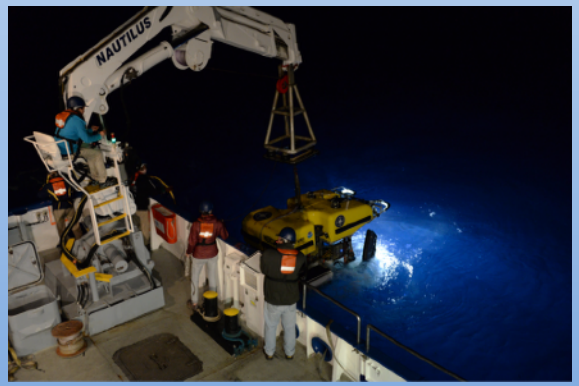
Fielding the field-to-fields pipeline:
From sampling through sequences in practice

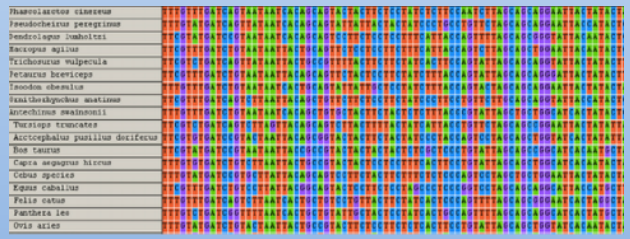
N. Dean Pentcheff & Regina Wetzer
Natural History Museum of Los Angeles County

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The digital collections landscape





NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

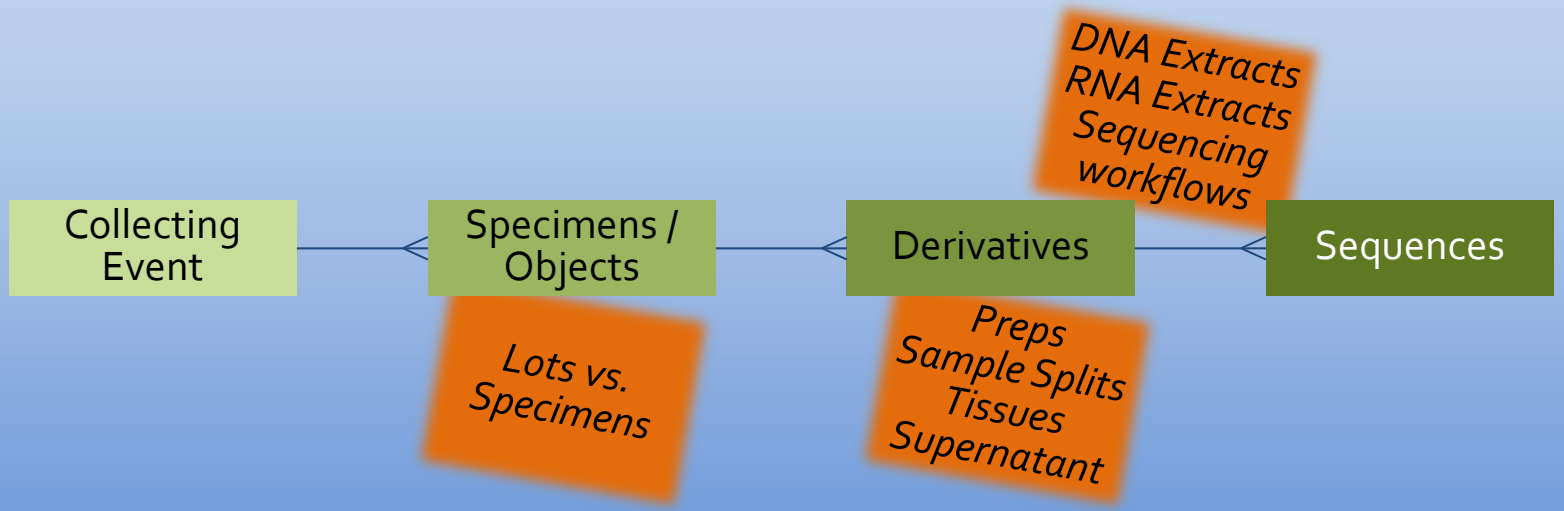
DISCO
Diversity Initiative for the Southern California Ocean

The data flow

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The data flow



Collectively how are we doing on this data landscape?

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The landscape in practice

It’s a bit sparse

The landscape in practice

Few institutions have the full pipeline

Photo of L.A. Water flowing above the Owens Valley.

