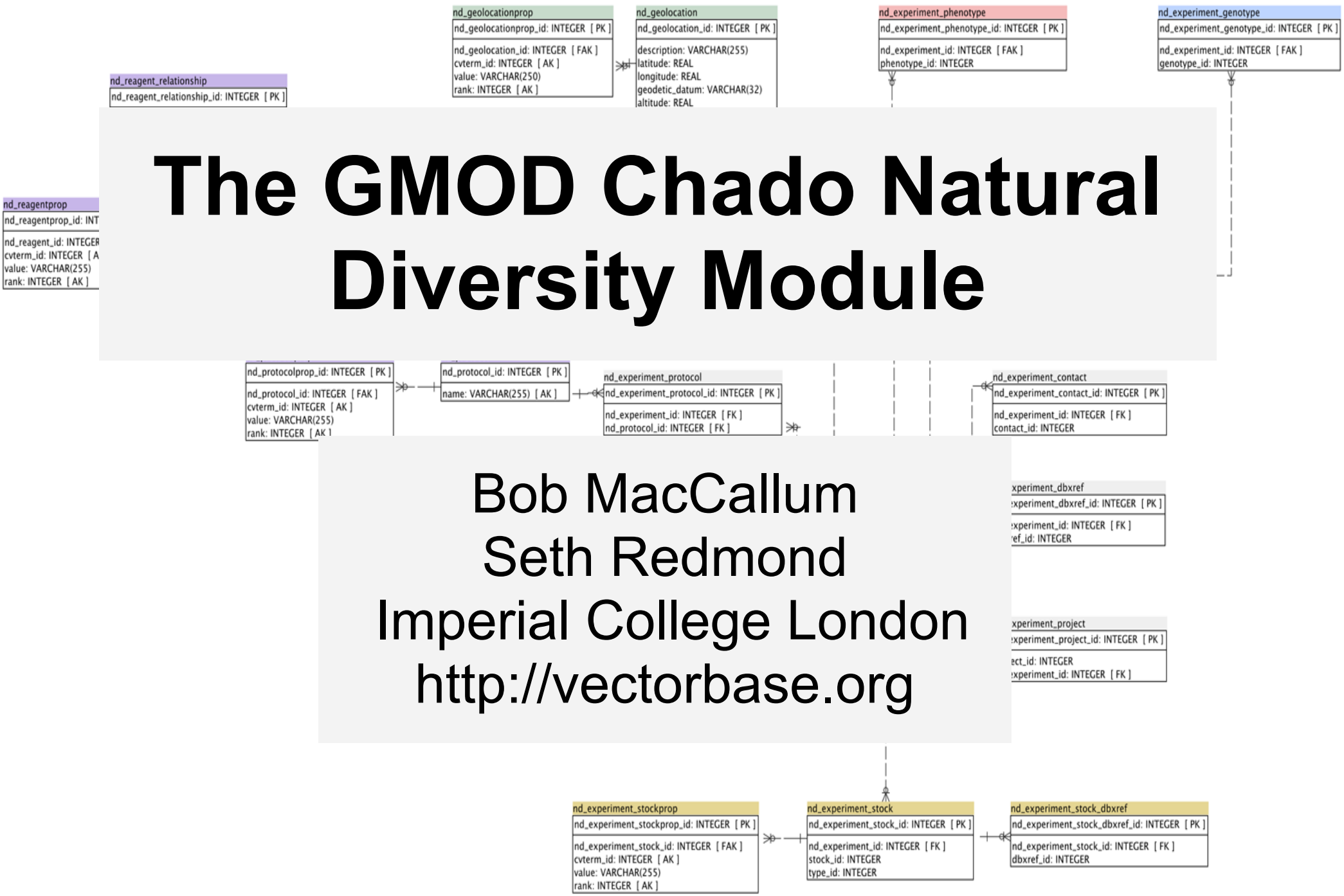



# The GMOD Chado Natural Diversity Module

Bob MacCallum
Seth Redmond
Imperial College London
http://vectorbase.org

# Motivation

Manage phenotypic and genotypic data for both field collected and captive bred organisms

