

Ein einheitliches Datenformat zum Austausch von Versuchsflächendaten der Sektion Ertragskunde des DVFFA - DVFFA_Data_V1.0 -

*Jürgen Nagel und Susanne Sprauer,
Nordwestdeutsche Forstlichen Versuchsanstalt, Abt. Waldwachstum, Göttingen*

Einleitung

Auf der letzten Ertragskunde-Tagung in Lenzen wurden wir von den beiden Obleuten "gebeten", einen Vorschlag für ein einheitliches Datenaustauschformat zu machen, damit in Zukunft Versuchsflächendaten zwischen den Versuchsanstalten reibungsloser und einfacher ausgetauscht werden können. Einen ersten Vorschlag hatten wir 2015 den Versuchsanstalten zu Stellungnahme geschickt. Daraus ist die im Folgenden beschriebene Version 1.0 entstanden.

Das Format ist zunächst einmal „recht einfach“ gehalten und für den Datenaustausch konzipiert. Es beansprucht nicht als Datenbankformat nutzbar zu sein. Das Datenformat lässt sich erweitern, wenn der Bedarf an zusätzlichen Datenfeldern entstehen sollte. Änderungen sollten dann über die Sektion bekanntgemacht werden.

Nach dem ersten Vorschlag fand die Verwendung von XML (Extensible Markup Language) für den Datenaustausch breite Zustimmung. In einer XML-Datei lassen sich die Daten mit Hilfe von sogenannten Tags beschreiben und in Hierarchieebenen ablegen. Darüber hinaus lassen sich XML-Dateien mit jedem Texteditor und auf allen Betriebssystemen bearbeiten. Man kann sie problemlos über das Internet verschicken. Die Erstellung und Bearbeitung einer XML-Datei aus einer Datenbankanwendung, die für einen Programmierer keine besondere Herausforderung ist, kann für den einfachen Datenbanknutzer aber eine unlösbare Hürde bedeuten. Damit das Datenaustauschformat trotzdem jedermann zugänglich ist, haben wir uns zusätzlich das Ziel gesetzt, dass das Datenaustauschformat für die Bearbeitung in eine einfache Datenbank ein- und ausgelesen werden kann.

Die Beschreibung des Datenaustauschformats, Beispieldateien, die Baumartencodierung nach der Bundeswaldinventur³ und ein Konvertierungsprogramm mit Programmiercode können von der Internetseite der Sektion Ertragskunde heruntergeladen werden.

DVFFA-Datenaustauschformat

In einer XML-Datei werden die Daten mit sogenannten Tags gekennzeichnet. Der Beginn von Daten wird durch den Anfangs-Tag (spitze Klammern mit einem frei wählbaren Namen, Beispiel: <Parzelle>) ausgedrückt und das Ende der zugehörigen Daten durch den End-Tag (entspricht dem Anfangs-Tag mit einem vorangestellten Schrägstrich, Beispiel: </Parzelle>). Für jeden beginnenden Tag muss ein endender Tag vorhanden sein, damit die XML-Datei später verarbeitet werden kann. Die Tags können geschachtelt werden, um die Daten zu strukturieren (Beispiel: <Parzelle> <Aufnahme> </Aufnahme></Parzellen>). In diesem Fall ähnelt die XML-Datei einer „Baumstruktur“ mit Knoten an den Verzweigungspunkten.