Rob Reynolds

The two cultures

Collections/Specimens

Genetics/Sequences

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Collection –> Specimens –> Derivatives

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Collection –> Specimens –> Derivatives

Collection Data

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Collection –> Specimens –> Derivatives

Sequence management

Sequences <– Repository

Collection Data

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Sequence management

Collection Data

Sequence Data

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Sequence management

Collection Data

Sequence Data

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

The two cultures

Collections/Specimens

Genetics/Sequences

Collections databases

Sequence management

Collection Data

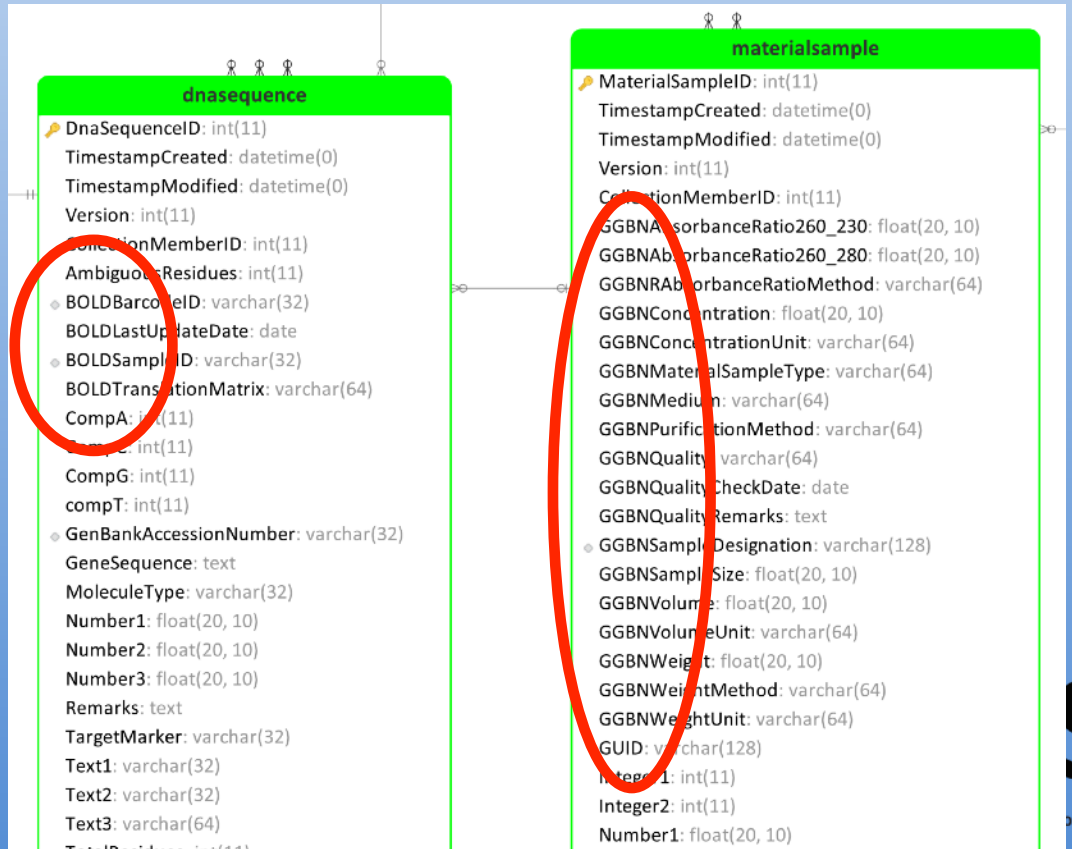
Sequence Data

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

This is why we have friction

Example:
Specify



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

SCO
or the Southern California Ocean

This is why we have friction

Example:
NHM
DISCO

Field Name	Type	Options / Comments (Click to toggle)
specimen ID	Number	Indexed, Auto-enter Serial, Always Validate, Required
parent	Number	Indexed, Always Validate, By Value List, Numeric Only
collection ID	Text	Indexed, Auto-enter Calculation replaces existing valu
taxon ID	Number	Indexed, By Value List, Message