Store collection site information for growing "next gen"-based variation data

Leverage existing/future Chado modules, GMOD tools and know-how

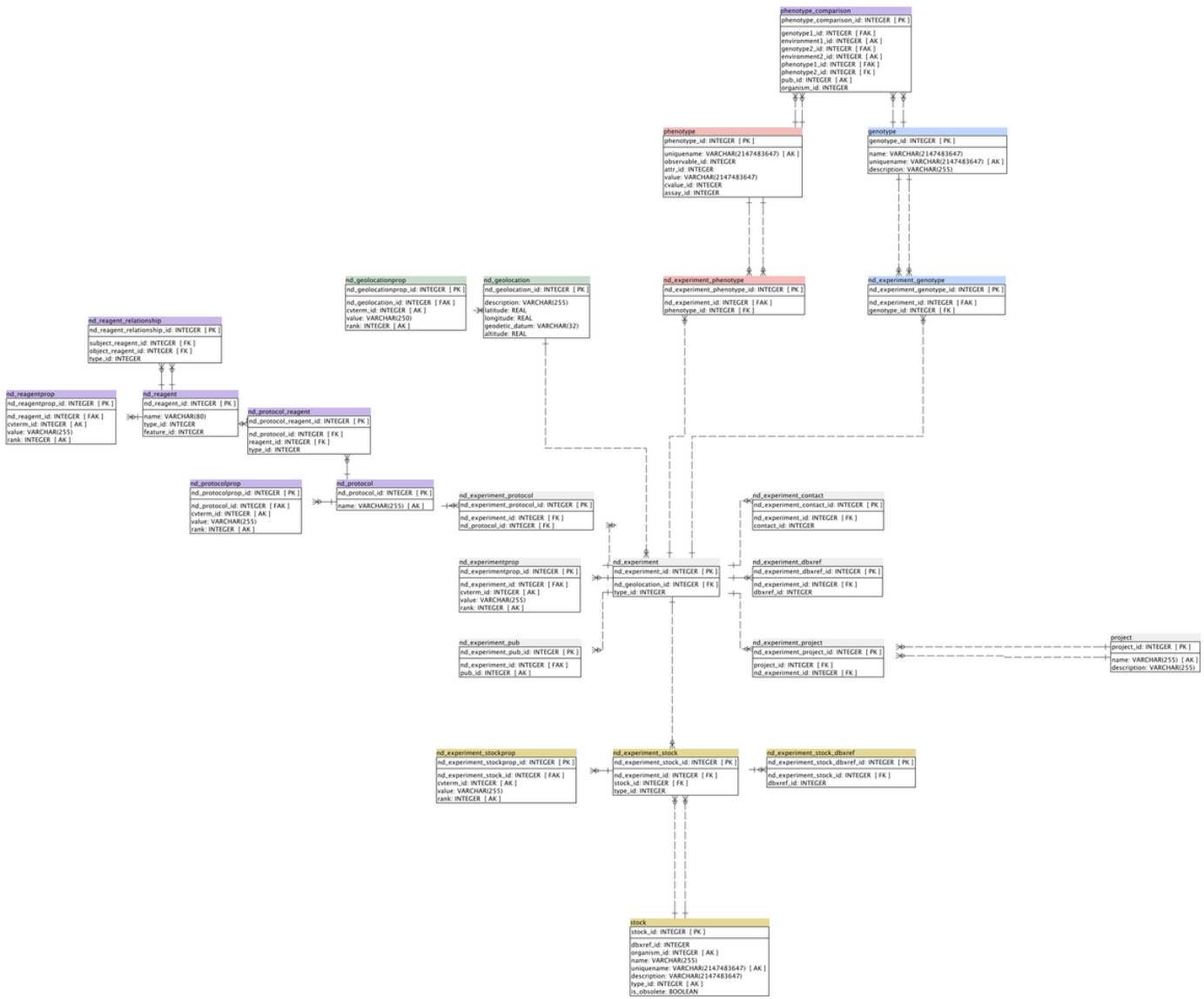
# Developmental history

| | |
|---|---|
| 2007 | Early version:<br>HeliconiusDB @ NESCent<br>(National Evolutionary Synthesis Center)<br>http://nescent.org<br><br>Inspired by:<br>GDPDM (The Genomic Diversity and Phenotype Data Model)<br>http://www.maizegenetics.net/gdpdm/  |
| 2009-2010 | Reincarnation spearheaded by:<br>Sook Jung @ Washington State University<br>GDR (Genome Database for Rosaceae)<br>http://www.rosaceae.org/  <br><br>GMOD working group formed:<br>http://gmod.org/wiki/Chado_Natural_Diversity_Module_Working_Group |
| August 2010 | Natural Diversity module merged into Chado svn trunk |

