

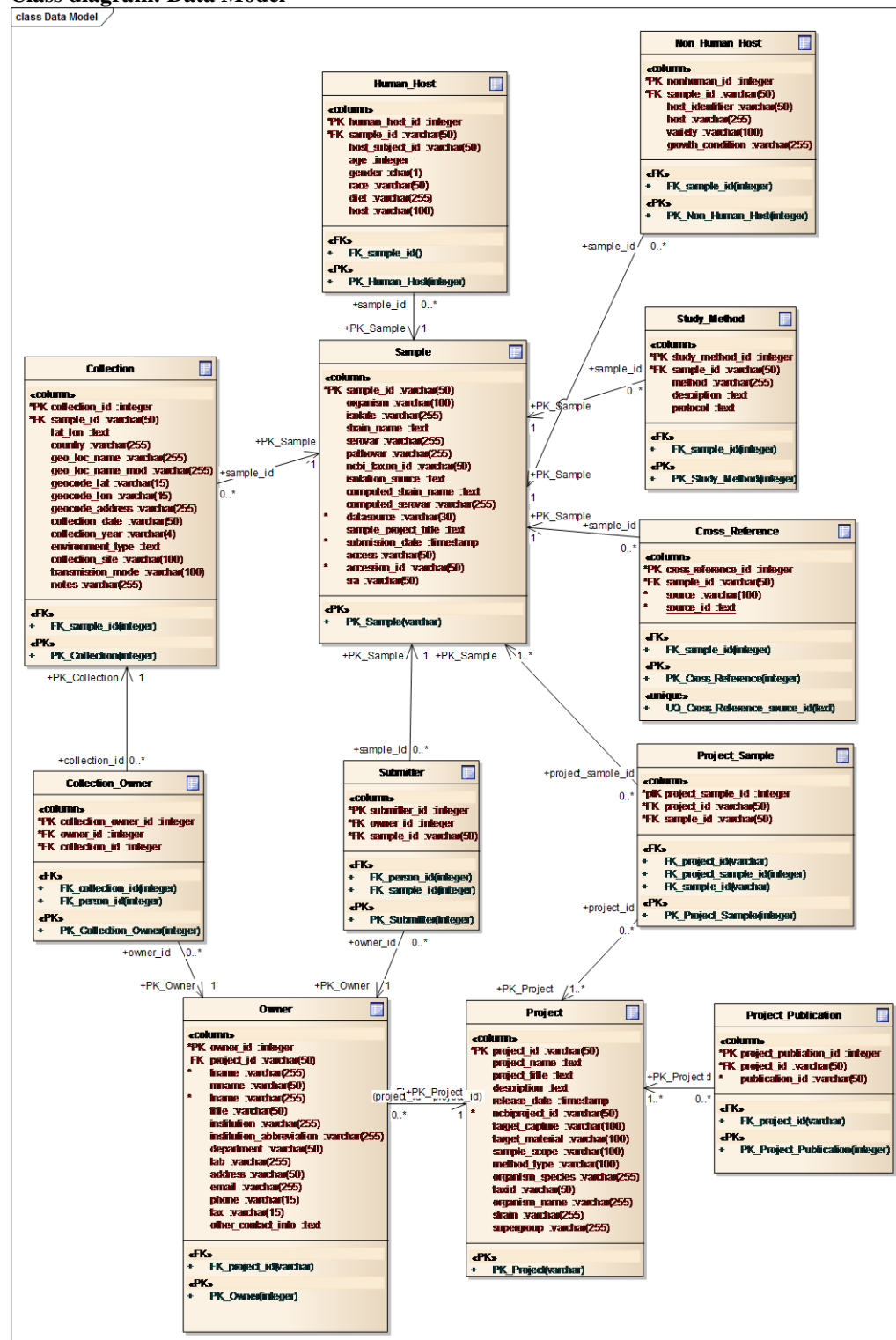
Table of Contents

Data Model.....	2
Collection	3
Collection_Owner.....	4
Cross_Reference	5
Human_Host.....	5
Non_Human_Host	6
Owner	7
Project.....	8
Project_Publication	10
Project_Sample.....	11
Sample.....	11
Study_Method.....	13
Submitter	14