determined by	Text	Indexed
determination date	Date	Indexed
determination reliable	Text	Indexed, Auto-enter Data, Always Validate, By Value L
determination remarks	Text	Indexed
determination timestamp	Timestamp	
determination history	Text	Indexed
genus species description	Text	Indexed
taxon description	Text	Indexed
molecule usefulness	Text	Indexed
physical location	Text	Indexed
specimen count description	Text	Indexed
specimen exists	Text	Indexed, Auto-enter Data, Always Validate, By Value L
BOLD voucher status	Text	Indexed, Auto-enter Data, Always Validate, By Value L
remarks	Text	Indexed
export marker	Text	Indexed
creation	Timestamp	Indexed, Creation Timestamp (Date and Time)
change	Timestamp	Indexed, Modification Timestamp (Date and Time), Ca
BOLD Project	Text	Indexed
BOLD Collection Code	Text	Indexed, Auto-enter Data
BOLD Institution Storing	Text	Indexed, Auto-enter Data
BOLD Identifier Institution	Text	Indexed, Auto-enter Data
BOLD Extra Info	Text	Indexed
BOLD Museum ID OVERRIDE	Text	Indexed
BOLD Museum ID	Calculation	Indexed, from Specimen Database, = If (IsEmpty (Tri
BOLD Process ID	Text	Indexed
specimen label cache	Text	Indexed
specimen label	Calculation	Unstored, from Specimen Database, = /* if no taxonic
islisted	Calculation	Unstored, from Specimen Database, = If (Taxonomy::

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

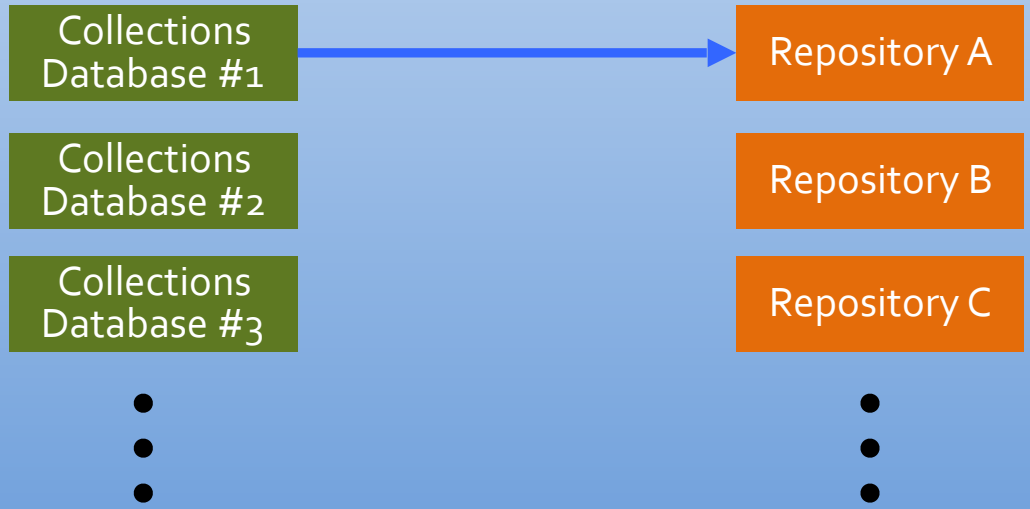
DISCO
Diversity Initiative for the Southern California Ocean

