

Beschreibung der Tabellenstruktur für eine Sammeltabelle für den Upload in die FOR816 Datenbank

Stand: 2010-06-07
Autor: Thomas Lotz

Im folgenden wird die Struktur beschrieben, in der nach dem neuen „Sammeltabellen-Format“, Excel oder OpenOfficeCalculator Dateien für den Upload in die Datenbank vorbereitet werden soll.

Unter Sammeltabellen verstehen wir folgendes:

- An verschiedenen räumlichen Standorten (Subplots oder Messpunkte) wurde zu verschiedenen Sammel-/ Aufnahmeobjekten ein einzelner Parameter aufgenommen, wie z.B.:
 - die „Häufigkeit verschiedener Tierarten“ gezählt.
 - der „prozentuale Bedeckungsgrad unterschiedlicher Pflanzenarten“ geschätzt.
 - der „Wert für die Menge diverser membrangebundener Fettsäuren im Boden“ gemessen.

Annahmen:

- An jedem Standort wurde zu den Objekten (Tierarten oder Pflanzenarten oder Fettsäuren...) die gleiche Information zum gleichen Zeitpunkt aufgenommen.
- Für alle Standorte gelten die selben Metainformationen.
- In der Datenbank wird nur ein (Listen basiertes) dataset kreiert, dass nur eine ID bekommt.
- Es wird in einer zukünftigen Version des FOR816 Data Warehouses möglich sein, das dataset wieder in Form einer Tabelle herunter zu laden.

Im Prinzip ist eine Sammeltabelle, die als ein Dataset in der Datenbank gespeichert werden soll, aus zwei Tabellen (Matrizen) zusammen gesetzt:

1. Sekundärmatrix

- enthält plotbezogene Informationen wie z.B. Plotbeschreibung, Exposition, Bewuchsdichte mit Kräutern etc. (das sind NICHT die Metainformationen für den gesamten, in der Datenbank zusätzlich zu speichernden, Datensatz -> vgl. allgemeine Uploadbeschreibung)

2. object x sample Matrix

- enthält die Datenwert: Standorte als Spalten und Informationen zu den Objekten in den Zeilen

Diese zwei Matrizen müssen vom Uploader in die im Folgenden beschriebene Form gebracht werden, sodass sie anschließend in dieser Struktur als eine csv-Datei an den Uploadprozess übergeben werden können.

Wenn der Datensatz mehr als 256 Spalten (Plots) enthält, kann das Erstellen dieser Datei in einem Tabellenkalkulationsprogramm (Excel vor Version XP oder OO-Calc) schwierig werden. In diesem Fall kann man sich mehrere Dateien erstellen und sie in einem geeigneten Texteditor zusammenfügen. Jedes Statistikprogramm und jede Datenbankanwendung wird eine Möglichkeit bieten, tabellarische Daten in eine geeignete Textdatei auszugeben. Generell können beliebig viele Spalten in die Datenbank hochgeladen werden.

	A	B	C	D	E	F	G
1	<i>PlotName</i>		SPLIT_COL	PlotA	PlotB	PlotC	
2	<i>PlotMeta1</i>			NW	N	O	
3	<i>PlotMeta2</i>			14	30	20	
4	<i>PlotMeta3</i>			open	closed	mixed	
5	<i>datetime</i>			2010-03-16	2010-03-16	2010-03-16	
6	<i>geocoordinate_latitude_min</i>			-3.97160	-3.97500	-4.00061	
7	<i>geocoordinate_latitude_max</i>			-3.97160	-3.97300	-4.00052	
8	<i>geocoordinate_longitude_min</i>			-79.07920	-79.07950	-79.07930	
9	<i>geocoordinate_longitude_max</i>			-79.07920	-79.07840	-79.07920	
10	<i>geocoordinate_amsl</i>			3850	NULL	3010	
11	SPLIT_ROW						
12	<i>Species</i>	<i>Genus</i>		<i>Spec_N</i>	<i>Spec_N</i>	<i>Spec_N</i>	
13	Spec1	GenA		2	NULL	4	
14	Spec2	GenA		0	5	2	
15	Spec3	GenC		0	NULL	NULL	
16	Spec4	GenB		12	33	NULL	
17							
18							

<i>Attribute (in Attributelist defined)</i>
<i>Attribute (required)</i>
KEYWORD
Value

Abb. 1: Collection_Table_Upload_Template

Beschreibung der Elemente der Sammeltabelle-Tabelle (vgl. Abb.1):