BioAtt Database Documentation

Data Model

Class diagram: Data Model



Collection

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model
Detail: Created on 2/28/2012. Last modified on 9/20/2014.
Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	collection_id	integer	True	False					
False	sample_id	varchar	True	False	50				Foreign key from sample table
False	lat_lon	text	False	False					Value from BioSample <Attribute> with “lat_lon” as the “harmonized_name” attribute.
False	country	varchar	False	False	255				Placeholder for country
False	geo_loc_name	varchar	False	False	255				Value from BioSample <Attribute> with “geo_loc_name” as the “harmonized_name” attribute.
False	geo_loc_name_mod	varchar	False	False	255				If there is a mapping for the value contained in “geo_loc_name”, then use this as the modified name.
False	geocode_lat	varchar	False	False	15				A placeholder for the latitude after geocoding the geo_loc_name.
False	geocode_lon	varchar	False	False	15				A placeholder for the longitude after geocoding the geo_loc_name.
False	geocode_address	varchar	False	False	255				A placeholder for the actual address corresponding to the latitude/longitude found by the geocoder.
False	collection_date	varchar	False	False	50				Value from BioSample <Attribute> with “collection_date” as the “harmonized_name” attribute.
False	collection_year	varchar	False	False	4				Value extracted from the collection_date; expected to be the

									year the collection was made.
False	environment_type	text	False	False					Value from BioSample <Attribute> with “environment” as the “harmonized_name” attribute.
False	collection_site	varchar	False	False	100				Value from BioSample <Attribute> with “body_sites” as the “harmonized_name” attribute.
False	transmission_mode	varchar	False	False	100				unused
False	notes	varchar	False	False	255				other collection information (unused)

Constraints

Name	Type	Columns	Initial Code	Notes
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Collection	Public	collection_id		

Relationships

Columns	Association	Notes
	0..* Collection_Owner.collection_id 1 Collection.PK_Collection	
	0..* Collection.sample_id 1 Sample.PK_Sample	

Collection_Owner

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 3/1/2012. Last modified on 8/30/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	collection_owner_id	integer	True	False					Bridge between owner and collection tables
False	owner_id	integer	True	False					Foreign key from owner table
False	collection_id	integer	True	False					Foreign key from collection table

Constraints

Name	Type	Columns	Initial Code	Notes
FK_collection_id	Public	collection_id		Foreign key constraint
FK_person_id	Public	owner_id		Foreign key constraint
PK_Collection_Owner	Public	collection_owner_id		

Name	Type	Columns	Initial Code	Notes

Relationships

Columns	Association	Notes
	0..* Collection_Owner.collection_id 1 Collection.PK_Collection	
	0..* Collection_Owner.owner_id 1 Owner.PK_Owner	

Cross_Reference

Database: PostgreSQL, *Stereotype*: «table», *Package*: Data Model

Detail: Created on 3/6/2012. Last modified on 3/21/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	cross_reference_id	integer	True	False					
False	sample_id	varchar	True	False	50				Foreign key from sample table
False	source	varchar	True	False	100				"db" attribute from the <Id> tag.
False	source_id	text	True	True					Value from the same <Id> tag.

Constraints

Name	Type	Columns	Initial Code	Notes
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Cross_Reference	Public	cross_reference_id		
UQ_Cross_Reference_source_id	Public	source_id		

Relationships

Columns	Association	Notes
	0..* Cross_Reference.sample_id 1 Sample.PK_Sample	

Human_Host

Database: PostgreSQL, *Stereotype*: «table», *Package*: Data Model

Detail: Created on 3/1/2012. Last modified on 10/28/2014.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	human_host_id	integer	True	False					
False	sample_id	varchar	True	False	50				Foreign key from sample table
False	host_subject_id	varchar	False	False	50				Value from BioSample <Attribute> with "host_subject_id" as the

									"harmonized_name" attribute.
False	age	integer	False	False					Placeholder for age
False	gender	char	False	False	1				Value from BioSample <Attribute> with "sex" or "host_sex" as the "harmonized_name" attribute. Either M(ale) or F(emale) or U(nknown)
False	race	varchar	False	False	50				Value from BioSample <Attribute> with "race" as the "harmonized_name" attribute.
False	diet	varchar	False	False	255				Value from BioSample <Attribute> with "diet" as the "harmonized_name" attribute.
False	host	varchar	False	False	100				Value from BioSample <Attribute> with "host" as the "harmonized_name" attribute.

Constraints

Name	Type	Columns	Initial Code	Notes
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Human_Host	Public	human_host_id		

Relationships

Columns	Association	Notes
	0..* Human_Host.sample_id 1 Sample.PK_Sample	

Non_Human_Host

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 3/6/2012. Last modified on 10/28/2014.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	nonhuman_id	integer	True	False					
False	sample_id	varchar	True	False	50				Foreign key from the sample table
False	host_identifier	varchar	False	False	50				Value from BioSample

									<Attribute> with “host_taxid” as the “harmonized_name” attribute.
False	host	varchar	False	False	255				Value from BioSample <Attribute> with “host” as the “harmonized_name” attribute.
False	variety	varchar	False	False	100				Value from BioSample <Attribute> with “variety” as the “harmonized_name” attribute.
False	growth_condition	varchar	False	False	255				Value from BioSample <Attribute> with “growth_condition” as the “harmonized_name” attribute.