In der Tabelle 1 ist das Datenaustauschformat mit seinen Knoten als der „Baumstruktur“ dargestellt. Das sogenannte Root-Element **<Parzellen>** umschließt alle Daten, die ausgetauscht werden sollen. Es können die Daten einer oder mehrerer Parzellen ausgetauscht werden. Der Tag **<Parzelle>** umfasst alle Daten, die für eine einzelne Parzelle ausgetauscht werden. Unter dem Tag **<Metadaten>** werden allgemeine Informationen zur Parzelle abgelegt. Im Bereich des Tags **<Baumarten>** wird die Kodierung der Baumarten beschrieben, wobei jede einzelne Art über den Tag **<Baumart>** definiert wird. Die Beschreibung der Baumarten ist nötig, weil die verschiedenen Forschungs- und Versuchsanstalten unterschiedliche Baumartenschlüssel verwenden und so die Baumartencodierung innerhalb der Daten für eine Parzelle eindeutig ist, ohne dass der Benutzer wissen muss, woher die Daten stammen. Mit dem Tag **<Aufnahmen>** werden alle auf die Aufnahmezeitpunkte bezogenen Parzellendaten gebündelt, welche sich auf die einzelnen Aufnahmen durch das Tag **<Aufnahme>** gliedern. Für jede Aufnahme werden die Parzelleneckpunkte unter dem Tag **<Eckpunkte>** und für den einzelnen Parzelleneckpunkt unter **<Eckpunkt>** eingetragen. Die Parzelleneckpunkte müssen für jede Aufnahme neu angegeben werden, weil es zum Teil bei den Versuchsanlagen zu Veränderungen der Parzellengrenzen durch Flächenverkleinerung oder -vergrößerung kommt. Unter dem Tag **<Bestandeswerte>** werden baumarten- und schicht-

weise die Bestandeswerte unter **<Bestandeszeile>** ausgegeben. Mit dem Austauschformat können neben den Bestandeswerten auch die Rohdaten der Einzelbäume weitergegeben werden. Die einzelnen Bäume einer Aufnahme finden sich unter dem Tag **<Baumwerte>**, wobei die einzelnen Bäume durch den Tag **<Baum>** gekennzeichnet sind.

Tabelle 1: Datenaustauschformat mit Knoten als „Baumstruktur“. Fett gedruckt sind Knoten, die Datenattribute und -werte enthalten.

Ebene	Knoten (Tag)	Beschreibung
Root	<Parzellen>	Zwischen dem Root-Tag stehen sämtliche Daten
1	<Parzelle>	Mit jedem neuen Tag Parzelle beginnt eine neue Parzelle
2	<Metadaten>	Allgemeine Angaben zur Parzelle
2	<Baumarten>	In der Parzelle vorkommende Baumarten
3	<Baumart>	Beschreibung der Baumartenkodierung einer Baumart
2	<Aufnahmen>	Alle auf die Aufnahmezeitpunkte einer Parzelle
3	<Aufnahme>	Alle auf einen Aufnahmezeitpunkt bezogenen Daten einer Parzelle
4	<Eckpunkte>	Informationen zu der Lage der Parzelleneckpunkte
5	<Eckpunkt>	Informationen zu einem Parzelleneckpunkt
4	<Bestandeswerte>	Bestandeswerte einer Aufnahme (sofern berechnet)
5	<Bestandeszeile>	Bestandeswerte einer Aufnahme gegliedert nach Baumart (sofern berechnet)
4	<Baumwerte>	Alle Baumdaten einer Aufnahme
5	<Baum>	Alle Werte eines Baumes

In den Tabellen 2-7 sind sämtliche Attribute bzw. Variablen (Tags) und deren Maßeinheiten aufgelistet, die mit dem Datenaustauschformat übergeben werden können. Obwohl die Daten in dem vorgestellten XML-Format weitgehend definiert und auch selbsterklärend sind, müssen für einen reibungslosen Datenaustausch einige **zusätzliche Vereinbarungen** festgelegt werden. Grundsätzlich gilt:

- Die Bestandeswerte sind berechnete Werte, die abhängig vom Auswerteverfahren unterschiedlich sein können. In der Variablen DeserNorm sollte daher das Verfahren beschrieben werden.
- Bei den Baumwerten handelt es sich um Rohdaten, das heißt, es werden nur gemessene Werte eingetragen.
- Ein fehlender oder nicht gemessener Wert wird in der XML-Datei grundsätzlich durch den Eintrag `<tag>xsi:nil="true"</tag>` gekennzeichnet.
- In der XML-Datei wird als Dezimaltrennzeichen ein Punkt verwendet.
- Das Datum wird wie ein Text gespeichert. Daher gilt für diesen Datentyp die Vereinbarung, dass ein Datum in dem Format **jjjj-mm-dd** eingegeben sein muss. Das Format jjjj-mm-dd bedeutet: zuerst die vierstellige Jahreszahl, dann nach einem Trennstrich die zweistellige Monatszahl und schließlich nach einem weiteren Trennstrich die zweistellige Tageszahl des Monats (Beispiel 14. März 2016: `<datum>2016-03-14</datum>`). Die englische Schreibweise erleichtert bei den meisten Programmen das Verarbeiten der Datumsangabe.

Tabelle 2: Variablen und deren Maßeinheiten des Knotens *Metadaten*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Name>	Parzellenname bzw. Nummer oder Kennung	Text	-
<JahrBegrueudung>	Jahr der Bestandesbegründung des Bestandes	Integer	jjjj
<AlterBeiBegrueudung>	Alter bei Bestandesbegründung des Bestandes	Integer	
<JahrAnlage>	Jahr der Parzellenanlage	Integer	jjjj
<JahrAufgabe>	Jahr der Parzellenaufgabe	Integer	jjjj
<GeoReferenz>	Koordinatensystem für Geo-Punkte (UTM / ETRS89)	Text	
<Referenzpunkt>	Referenzpunkt der folgenden Koordinate (ungefähr, Zentrum, Ecke1, Ecke Nordost, etc..)	Text	
<Latitude>	Koordinate des Referenzpunktes, Breitenangabe	Double	m
<Longitude>	Koordinate des Referenzpunktes, Längenangabe	Double	m
<HoeheNN_m>	Koordinate des Referenzpunktes, Höhe über NN	Double	m
<Behandlungstyp>	z.B. A-Grad, B-Grad, etc.	Text	
<Versuchsziel>	z. B. Durchforstung, Herkunft, Düngung, etc.	Text	
<ErstelltAm>	Datum der Datenzusammenstellung	Date	jjjj-mm-tt
<ErstelltMit>	Programmname	Text	
<Datenbesitz>	Urheber (z.B. NW-FVA)	Text	
<DVFFADataFormat>	Version	Text	

Tabelle 3: Variablen und deren Maßeinheiten des Knotens *Baumart*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Code>	Codierung der Baumzeilen und Bäume (interne Schlüssel)	Text	-
<CodeBWI>	Baumartencode BWI	Integer	
<DtName>	Deutscher Name	Text	
<LatName>	Botanischer Name	Text	

Tabelle 4: Variablen und deren Maßeinheiten des Knotens *Aufnahme*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Datum>	Datum der Aufnahme	Date	jjjj-mm-tt
<Typ>	(1=Vollaufnahme, 2=repräsentative Aufnahme, 3=Teilaufnahme nur ausscheidender Bestand, 4=Teilaufnahme nur Z-Bäume, 5=Teilaufnahme nur Oberstand)	Integer	
<Kluppschwelle_cm>	Kluppschwelle (z.B. 7.0 cm)	Double	
<Flaechengroesse_ha>	Flächengröße der Parzelle	Double	ha
<Polygon>	1 = Eckpunkte vorhanden, 0 = Probekreis	Integer	
<Eingriff>	Läuterung, Hochdurchforstung	Text	
<Eingriffstaerke>	Kein Eingriff, mäßig, stark, 50%	Text	

Tabelle 5: Variablen und deren Maßeinheiten des Knotens *Eckpunkt*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Name>	Name des Eckpunktes (z.B. Ecke1)	Text	
<XEckpunkt_m>	X-Koordinate relativ zum Referenzpunkt	Double	m
<YEckpunkt_m>	Y-Koordinate relativ zum Referenzpunkt	Double	m
<ZEckpunkt_m>	Höhen-Koordinate relativ zum Referenzpunkt	Double	m