- Es gibt eine Spalte (SPLIT_COL) und eine Zeile (SPLIT_ROW), die als Trennfelder in der csv-Datei unbedingt vorhanden sein müssen.
- Diese Trennfelder teilen die Tabelle in **vier Quadranten**, wovon jeweils zwei zu einer der oben genannten Matrizen gehören
 - **links oben** --> Sekundärmatrix
 - erste Spalte: Attributnamen der plotspezifischen Informationen
 - alle weiteren Spalten: leer
 - **rechts oben** --> Sekundärmatrix
 - Werte der plotspezifischen Informationen (in der Maßeinheit des Attributs)
 - jede Spalte enthält einen Standort (hat individuelle Koordinaten)
 - **links unten** --> object x sample Matrix
 - erste Zeile: Attributnamen der zeilenbeschreibenden Objekte und ggf. in weiteren Spalten Attribute zur weiteren Klassifikationen der Objekte
 - alle weiteren Zeilen: Werte (Bezeichnung, Name) der zeilenbeschreibenden Objekte
 - **rechts unten** --> object x sample Matrix
 - in der ersten Zeile steht der Attributname der Information (Parameter/ Meßwert), der für die einzelnen Objekte im Quadranten links unten aufgenommen wurde (hier muss für alle Spalten immer der selbe attributeName verwendet werden)
 - alle Felder ab der zweiten Zeile enthalten die Werte

Alle Attribute, egal in welchem Quadranten sie stehen, müssen in der Attributliste definiert sein, die auf der Webseite im Uploadbereich veröffentlicht ist. Der Attributname muss dabei exakt dem „attributeName“ in der Attributliste entsprechen.

Es ist zu beachten, dass es für den Quadranten links oben sechs Pflichtattribute (rot markiert) gibt.

Zu diesen Attributen müssen für jeden Standort Werte eingegeben werden.

Wertefelder von anderen Attributen dürfen auch leer sein („NULL“, s.u.).

Die **sechs Pflichtattribute** beschreiben zu jedem Wert, wann und wo er aufgenommen wurde.

Damit werden die Informationen in dem Primärdatensatz auch für andere Nutzer wertvoll, da eine Suche an Hand von Raum- und Zeitkriterien ermöglicht wird. Wie schon erwähnt wird bei diesem Uploadformat davon ausgegangen, dass alle Daten an einem Standort zum selben Zeitpunkt erhoben wurden.

- **datetime:**

Zeitpunkt, an dem die Daten aufgenommen wurden. Dieser kann theoretisch in einigen Fällen nicht auf einen Tag beschränkt werden, da über Wochen hinweg gesammelt wurde, aber der Autor muss sich hier für einen repräsentativen Tag entscheiden.

Mögliche Formate: YYYY-MM-DD
 YYYY-MM-DD hh:mm
 YYYY-MM-DD hh:mm:ss

- **geocoordinate_:**

Vier geographische Koordinaten beschreiben die „BoundingBox“ des Plots (_latitdue_min: südliche Begrenzung, _latitdue_max: nördliche Begrenzung, _longitdue_min: westliche Begrenzung, _longitdue_max: östliche Begrenzung).

Ist der Standort nur ein **räumlicher Punkt** und keine Fläche, so gilt

_latitude_min = _latitude_max und _longitude_min = _longitude_max

Es wird dann zweimal der gleiche Koordinatenwert eingetragen.

Das Format der geographischen Koordinate ist in Dezimalgrad nach WGS84 anzugeben mit maximal fünf Nachkommastellen ($\pm xx.xxxxx$), da für Südecuador 0.00001° etwa 1.11 Meter entspricht.

Da das Projektgebiet auf der Südhalbkugel liegt und wir uns westlich vom Nullmeridian befinden, sind alle geographischen Koordinaten negativ. Z.B. gilt für die ECSF-Station

_latitdue_min: südliche Begrenzung (-3.97254)
_latitdue_max: nördliche Begrenzung (-3.97238)
_longitdue_min: westliche Begrenzung (-79.07630)
_longitdue_max: östliche Begrenzung (-79.07600)

- **geocoordinate_amsl**

Höhe über Meeresspiegel in Meter (**above mean sea level**) (nur ganze Zahlen)

Beschreibt die Höhe, in welcher der Wert aufgenommen wurde. Diese kann in diesem Format nur plotweise übergeben werden. Handelt es sich bei einem Plot um ein Transekt, muss eine repräsentative mittlere Höhe angegeben werden.

Ist diese Information nicht bekannt oder nicht relevant, muss „NULL“ eingetragen werden

NULL oder 0 ?

„NULL“ steht in einem Tabellenfeld für welches keine Informationen vorliegen. Auch ein leeres Feld (ohne jeglichen Text) wird als NULL interpretiert und gespeichert.

Der **Wert 0** hingegen ist eine Information. Das heißt, es wurde an diesem Plot nach einem bestimmten Objekt (eine Art) gesucht, aber das Objekt (die Art) wurde hier mit den angegebenen Aufnahmemethoden zu diesem Zeitpunkt nicht gefunden.

Bei gemessenen Werten ist der Unterschied zwischen 0 und NULL offensichtlicher.

0 - es wurde gemessen und der Wert ist 0.00

NULL - es wurde an diesem Standort dieses Attribut nicht gemessenen (keine Information)