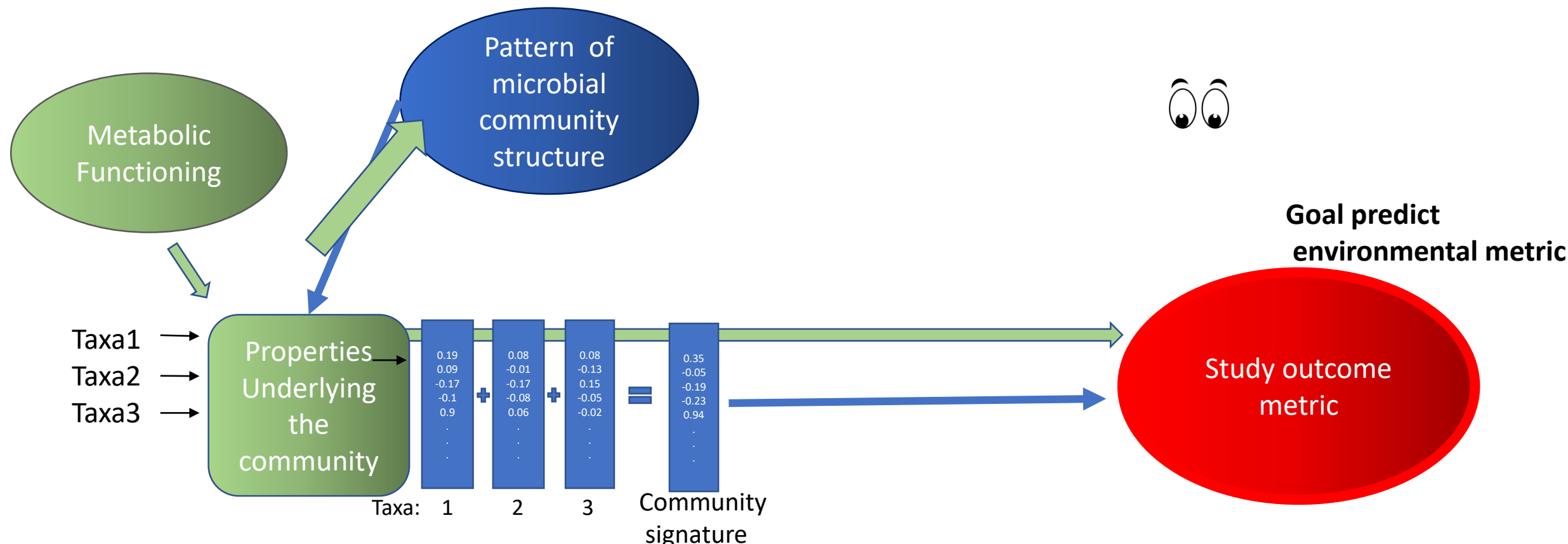


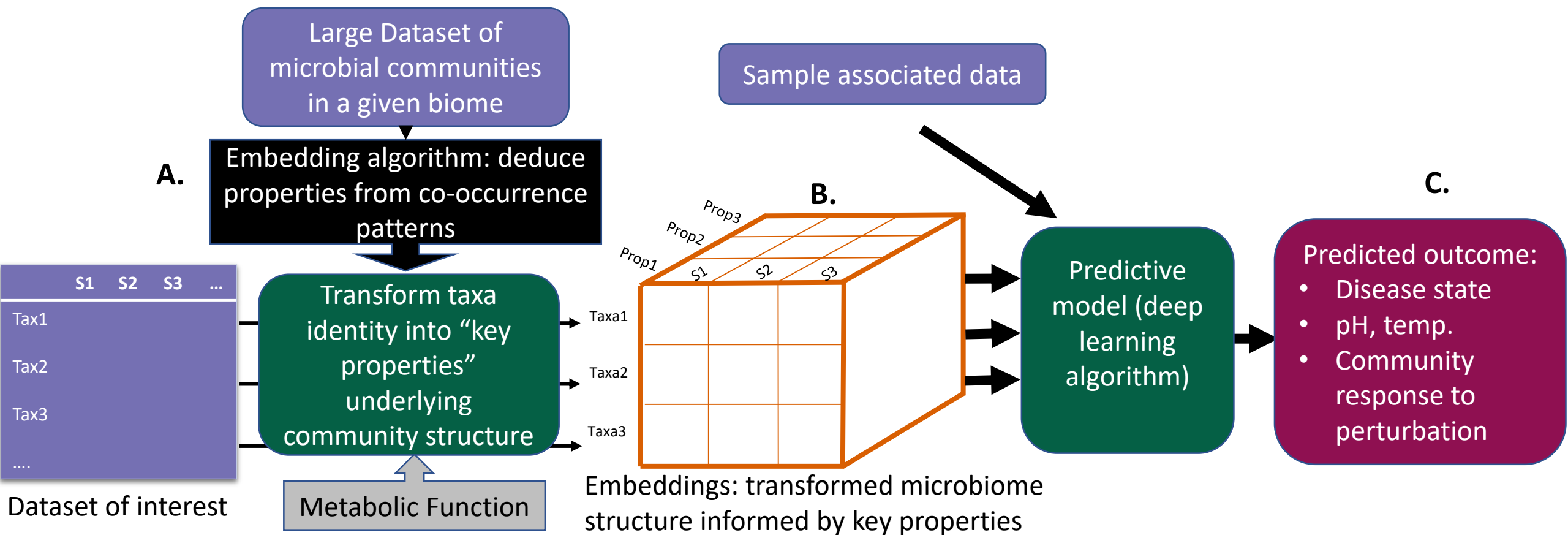
Model of information flow underlying framework