So we build adapters

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

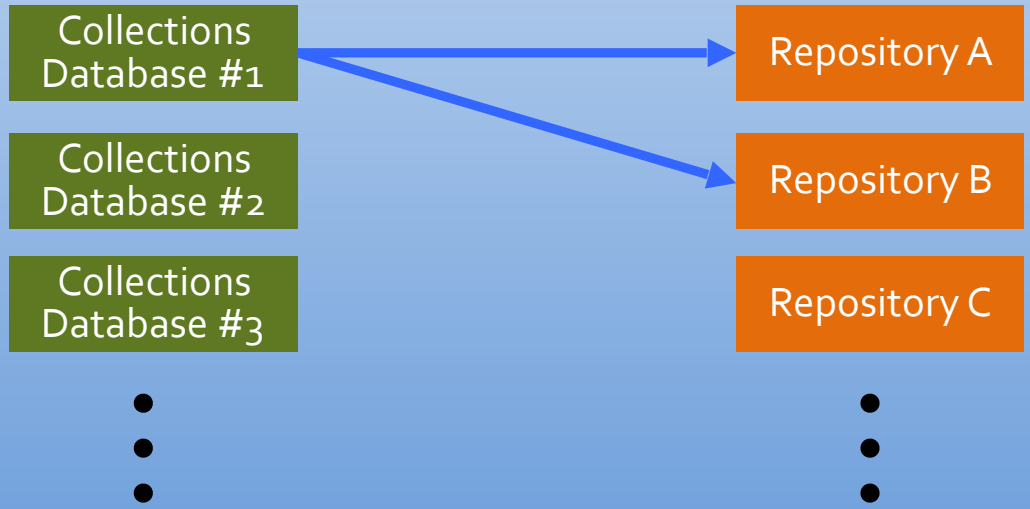
So we build adapters



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

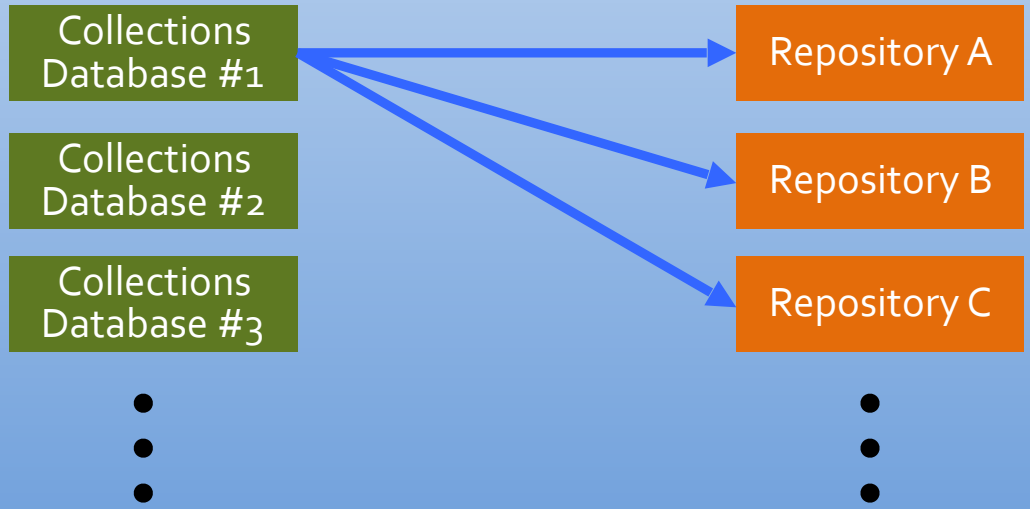
So we build adapters



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

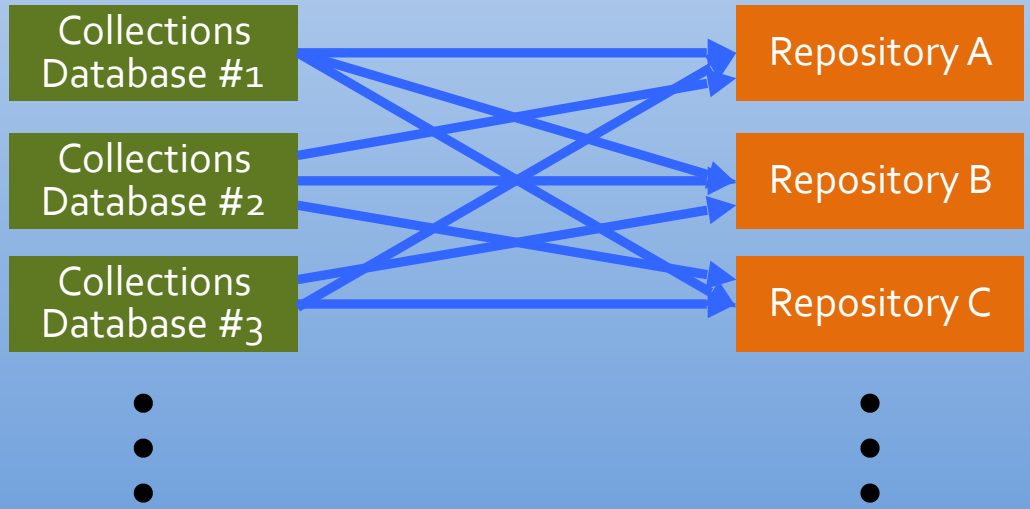
So we build adapters



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

So we build adapters



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

Adapters are horrid time sinks

Recast identifiers,
combine fields,
convert units,
fix filename limits,
massage taxonomy,
convert date format,
etc.

