

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

gi_to_text (For all NCBI GIs)	One to one
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	

text_to_matches (Only for sources found)	One to many
"the deepest sea mud of the Mariana Trench" ->	[ENV0:123123, ENV0:123777, ENV0:123777]
"isolated from terrestorial samples near ..." ->	[ENV0:23452345]

text_to_counts	One to one to many
"the deepest sea mud of the Mariana Trench" ->	(ENV0:00002007 -> 1, ENV0:00002041 -> 2)
"isolated from terrestorial samples near ..." ->	(ENV0:00001993 -> 1, ENV0:00003452 -> 1)

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

serial_to_concept
lorem

child_to_parents
lorem

concept_to_name
lorem

df_abundances
lorem

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