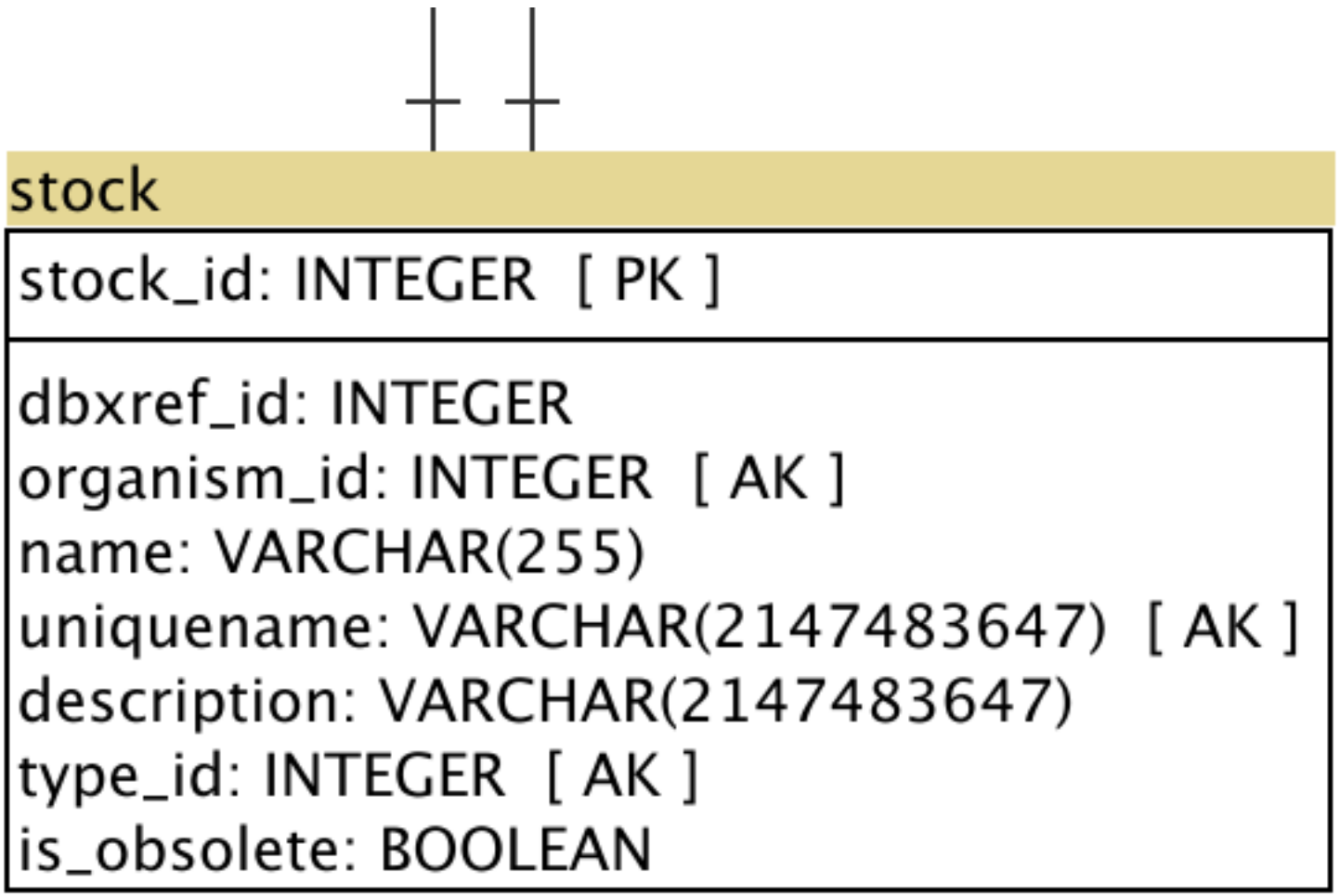
# Current developers

| Name | Affiliation |
|---|---|
| Dave Clements (organizer) | NESCent, GMOD |
| Sook Jung | Washington State University, GDR |
| Dorrie Main | Washington State University, GDR |
| Stephen Ficklin | CUGI |
| Meg Staton | CUGI |
| Scott Cain | OICR / GMOD |
| Genevieve DeClerk | Cornell / Gramene |
| Bob MacCallum | VectorBase, Imperial, London |
| Seth Redmond | VectorBase, Imperial, London |
| Naama Menda | Sol Genomics Network / SGN |
| Maren Friesen | Medicago ecological genomics, University of Southern California |
| Yuri Bendana | Medicago ecological genomics, University of Southern California |
| Pantelis Topalis | VectorBase, IMBB, Crete |
| Emmanuel Dialynas | VectorBase, IMBB, Crete |
| Rob Buels | Sol Genomics Network / SGN |

