

seq_to_gis	
C0 ->	[343206452, 343204350, 199999999]
C1 ->	[234523451, 343204354]
C2 ->	[]

One to many

All numbers are
fake
This document
just describes a
plan for
processing data.

gi_to_text (For all NCBI GIs)	
343206452 ->	"the deepest sea mud of the Mariana Trench"
343204354 ->	"isolated from terrestorial samples near ..."
199999999 ->	""
343204350 ->	"the deepest sea mud of the Mariana Trench"
... etc etc	

One to one

gi_to_matches (For only GIs found)	
343206452 ->	[ENV0:123123, ENV0:123777, ENV0:123777]
343204354 ->	[ENV0:23452345]
199999999 ->	[]
343204350 ->	[ENV0:123123, ENV0:123777, ENV0:123777]

One to many

gi_to_counts	
343204354 ->	(ENV0:00002007 -> 1, ENV0:00002041 -> 2)
199999999 ->	(ENV0:00001993 -> 1, ENV0:00003452 -> 1)
343204350 ->	(ENV0:00002007 -> 1, ENV0:00002041 -> 2)

One to one to many

seq_to_counts	
C0 ->	(ENV0:00002007 -> 0.667, ENV0:00002041 -> 0.333)
C1 ->	(ENV0:00002007 -> 0.123, ENV0:00002041 -> 0.821)
C2 ->	(ENV0:00002007 -> 0.667, ENV0:00002041 -> 0.333)

One to one to many

serial_to_concept
lorem

child_to_parents
lorem

concept_to_name
lorem

df_abundances
lorem