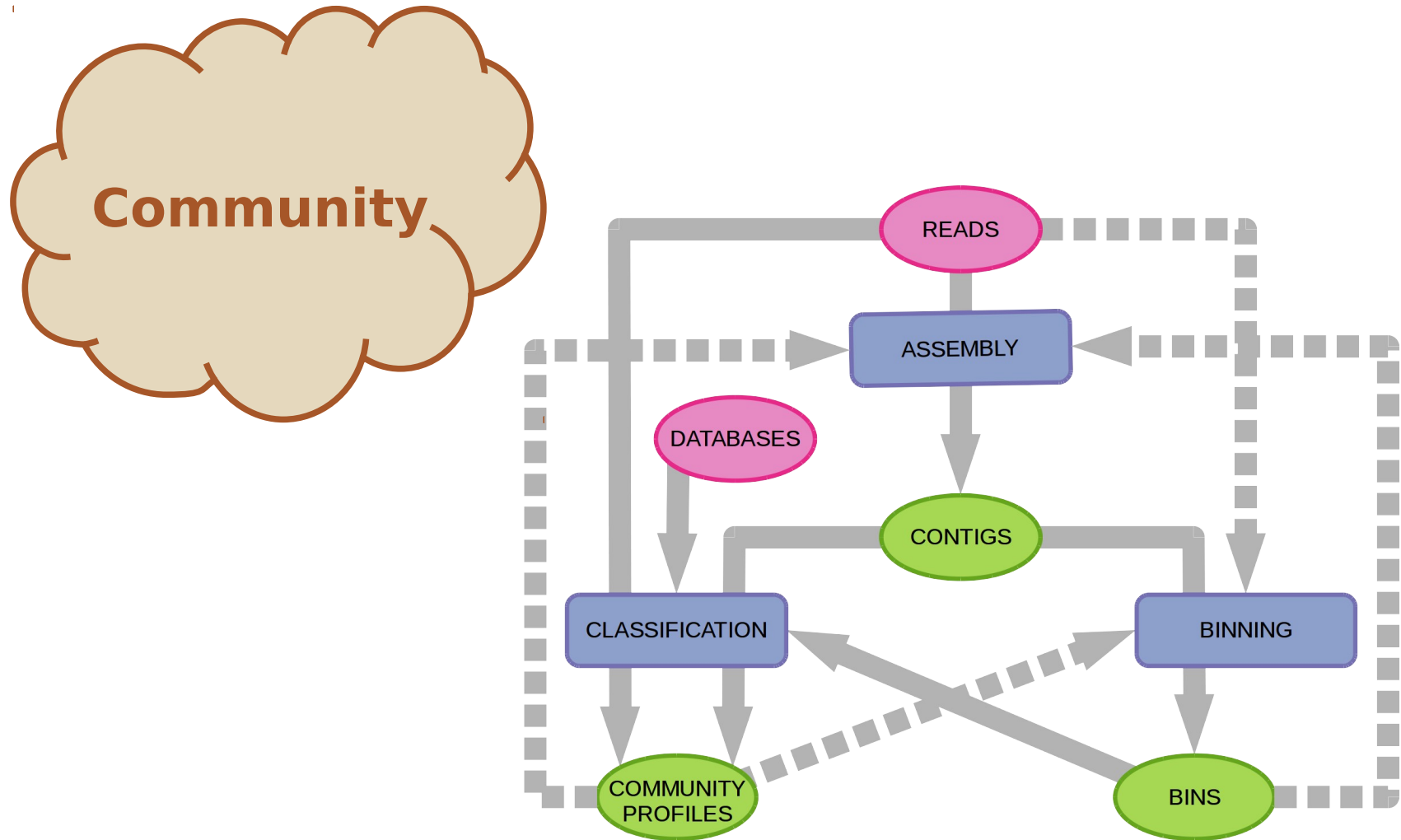


Please don't ask me about these guys

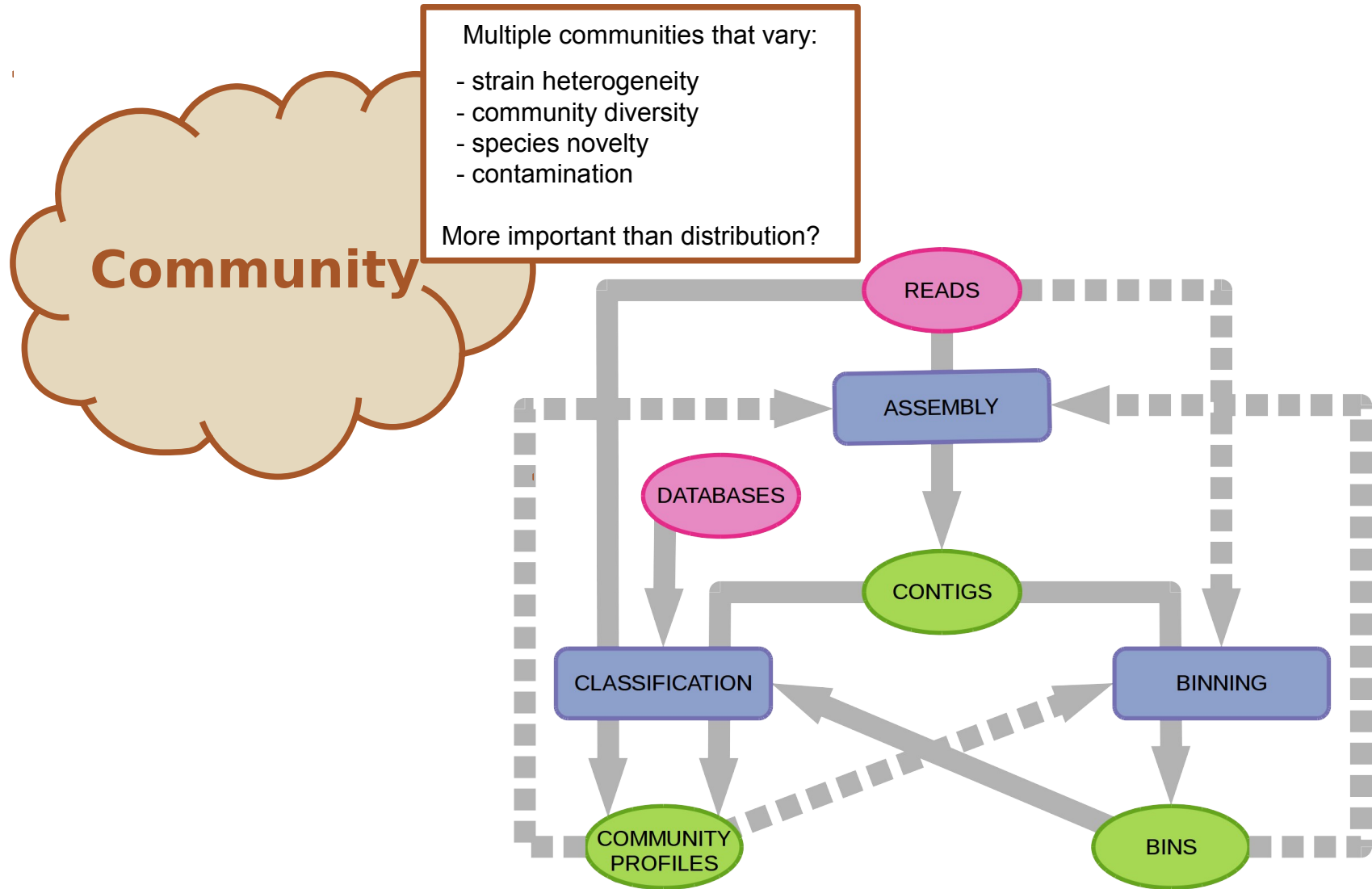
It all comes down to the test data



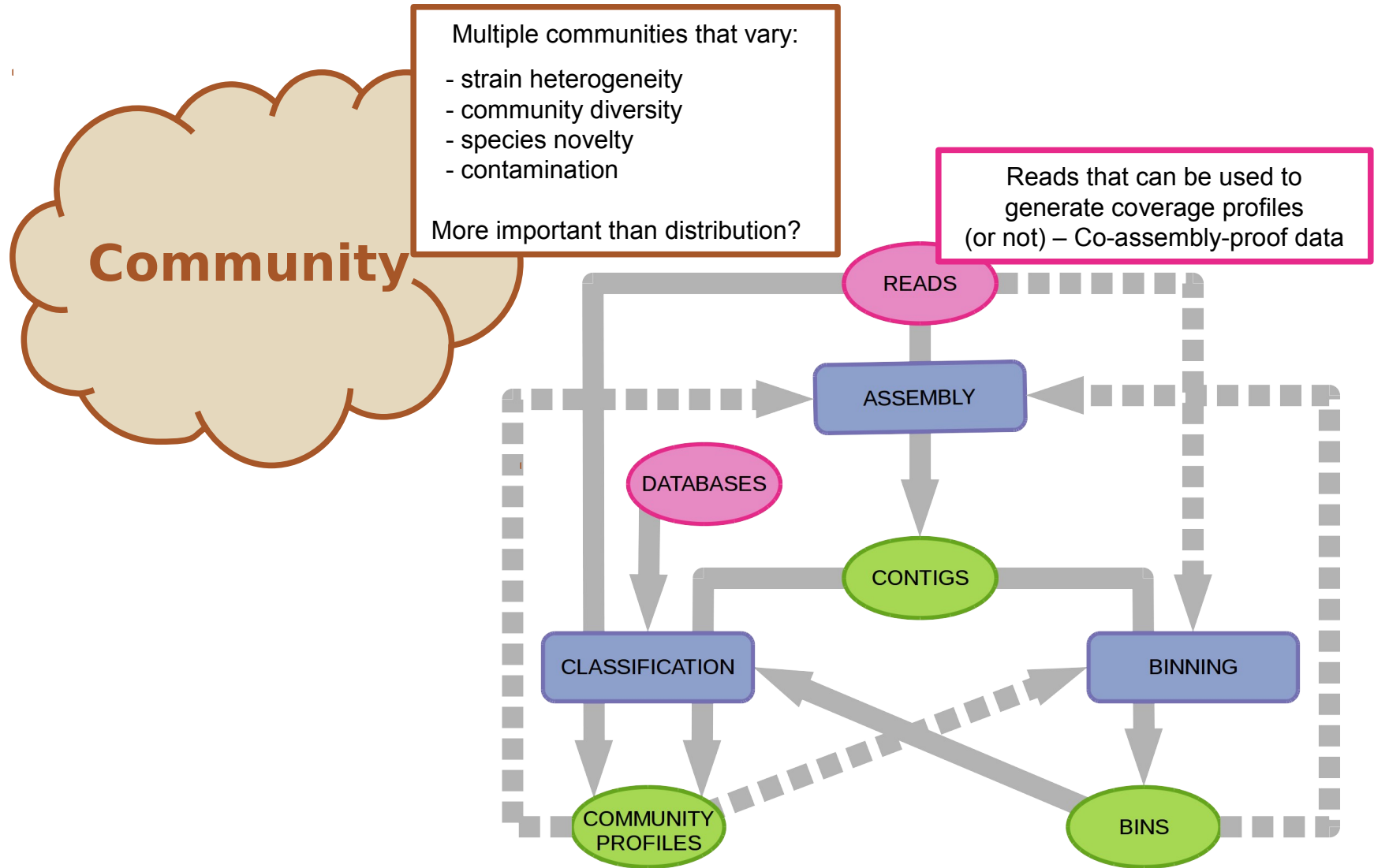
It all comes down to the test data