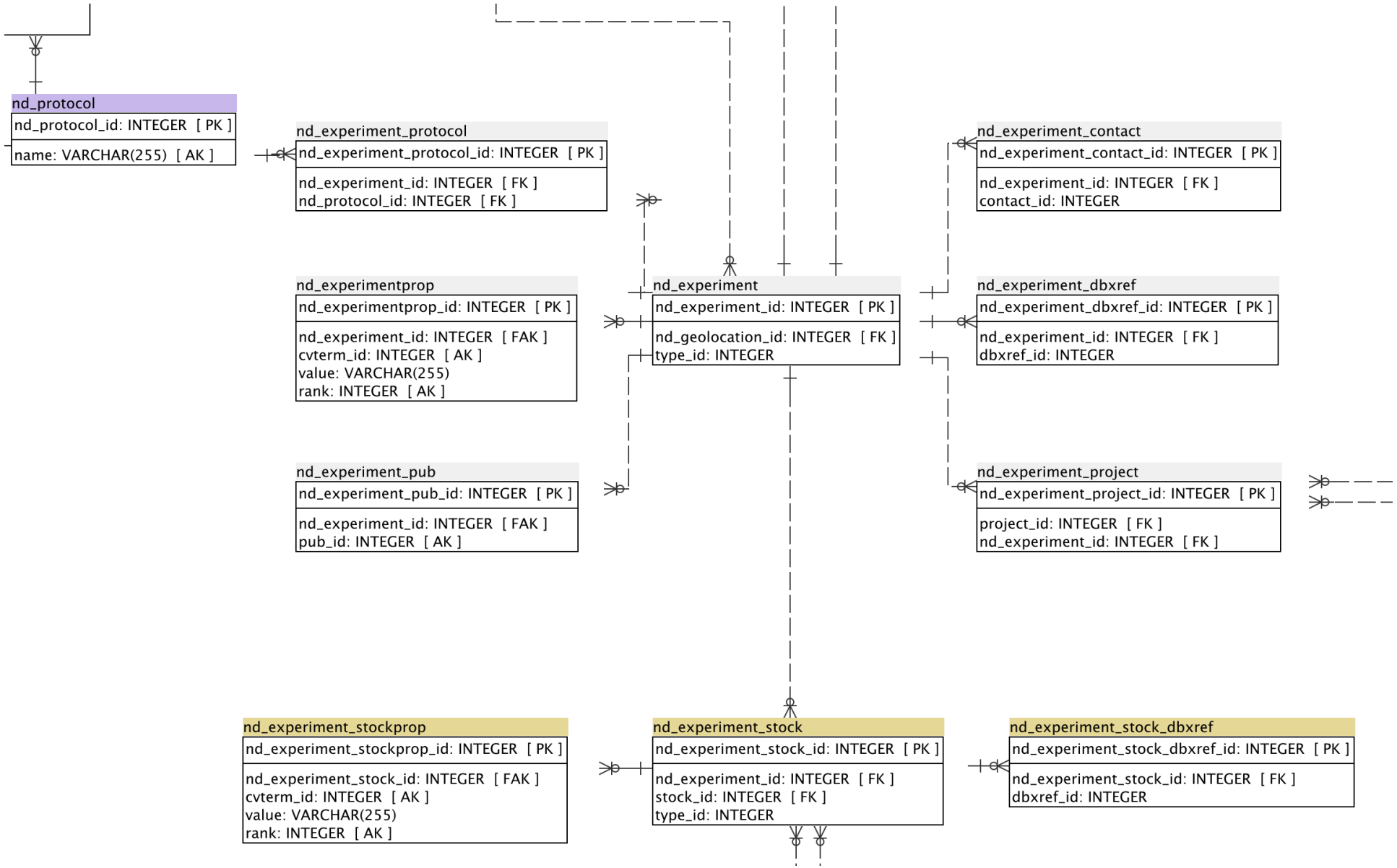
# The Schema


phenotype_comparison
phenotype_comparison_id: INTEGER [ PK ]
genotype1_id: INTEGER [ FAK ]
environment1_id: INTEGER [ AK ]
genotype2_id: INTEGER [ FAK ]
environment2_id: INTEGER [ AK ]
phenotype1_id: INTEGER [ FAK ]
phenotype2_id: INTEGER [ FK ]
pub_id: INTEGER [ AK ]
organism_id: INTEGER
phenotype
phenotype_id: INTEGER [ PK ]
uniquename: VARCHAR(2147483647) [ AK ]
observable_id: INTEGER
attr_id: INTEGER
value: VARCHAR(2147483647)
cvalue_id: INTEGER
assay_id: INTEGER
genotype
genotype_id: INTEGER [ PK ]
name: VARCHAR(2147483647)
uniquename: VARCHAR(2147483647) [ AK ]
description: VARCHAR(255)
nd_geolocationprop
nd_geolocationprop_id: INTEGER [ PK ]
nd_geolocation_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(250)
rank: INTEGER [ AK ]
nd_geolocation
nd_geolocation_id: INTEGER [ PK ]
description: VARCHAR(255)
latitude: REAL
longitude: REAL
geodetic_datum: VARCHAR(32)
altitude: REAL
nd_experiment_phenotype
nd_experiment_phenotype_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
phenotype_id: INTEGER [ FK ]
nd_experiment_genotype
nd_experiment_genotype_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
genotype_id: INTEGER [ FK ]
nd_reagent_relationship
nd_reagent_relationship_id: INTEGER [ PK ]
subject_reagent_id: INTEGER [ FK ]
object_reagent_id: INTEGER [ FK ]
type_id: INTEGER
nd_reagentprop
nd_reagentprop_id: INTEGER [ PK ]
nd_reagent_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_reagent
nd_reagent_id: INTEGER [ PK ]
name: VARCHAR(80)
type_id: INTEGER
feature_id: INTEGER
nd_protocol_reagent
nd_protocol_reagent_id: INTEGER [ PK ]
nd_protocol_id: INTEGER [ FK ]
reagent_id: INTEGER [ FK ]
type_id: INTEGER
nd_protocolprop
nd_protocolprop_id: INTEGER [ PK ]
nd_protocol_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_protocol
nd_protocol_id: INTEGER [ PK ]
name: VARCHAR(255) [ AK ]
nd_experiment_protocol
nd_experiment_protocol_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
nd_protocol_id: INTEGER [ FK ]
nd_experiment_contact
nd_experiment_contact_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
contact_id: INTEGER
nd_experimentprop
nd_experimentprop_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_experiment
nd_experiment_id: INTEGER [ PK ]
nd_geolocation_id: INTEGER [ FK ]
type_id: INTEGER
nd_experiment_dbxref
nd_experiment_dbxref_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
dbxref_id: INTEGER
nd_experiment_pub
nd_experiment_pub_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
pub_id: INTEGER [ AK ]
nd_experiment_project
nd_experiment_project_id: INTEGER [ PK ]
project_id: INTEGER [ FK ]
nd_experiment_id: INTEGER [ FK ]
project
project_id: INTEGER [ PK ]
name: VARCHAR(255) [ AK ]
description: VARCHAR(255)
nd_experiment_stockprop
nd_experiment_stockprop_id: INTEGER [ PK ]
nd_experiment_stock_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_experiment_stock
nd_experiment_stock_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
stock_id: INTEGER [ FK ]
type_id: INTEGER
nd_experiment_stock_dbxref
nd_experiment_stock_dbxref_id: INTEGER [ PK ]
nd_experiment_stock_id: INTEGER [ FK ]
dbxref_id: INTEGER
stock
stock_id: INTEGER [ PK ]
dbxref_id: INTEGER
organism_id: INTEGER [ AK ]
name: VARCHAR(255)
uniquename: VARCHAR(2147483647) [ AK ]
description: VARCHAR(2147483647)
type_id: INTEGER [ AK ]
is_obsolete: BOOLEAN

