

## VegBank - Bug #786

### Update the Plots datamodel to allow for individual trees

11/13/2002 01:53 PM - Michael Lee

<b>Status:</b>	Resolved	<b>Start date:</b>	11/13/2002
<b>Priority:</b>	Immediate	<b>Due date:</b>	
<b>Assignee:</b>	Michael Lee	<b>% Done:</b>	0%
<b>Category:</b>	misc	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	UnMilestoned	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	786		
<b>Description</b>			
Discussion with Boyle & Enquist led to realization that we can improve the data model for individual tree stems to better meet the needs of the stem-map community. This can wait until early 2003. However, in as much as trees are not being loaded now, would make sense to do this before we really load trees. Probable solution is separate table for stems attached to TaxonObservation, but for performance may need separate table. -- DECIDE THIS ISSUE!			
<b>Related issues:</b>			
Blocked by VegBank - Bug #788: Load tree-related data		<b>Resolved</b>	<b>11/13/2002</b>
Blocked by VegBank - Bug #1169: Upgrade VegBranch to new 1.0.2 Vegbank data m...		<b>Resolved</b>	<b>10/02/2003</b>
Blocked by VegBank - Bug #870: New XML export and import into the VegBank system		<b>Resolved</b>	<b>11/13/2002</b>

#### History

##### #1 - 11/27/2002 01:40 PM - Michael Lee

The current solution will be for us to add one field to TaxonObservation

taxonObsType varchar(30) [nulls are OK]

acceptable values are (currently):

"collective" -- taxonObs applies to a collection of individuals

"individual" -- taxonObs applies to an individual

"collective" should be the default value unless 'individual' is specified

##### #2 - 01/08/2003 10:34 PM - John Harris

An example dataset would be good.

##### #3 - 01/15/2003 10:15 AM - Michael Lee

This db is scarcely populated. Data that concern us are in TaxonObs, StemCount, and StemLocation. To download, you'll get download.cgi or something, then you'll need to change the extension to .zip and then extract the db therein.

##### #4 - 01/15/2003 10:56 AM - Robert Peet

Various users need individual tree attributes to be contained within the database. They cite the case of each tree having a voucher museum collection (or even multiple vouchers).

Because of the potential existence of multiple vouchers, we agreed to add a table for vouchers.

The essential problem with the current data model is lack of taxonomic determination of individual trees. Two primary solutions were discussed.

(1) Create a table parallel to Taxon\_Observation for individuals, called perhaps Individual\_Observation. The problem here is that we would need a second Individual\_ Interpretation table, or we would need to place a switch in Taxon\_ Interpretation so that a record could point to a record in either of the two tables. Similarly, the Voucher table would need a switch so that a record could apply to either the Taxon\_Observation or

## Individual\_Observation

(2) Add an Observation\_Type to Taxon\_Observation which indicates whether this is an individual stem or a collective record. If a collective observation, then current fields like cover could be populated, and if an individual observation then a record could be placed in a linked stems table. This solution has a performance price in that there will be millions of added entries in the table to handle that many individual stems.

We recognized that solution #2 is the best for use in a conceptual data model, while recognizing that the actual physical model implemented may need to be closer to #1 for performance reasons.

I think we agreed to start with solution #2, but we could revisit this. I am anything but sure about it being the better choice.

Here is the basic data model (replacing TaxonObs, StemCount and StemLocation for solution #2.

```
TABLE: TaxonObservation) ---- A slight revision of hte table
TaxonObservation_ID = PK
Observation_ID = FK to the plot
BiologicalType (Collective or Individual)
Cover
BasalArea
Density
Notes
```

```
TABLE: Stem (replaces current stemCount and stemLocation tables)
Stem_ID = PK
TaxonObservation_ID = FK
Stem_ID = Recursive FK for repeated observations
  (Each set would need a new Observation entry and thus
  A new Taxon_Observation_ID record)
StemDiameter
StemDiameterAccuracy
StemXPosition
StemYPosition
StemCode
```

```
TABLE: Voucher
Voucher_ID = PK
TaxonObservation_ID = FK
Party_ID = collector
CollectionNumber
Museum
AccessionNumber
CollectionDate
```

As soon as we agree that we want solution 2 rather than 1, you could build this into VegBank and then populate it with some example data.