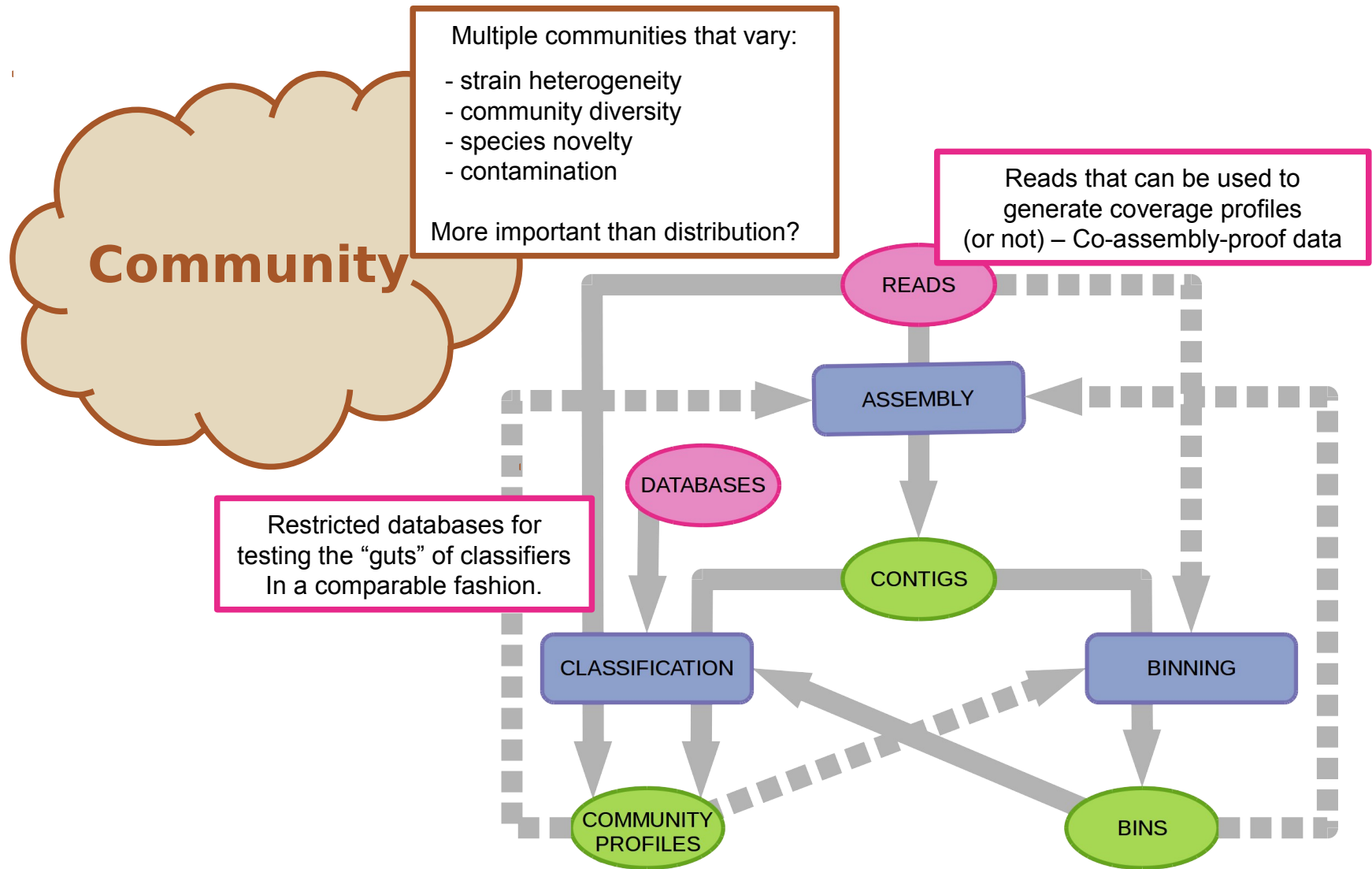
It all comes down to the test data



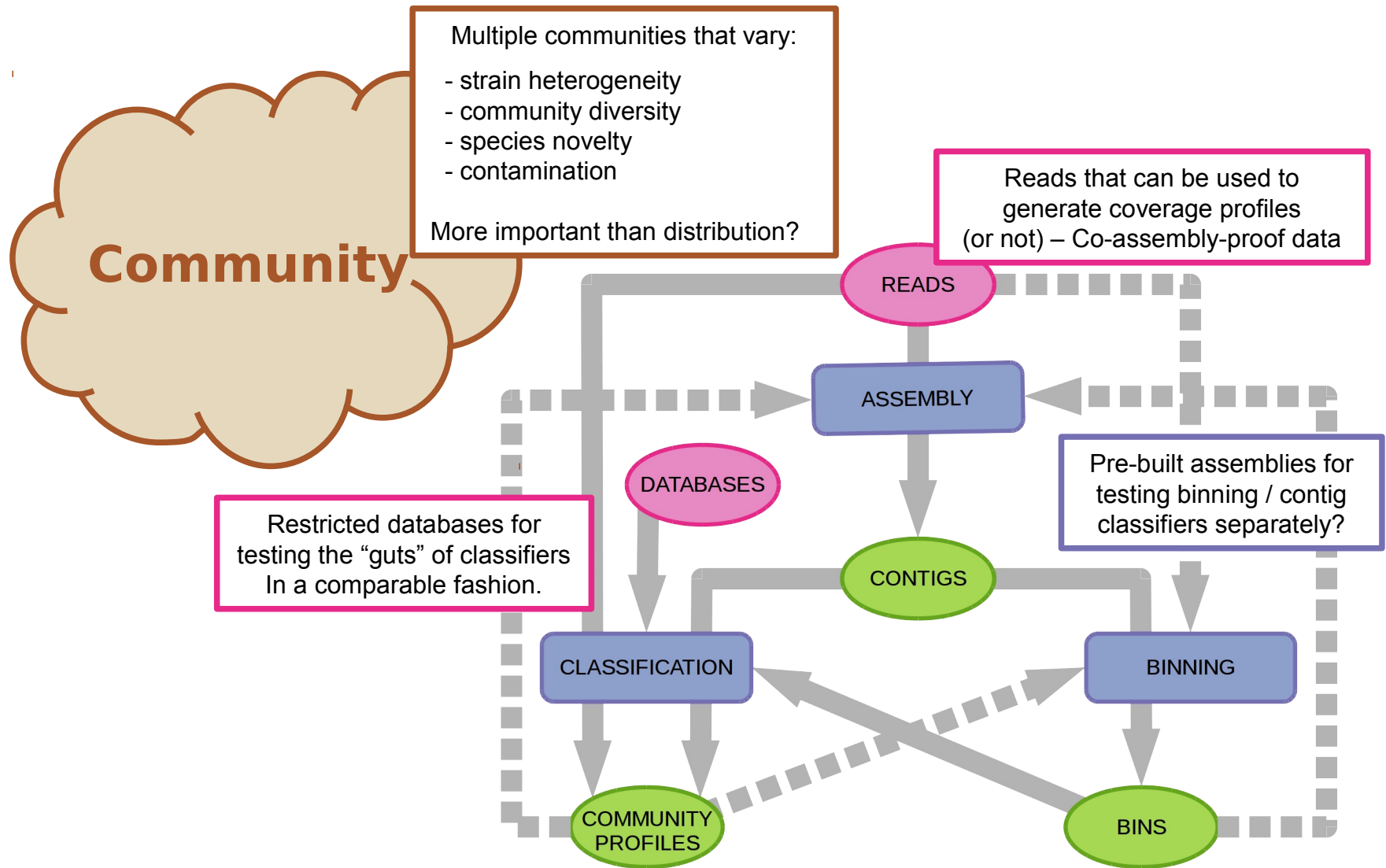
It all comes down to the test data



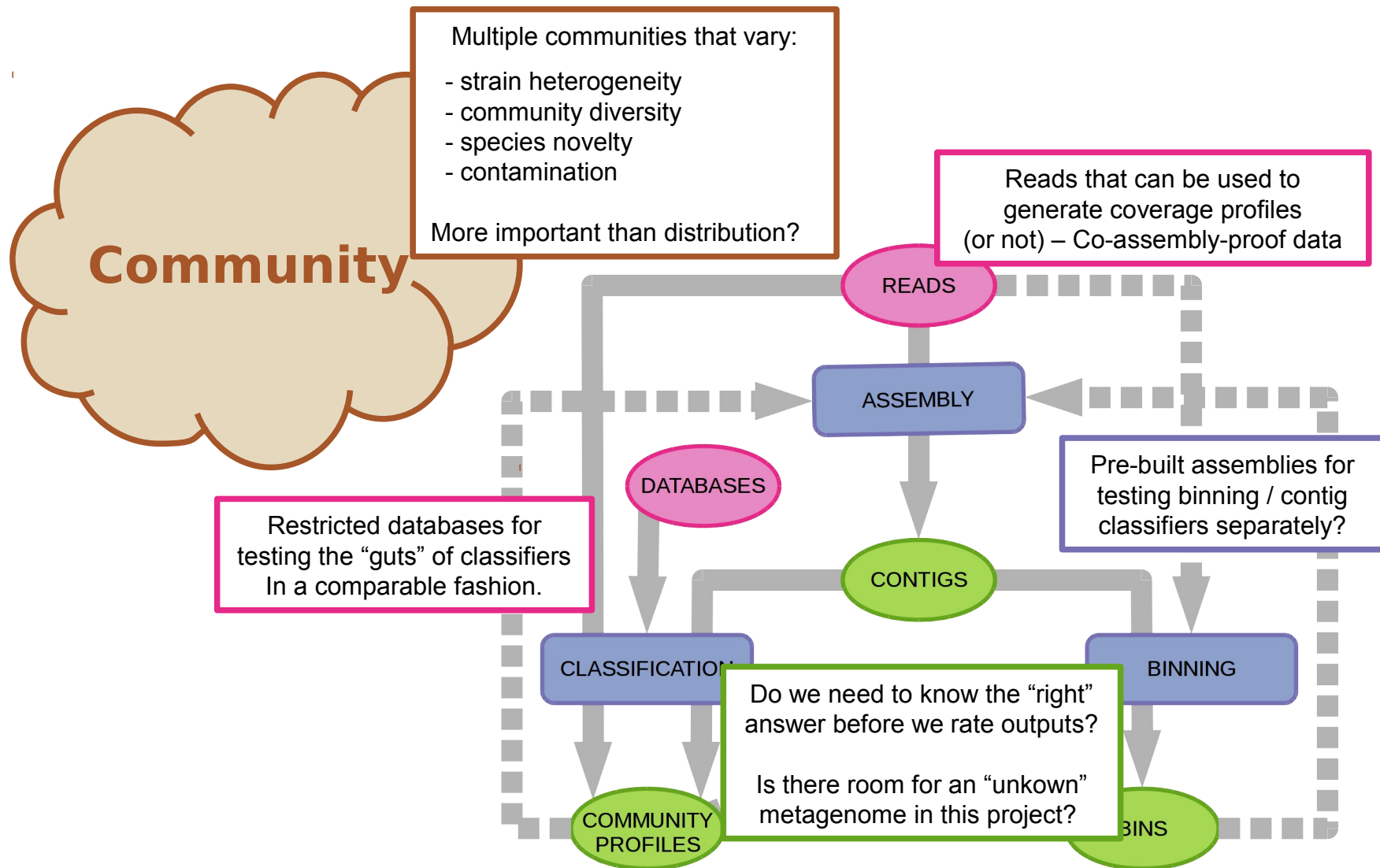
It all comes down to the test data



It all comes down to the test data



It all comes down to the test data



Test Data

Competition

Metrics

Test Data

- Intentional / unintentional biases
- Value of “real” Vs synthetic data
- Accessibility / scalability
(compute requirements etc.)

Competition

- Who are the consumers? What are their needs?

Developers Vs Users

Metrics

- Minimise mixing
- Accessibility / explanatory value

Final thoughts

An algorithm is not an island

Metric mixing must be managed

Good metrics should have clear domains

Knowing what we intend to measure
must *drive how* we build the test data

We shouldn't be distracted by “rare” or novel species

Strains are **the** issue facing metagenomics

Thanks for listening?

An algorithm is not an island

Metric mixing must be managed

Good metrics should have clear domains

Knowing what we intend to measure
must *drive how* we build the test data

We shouldn't be distracted by “rare” species

Strains are **the** issue facing metagenomics