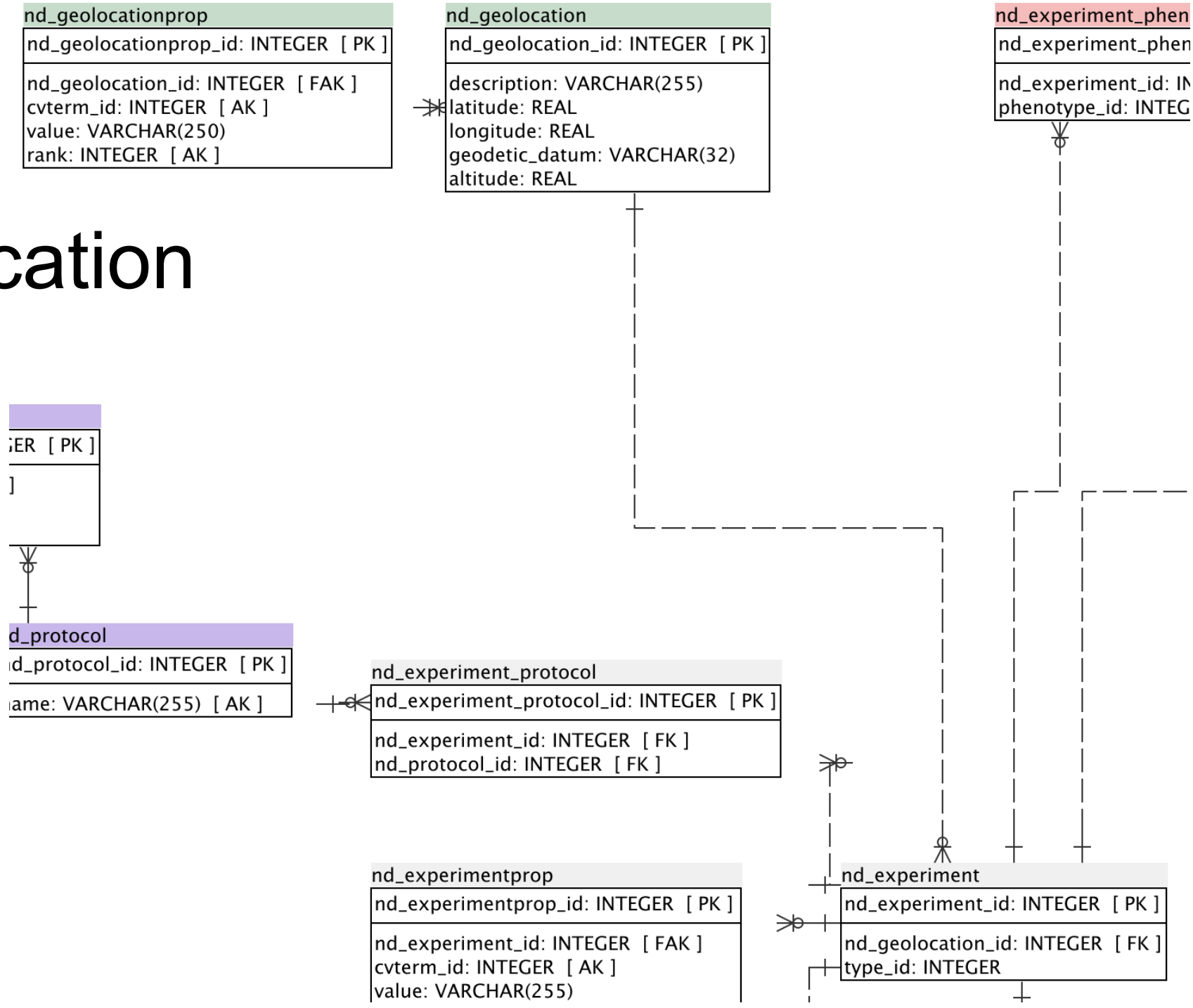
# Stock (pre-existing module)