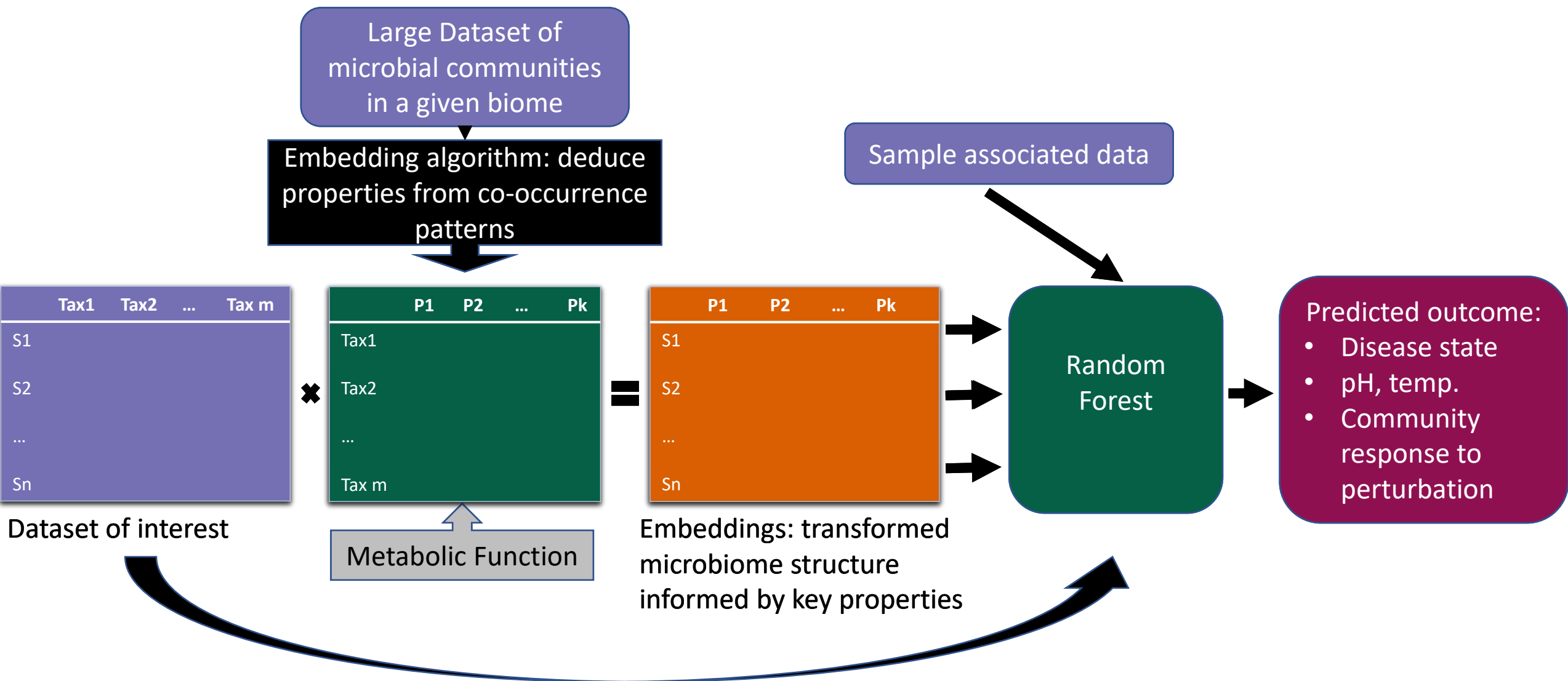
Green – Concepts indirectly deducible from data (Information flowing but data not observable)

Blue – Concepts directly accessible from data (Observed data and deduction from that data)

Red - goal







Public – no metadata

Large Dataset of microbial communities in the gut (AGP)

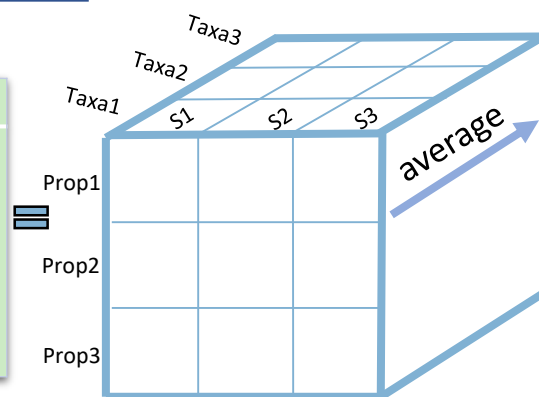
Embedding algorithm: deduce properties from co-occurrence patterns

Dataset of interest

	Tax1	Tax2	...	Tax m
S1				
S2				
...				
Sn				



	P1	P2	...	Pk
Tax1				
Tax2				
...				
Tax m				



	P1	P2	...	Pk
Tax1				
Tax2				
...				
Tax m				

ASV table transformed into embedding space

Labeled data

Sample associated data

Random Forest

IBD Status
0
1

m ~ 26,700 taxa
k ~ 100 properties

PCA
(k dim)

Normalize
(m dim)

- Untransformed data
- Embedded data
- PCA Reduced data
- No Dim. Red. data