### #5 - 07/07/2003 04:20 PM - Michael Lee

3 options, as outlined <http://bugzilla.ecoinformatics.org/attachment.cgi?id=86&action=view> are, to leave things the same and use business rules to deal with this issue. Option B is to add one field to differentiate between records that are dealing with all of a taxon on a plot, and those records that only deal with part of the whole taxon. Option C is to make all the changes requested by Bob. I lean towards A or B, with a slight preference for B so long as adding the field is trivial for Gabe. I think C is unnecessary right now.

Vouchers can be dealt with separately, since they would be completely new to the system. We could also use vouchers in a way similar to notes. Someone could then specify tableName and fieldName and PK# to add a voucher for anything in the database: plant, soil, etc.

### #6 - 07/12/2003 03:11 PM - Robert Peet

Are not tree stems largely new to the system in that we have never tried to load any in Vegbank, and I rather doubt any code has been written for them aside from the table construction.

I vote to do this once and do it correctly. Am open to discussion as to when

we do it.

I remain very open minded as to whether stems should be a child of taxonObservation or of a new table called individualObservation. I am currently leaning more toward having the parallel individualObservation. If we go this way we need to decide whether taxonInterpretation applies to both individualObservation and taxonObservation or if we need a new individualInterpretation. Makes no difference to me. The first is perhaps more elegant, but there would be less to reengineer if we constructed a new parallel stemInterpretation.

**#7 - 07/21/2003 10:01 AM - Michael Lee**

Given Bob's comment of "Let's do this once and do it right" it seems to me best to leave the model as it is for now and implement any additional changes in functionality by adding values to closed lists and/or business rules. This keeps Gabe happy, too, as model changes are cumbersome to implement with a populated database.

**#8 - 07/21/2003 10:19 AM - Michael Lee**

A clarification of my last comment:

Revising the structure of a populated database is somewhat difficult, even if the tables being changed are not the ones populated. We are attempting to ensure that everything flows from one source, an XML document describing our database. So the process to change the db structure would be to change that document, drop the current database, then create the new one. Obviously, before we drop the initial database, we have to backup the data and then restore the data to the new database, directly where possible. Where changes have been made to the structure, this becomes more difficult, as the backed up data may have to be fiddled with to get it back into the new database.

[the alternative would be to "hack" the new tables in manually at a SQL prompt, but this introduces the possibility that the implemented database aspects (java, sql, model, vegbranch) could disagree if there was a typo in the manually entered SQL. This possibility is greatly reduced if all things are generated from the same source.]

But the general idea with datamodel changes **should be** to add functionality not subtract it. So the "old data" may have no information about the new fields we might want to add, but they should stay completely intact. I think this is true for our stems model change, as all fields are kept, just the tables are being combined into only one table and maybe a few new fields will be added.

Furthermore, another reason to wait is that Gabe has no time left. There are more important things for him to deal with now and this is one of the several things that will need to be left for the next person. Especially since we don't seem to be convinced of how to deal with the taxonInterpretation w.r.t. stems specifically (ie 2 tables: individualInterpretation and taxonInterpretation or what?). This certainly needs resolving if we are to "fix it once and fix it right"

**#9 - 10/02/2003 08:34 AM - Michael Lee**

we should fix this as we are fixing the irregular taxa model change, which must be done by Nov. 1. This one requires adding fields to taxonInterpretation only. Must be mirrored in VegBranch, which is bug 1169

**#10 - 10/13/2003 04:02 PM - Michael Lee**

This change has been made and no java code should sit on this, save perhaps only importing via XML. At current time, we can't make a taxonInterpretation of an extant taxonObs. When this is made, we will have to allow for this to be attached to either TaxonObs or StemLocation.

**#11 - 01/31/2005 02:54 PM - Michael Lee**

changed from components that are to be deleted to "misc" so that bugs don't get deleted with component. Sorry for all the email.

**#12 - 03/27/2013 02:15 PM - Redmine Admin**

Original Bugzilla ID was 786

---

**Files**