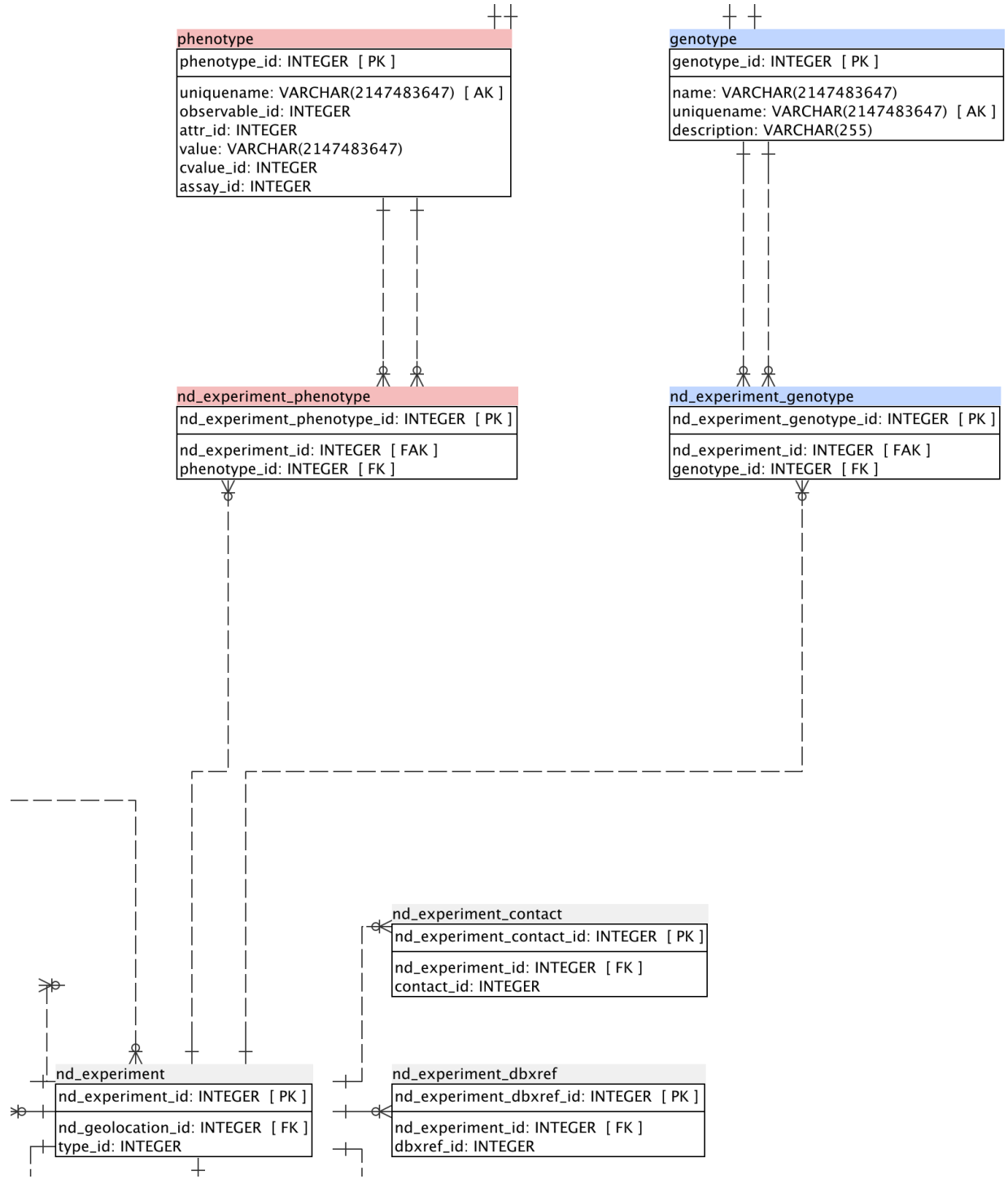
# Experiment


nd_protocol
nd_protocol_id: INTEGER [ PK ]
name: VARCHAR(255) [ AK ]
nd_experiment_protocol
nd_experiment_protocol_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
nd_protocol_id: INTEGER [ FK ]
nd_experiment_contact
nd_experiment_contact_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
contact_id: INTEGER
nd_experimentprop
nd_experimentprop_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_experiment
nd_experiment_id: INTEGER [ PK ]
nd_geolocation_id: INTEGER [ FK ]
type_id: INTEGER
nd_experiment_dbxref
nd_experiment_dbxref_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
dbxref_id: INTEGER
nd_experiment_pub
nd_experiment_pub_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FAK ]
pub_id: INTEGER [ AK ]
nd_experiment_project
nd_experiment_project_id: INTEGER [ PK ]
project_id: INTEGER [ FK ]
nd_experiment_id: INTEGER [ FK ]
nd_experiment_stockprop
nd_experiment_stockprop_id: INTEGER [ PK ]
nd_experiment_stock_id: INTEGER [ FAK ]
cvterm_id: INTEGER [ AK ]
value: VARCHAR(255)
rank: INTEGER [ AK ]
nd_experiment_stock
nd_experiment_stock_id: INTEGER [ PK ]
nd_experiment_id: INTEGER [ FK ]
stock_id: INTEGER [ FK ]
type_id: INTEGER
nd_experiment_stock_dbxref
nd_experiment_stock_dbxref_id: INTEGER [ PK ]
nd_experiment_stock_id: INTEGER [ FK ]
dbxref_id: INTEGER