Constraints

Name	Type	Columns	Initial Code	Notes
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Non_Human_Host	Public	nonhuman_id		

Relationships

Columns	Association	Notes
	0..* Non_Human_Host.sample_id 1 Sample.PK_Sample	

Owner

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 2/28/2012. Last modified on 9/19/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	owner_id	integer	True	False					
False	project_id	varchar	False	False	50				Foreign key from the project table
False	fname	varchar	True	False	255				Value from BioSample <Contact><Name><Last> tag.
False	mname	varchar	False	False	50				Placeholder for middle name
False	lname	varchar	True	False	255				Value from BioSample <Contact><Name><

False									First> tag.
False	title	varchar	False	False	50				Placeholder for contact's title
False	institution	varchar	False	False	255				Value from BioSample <Owner><Name> tag.
False	institution_abbreviation	varchar	False	False	255				"abbr" attribute from BioSample <Owner><Name> tag.
False	department	varchar	False	False	50				Placeholder for department
False	lab	varchar	False	False	255				Placeholder for lab
False	address	varchar	False	False	50				Placeholder for address
False	city	varchar	False	False	50				Placeholder for city
False	state	varchar	False	False	50				Placeholder for state
False	postal_code	varchar	False	False	50				Placeholder for postal code
False	country	varchar	False	False	100				Placeholder for country
False	email	varchar	False	False	255				"email" attribute from BioSample <Contact><Name> tag
False	phone	varchar	False	False	15				Placeholder for phone
False	fax	varchar	False	False	15				Placeholder for fax
False	other_contact_info	text	False	False					Value from the first BioSample <Provider> tag.

Constraints

Name	Type	Columns	Initial Code	Notes
FK_project_id	Public	project_id		Foreign key constraint
PK_Owner	Public	owner_id		

Relationships

Columns	Association	Notes
	0..* Submitter.owner_id 1 Owner.PK_Owner	
	0..* Collection_Owner.owner_id 1 Owner.PK_Owner	
(project_id = project_id)	0..* Owner.FK_project_id 1 Project.PK_Project	

Project

Database: PostgreSQL, *Stereotype*: «table», *Package*: Data Model

Detail: Created on 9/19/2012. Last modified on 9/7/2014.
Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	project_id	varchar	True	False	50				"id" attribute from BioProject <ArchiveID> tag prefixed with "BioProject".
False	project_name	text	False	False					Value from BioSample <ProjectDescr><Name> tag
False	project_title	text	False	False					Value from BioSample <ProjectDescr><Title> tag
False	description	text	False	False					Value from BioSample <ProjectDescr><description> tag
False	release_date	timestamp	False	False					Value from BioSample <ProjectDescr><ProjectReleaseDate> tag
False	ncbiproject_id	varchar	True	False	50				"id" attribute from BioProject <ArchiveID> tag
False	target_capture	varchar	False	False	100				"capture" attribute from BioProject <ProjectType><Target> tag
False	target_material	varchar	False	False	100				"material" attribute from BioProject <ProjectType><Target> tag
False	sample_scope	varchar	False	False	100				"sample_scope" attribute from BioProject <ProjectType><Target> tag
False	method_type	varchar	False	False	100				"method_type" attribute from BioProject <ProjectType><Method> tag
False	organism_species	varchar	False	False	255				"species" attribute from BioProject <ProjectType><Organism> tag
False	taxid	varchar	False	False	50				"taxid" attribute from BioProject <ProjectType><Organism> tag

False	organism_name	varchar	False	False	255				Value from BioProject <ProjectType><Organism><OrganismName> tag
False	strain	varchar	False	False	255				Value from BioProject <ProjectType><Organism><Strain> tag
False	supergroup	varchar	False	False	255				Value from BioProject <ProjectType><Organism><Supergroup> tag

Constraints

Name	Type	Columns	Initial Code	Notes
PK_Project	Public	project_id		

Relationships

Columns	Association	Notes
	0..* Project_Sample.project_id 1..* Project.PK_Project	
(project_id = project_id)	0..* Owner.FK_project_id 1 Project.PK_Project	
	0..* Project_Publication.project_id 1..* Project.PK_Project	

Project_Publication

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 10/24/2012. Last modified on 10/24/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	project_publication_id	integer	True	False					
False	project_id	varchar	True	False	50				"id" attribute from BioProject <ArchiveID> tag prefixed with "BioProject".
False	publication_id	varchar	True	False	50				"publication_id" attribute from BioProject <ProjectDescr><Publication> tags

Constraints

Name	Type	Columns	Initial Code	Notes
FK_project_id	Public	project_id		Foreign key constraint
PK_Project_Publication	Public	project_publication_id		

Name	Type	Columns	Initial Code	Notes

Relationships

Columns	Association	Notes
	0..* Project_Publication.project_id 1..* Project.PK_Project	