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

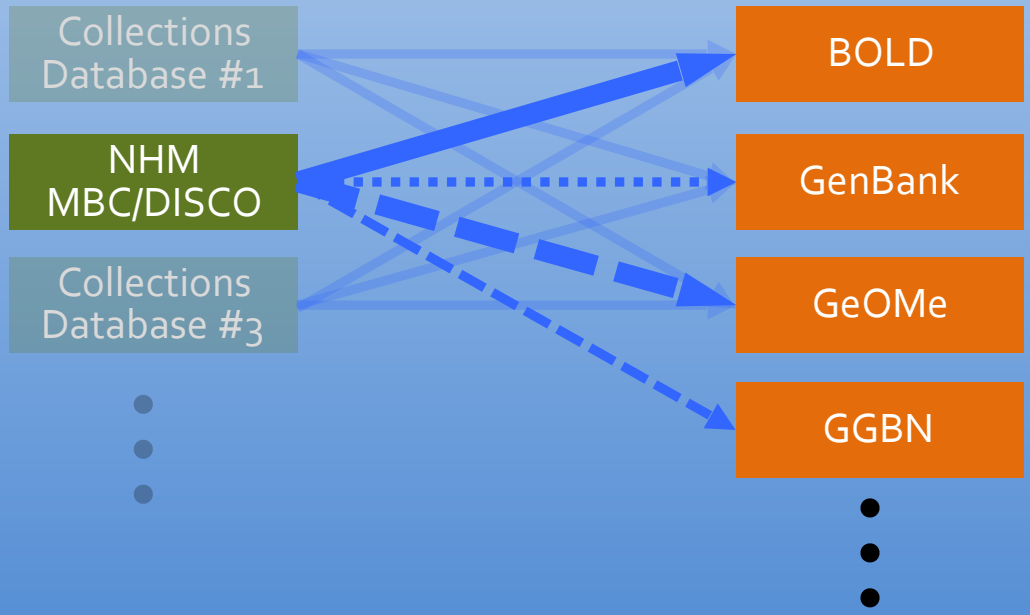
We do it under duress

In our case:

NATURAL HISTORY MUSEUM LOS ANGELES COUNTY

DISCO

Diversity Initiative for the Southern California Ocean



NATURAL HISTORY MUSEUM LOS ANGELES COUNTY

DISCO

Diversity Initiative for the Southern California Ocean

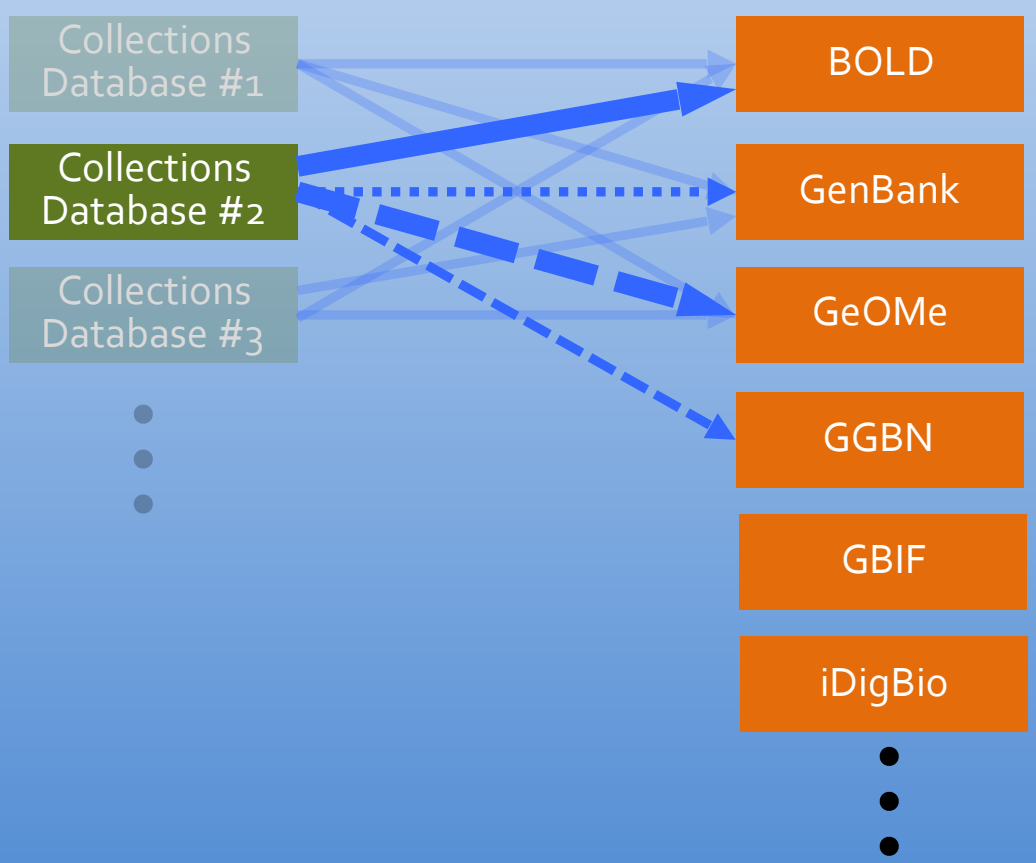
The landscape in practice

Few institutions have the full pipeline

Photo of L.A. Water flowing above the Owens Valley.

Rob Reynolds

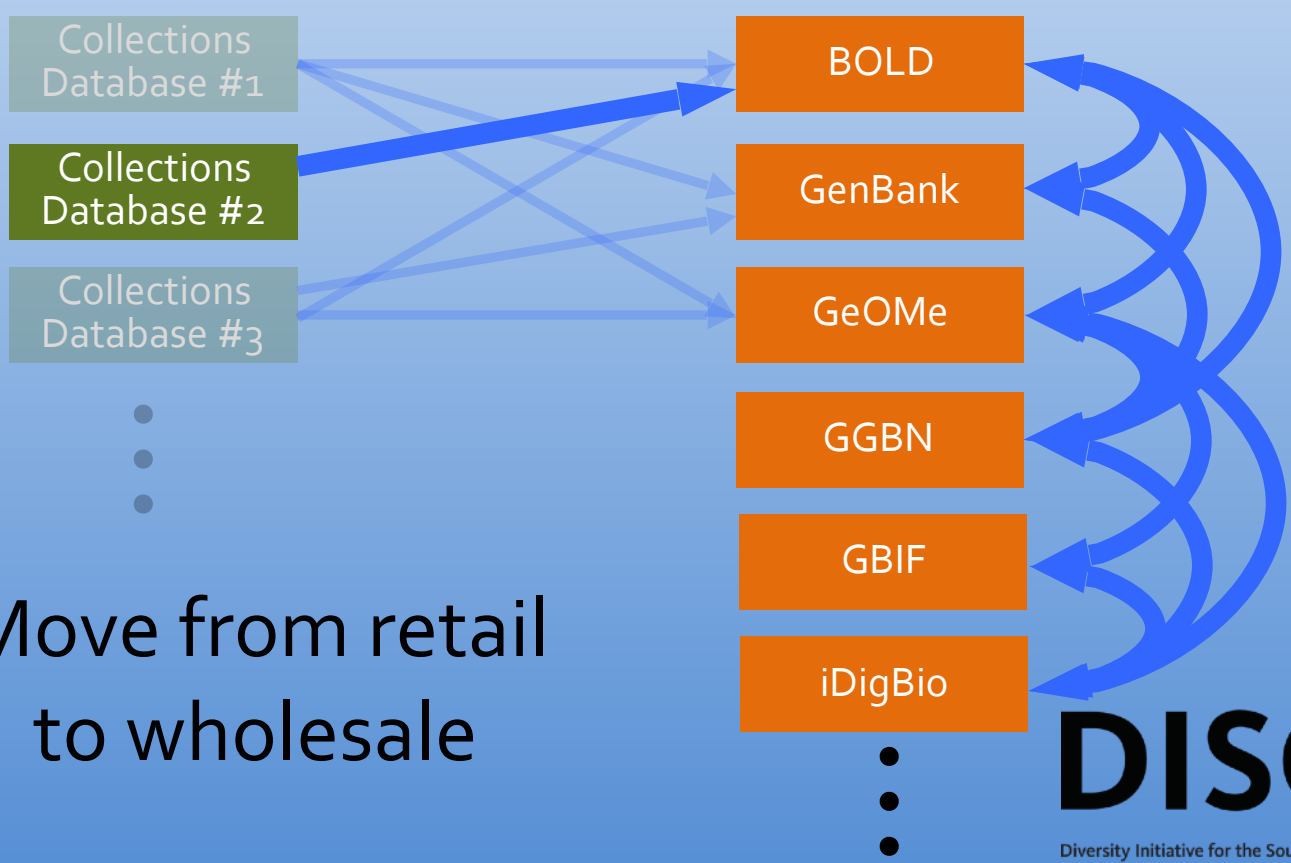
What is to be done?



NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

What is to be done?

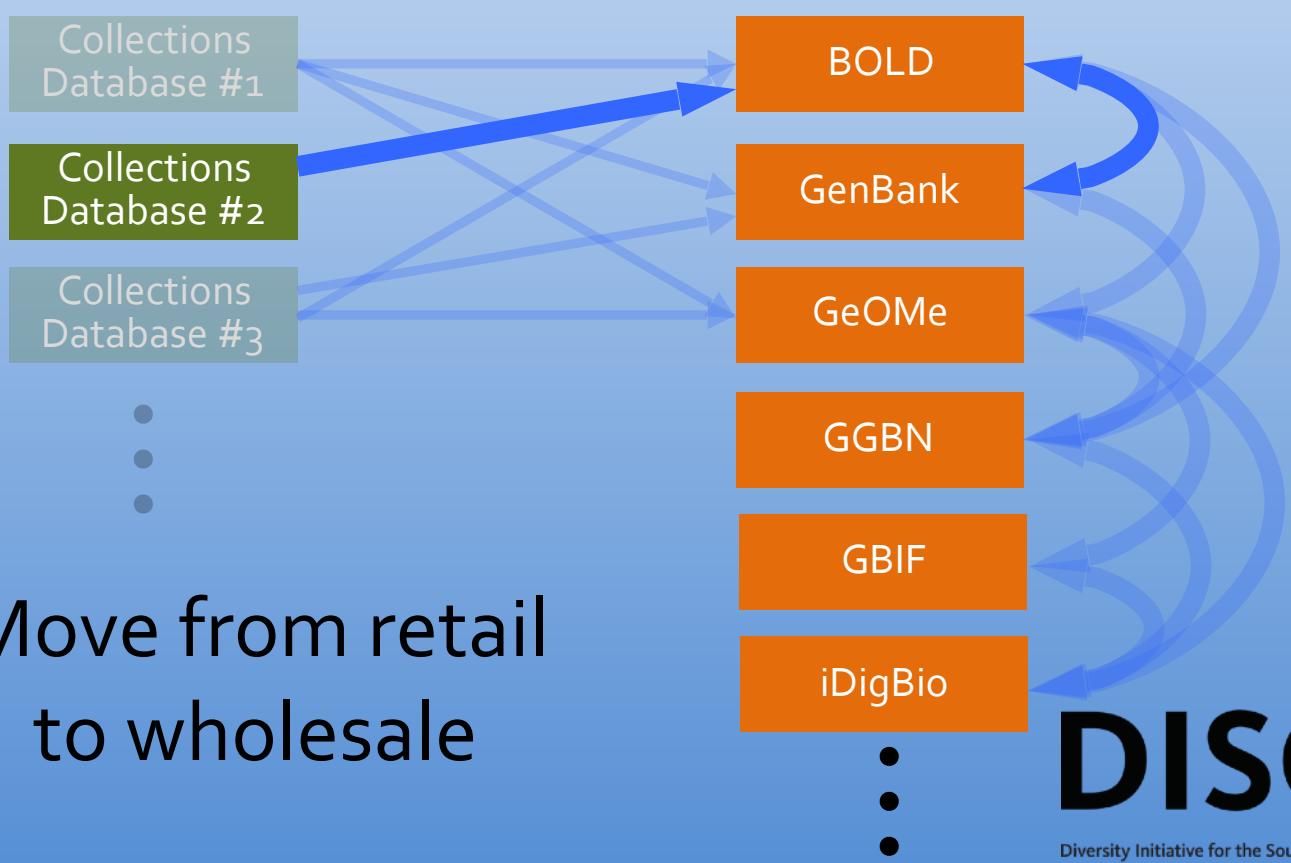


Move from retail to wholesale

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean

What is to be done?



Move from retail to wholesale

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

DISCO
Diversity Initiative for the Southern California Ocean


May our data flow as smoothly
as waves to the shore...
Thanks to:
GGBN
NMNH LAB
Dean Pentcheff pentcheff@nhm.org
Regina Wetzer wetzer@nhm.org
Diversity Initiative for the
Southern California Ocean