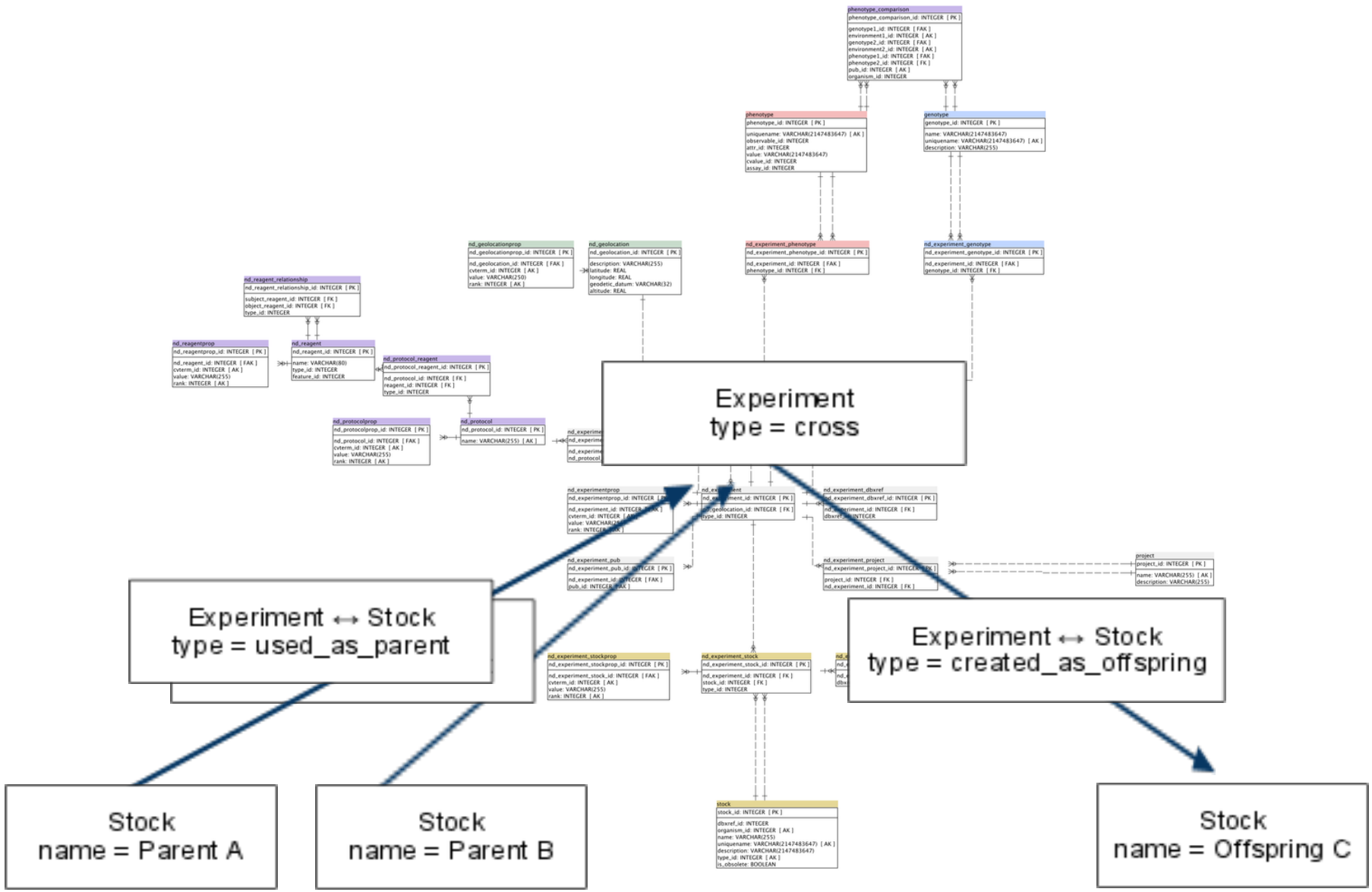
# Geolocation



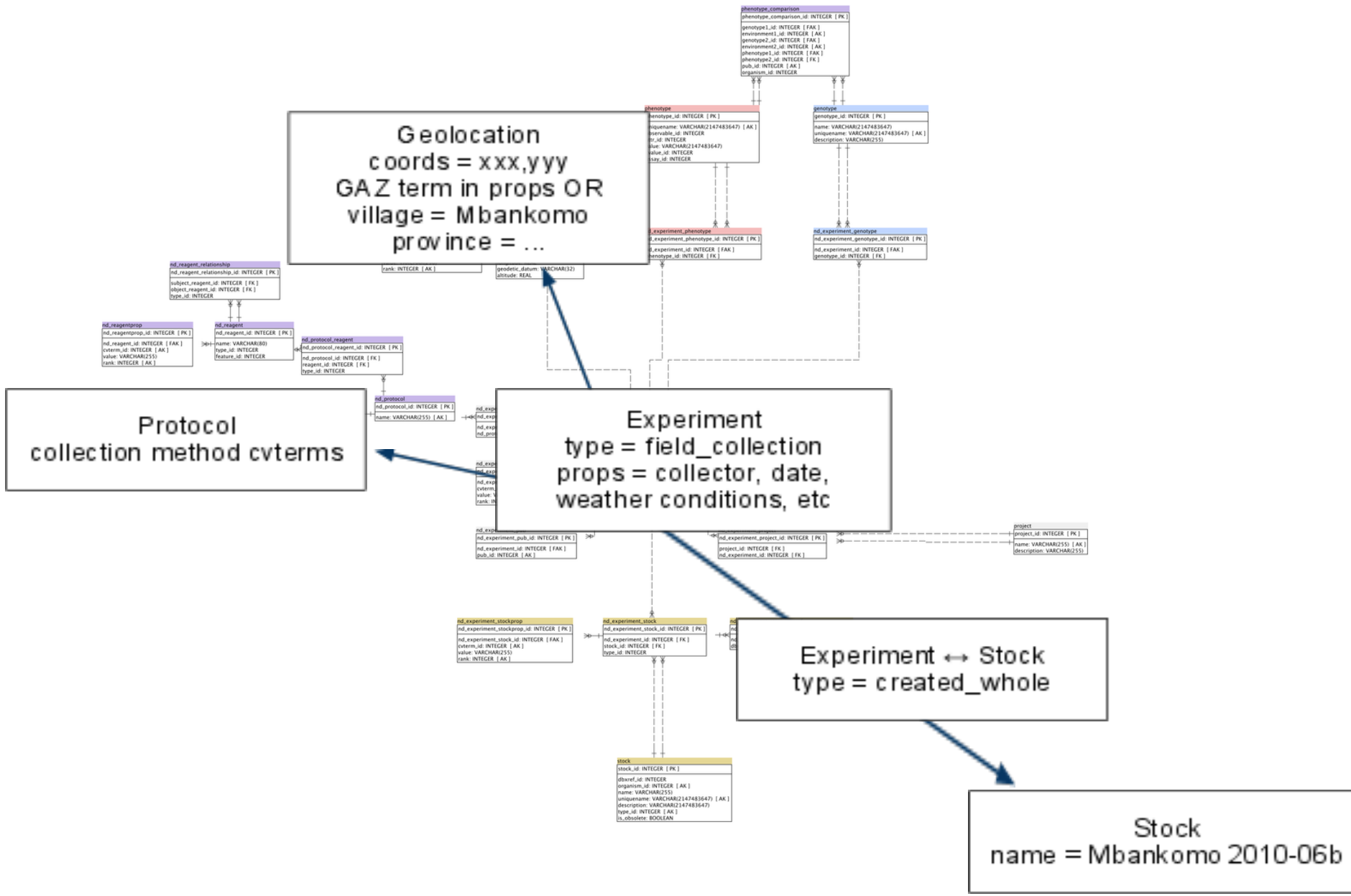
# Genotype & Phenotype



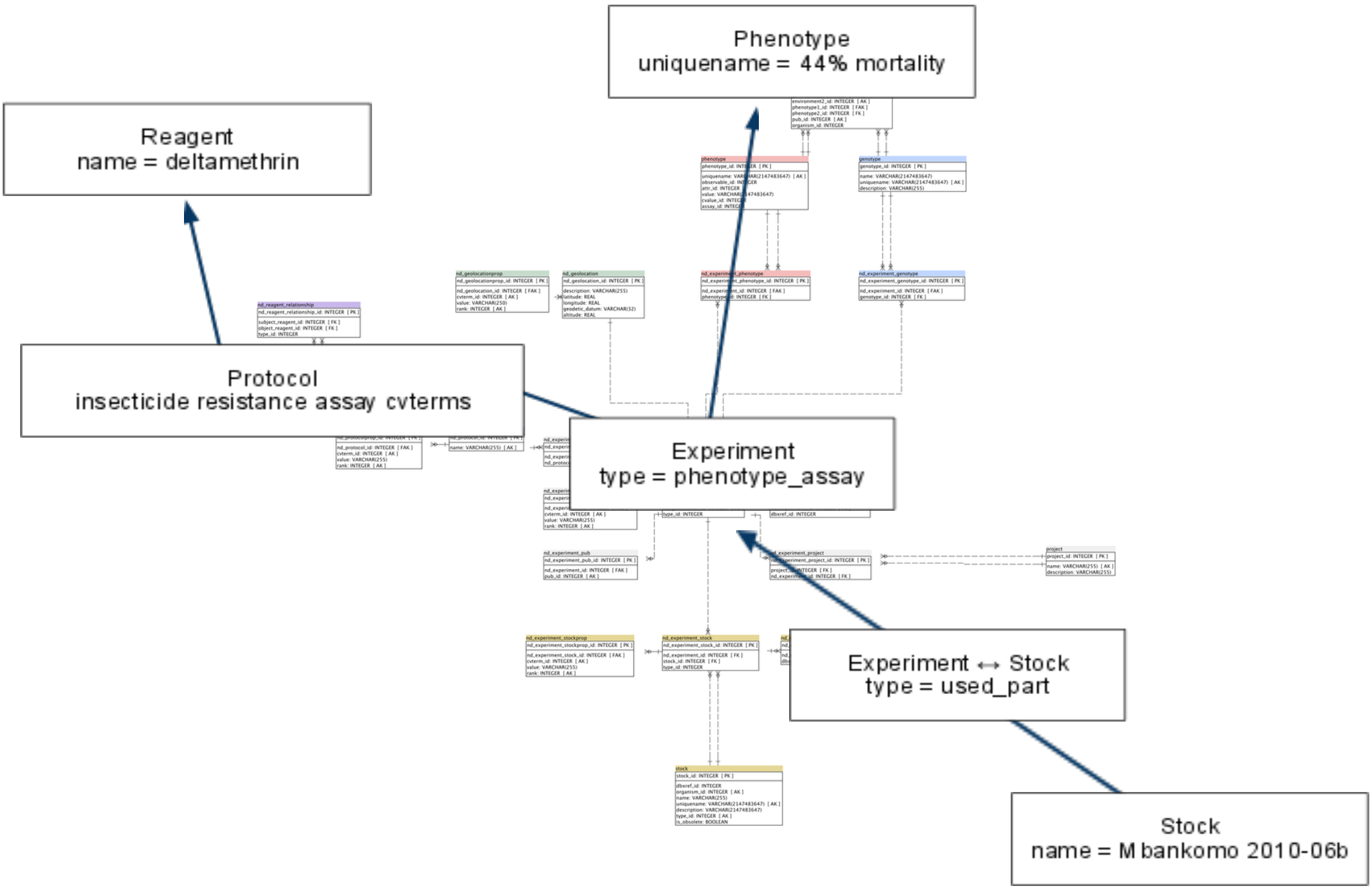
# Use case: cross experiment



# Use case: field collection



# Use case: phenotype assay



# CV terms

Schema is very flexible

nd_experiment.type and nd_experiment_stock.type are key

Several ways to do the same thing

Working group is hoping to agree on core CV terms to aid API development

# A simplified API