Project_Sample

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 9/19/2012. Last modified on 10/15/2014.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	project_sample_id	integer	True	False					Bridge between project and sample tables
False	project_id	varchar	True	False	50				Foreign key from project table
False	sample_id	varchar	True	False	50				Foreign key from sample table

Constraints

Name	Type	Columns	Initial Code	Notes
FK_project_id	Public	project_id		Foreign key constraint
FK_project_sample_id	Public	project_sample_id		Foreign key constraint
PK_Project_Sample	Public	project_sample_id		
FK_sample_id	Public	sample_id		Foreign key constraint

Relationships

Columns	Association	Notes
	0..* Project_Sample.project_id 1..* Project.PK_Project	
	0..* Project_Sample.project_sample_id 1..* Sample.PK_Sample	

Sample

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 2/28/2012. Last modified on 10/28/2014.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	sample_id	varchar	True	False	50				
False	organism	varchar	False	False	100				"taxonomy_name" attribute from the BioSample <Organism> tag
False	isolate	varchar	False	False	255				Value from

									BioSample <Attribute> with “isolate” as the “harmonized_name” attribute.
False	strain_name	text	False	False					Value from BioSample <Attribute> with “strain” as the “harmonized_name” attribute.
False	serovar	varchar	False	False	255				Value from BioSample <Attribute> with “serovar” as the “harmonized_name” attribute.
False	pathovar	varchar	False	False	255				Value from BioSample <Attribute> with “pathovar” as the “harmonized_name” attribute.
False	ncbi_taxon_id	varchar	False	False	50				“taxonomy_id” attribute from the BioSample <Organism> tag
False	isolation_source	text	False	False					Value from BioSample <Attribute> with “isolation_source” as the “harmonized_name” attribute.
False	computed_strain_name	text	False	False					If strain_name is found, then this is the same. Otherwise use regular expression “.*?(?:str.\s)(.*)” to extract from “taxonomy_name”
False	computed_serovar	varchar	False	False	255				If serovar is found, then this is the same. Otherwise use regular expression “.*?(?:serovar\s)((?: ?!\\sstr).*)” to extract from “taxonomy_name”
False	datasource	varchar	True	False	30				Always set to “BioSample” currently.
False	sample_project_title	text	False	False					Value from BioSample <Title>

									tag
False	submission_date	timestamp	True	False					"submission_date" attribute from BioSample root
False	access	varchar	False	False	50				"access" attribute from BioSample root
False	accession_id	varchar	True	False	50				"id" attribute from BioSample root
False	sra	varchar	False	False	50				Value of BioSample <Id> tag with "db" attribute equal to "SRA"

Constraints

Name	Type	Columns	Initial Code	Notes
PK_Sample	Public	sample_id		

Relationships

Columns	Association	Notes
	0..* Cross_Reference.sample_id 1 Sample.PK_Sample	
	0..* Collection.sample_id 1 Sample.PK_Sample	
	0..* Non_Human_Host.sample_id 1 Sample.PK_Sample	
	0..* Human_Host.sample_id 1 Sample.PK_Sample	
	0..* Study_Method.sample_id 1 Sample.PK_Sample	
	0..* Submitter.sample_id 1 Sample.PK_Sample	
	0..* Project_Sample.project_sample_id 1..* Sample.PK_Sample	

Study_Method

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 3/7/2012. Last modified on 3/21/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	study_method_id	integer	True	False					
False	sample_id	varchar	True	False	50				Foreign key from sample table
False	method	varchar	False	False	255				Placeholder for method
False	description	text	False	False					Placeholder for description
False	protocol	text	False	False					Values from all the BioSample <Description><Paragraph> tags appended together.

Constraints

Name	Type	Columns	Initial Code	Notes
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Study_Method	Public	study_method_id		

Relationships

Columns	Association	Notes
	0..* Study_Method.sample_id 1 Sample.PK_Sample	

Submitter

Database: PostgreSQL, *Stereotype:* «table», *Package:* Data Model

Detail: Created on 2/28/2012. Last modified on 5/3/2012.

Notes:

Columns

PK	Name	Type	Not Null	Unique	Len	Prec	Scale	Init	Notes
True	submitter_id	integer	True	False					Bridge between owner and sample tables
False	owner_id	integer	True	False					Foreign key from owner table.
False	sample_id	varchar	True	False	50				Foreign key into sample table extracted from "id" attribute of BioSample root and prefixed with "BioSample"

Constraints

Name	Type	Columns	Initial Code	Notes
FK_person_id	Public	owner_id		Foreign key constraint
FK_sample_id	Public	sample_id		Foreign key constraint
PK_Submitter	Public	submitter_id		

Relationships

Columns	Association	Notes
	0..* Submitter.owner_id 1 Owner.PK_Owner	
	0..* Submitter.sample_id 1 Sample.PK_Sample	