Tabelle 6: Variablen und deren Maßeinheiten des Knotens *Bestandeszeile*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Code>	Codierung der Baumart (interne Schlüssel, siehe Knoten Baumart)	Text	
<Schicht>	Schicht nach Forsteinrichtung (0 = Schichten nicht unterschieden, 1 = Hauptbestand, 2 = Unterstand, 3 = Verjüngung, 4 = Überhalt)	Integer	
<Alt>	Alter	Integer	Jahre
<NVerb_st_ha>	Stammzahl verbleibender Bestand	Integer	St/ha
<N7Verb_st_ha>	Stammzahl verbleibender Bestand größer Kluppschwelle	Integer	St/ha
<GVerb_qm_ha>	Grundfläche verbleibender Bestand	Double	m ² /ha
<VVerb_cbm_ha>	Volumen verbleibender Bestand	Double	m ³ /ha
<DgVerb_cm>	Dg verbleibender Bestand	Double	cm
<HgVerb_m>	Hg verbleibender Bestand	Double	m
<D100Verb_cm>	D100 verbleibender Bestand	Double	cm
<H100Verb_m>	H100 verbleibender Bestand	Double	m
<NAus_st_ha>	Stammzahl ausscheidender Bestand	Integer	St/ha
<N7Aus_st_ha>	Stammzahl ausscheidender Bestand größer Kluppschwelle	Integer	St/ha
<GAus_qm_ha>	Grundfläche ausscheidender Bestand	Double	m ² /ha
<VAus_cbm_ha>	Volumen ausscheidender Bestand	Double	m ³ /ha
<DgAus_cm>	Dg ausscheidender Bestand	Double	cm
<HgAus_m>	Hg ausscheidender Bestand	Double	m
<Desernorm>	Desernorm der Auswertung	Text	
<Bemerkung>	Bemerkung	Text	

Tabelle 7: Variablen und deren Maßeinheiten des Knotens *Baum*

<Tag>	Feldbeschreibung	Datentyp	Einheit
<Code>	Codierung der Baumart (interne Schlüssel, siehe Knoten Baumart)	Text	
<Kennung>	Kennung des Baumes, z.B. Baumnummer, Kennung in der Datenbank, RFID Nummer	Text	
<Schicht>	Schicht nach Forsteinrichtung (0 = Schichten nicht unterschieden, 1 = Hauptbestand, 2 = Unterstand, 3 = Verjüngung, 4 = Überhalt)	Integer	
<Alt>	Alter	Integer	
<AusGrund>	Ausschneide-Grund (0= Baum bleibt; 1= Durchforstung/Zielstärke; 2= zufällig (unbestimmt); 3= Wind/Sturm; 4= Schnee; 5= Insekten; 6= Pilze; 7= dürr; 8= fehlend, 9 = unterständig u. nicht weiter gemessen)	Integer	
<AusDatum>	Entnahmedatum (kann von Aufnahmedatum abweichen)	Date	jjjj-mm-dd
<DMesshoehe_m>	Höhe der Durchmessererhebung	Double	m
<D_cm>	Durchmesser in Messhöhe bei Umfangmessung	Double	cm
<DK1_cm>	Durchmesser in Messhöhe bei Klappung	Double	cm
<DK2_cm>	2. Durchmesser in Messhöhe bei Kreuzklappung	Double	cm
<Hoehe_m>	Baumhöhe	Double	m
<KAnsatz_m>	Kronenansatz	Double	m
<KBreite_m>	Kronenbreite	Double	m
<KBruch_m>	Höhe des Kronenbruchs	Double	m
<Kraft>	Baumklasse nach Kraft	Integer	
<ZBaum>	Z-Baum (0 = keine, 1= Z-Baum, xsi:nil="true" = nicht angesprochen)	Integer	
<DVFFA>	Baumklasse nach DVFFA	Text	
<IUFRO>	Baumklasse nach IUFRO	Text	
<X_m>	X-Koordinate relativ zum Referenzpunkt	Double	m
<Y_m>	Y-Koordinate relativ zum Referenzpunkt	Double	m
<Z_m>	Höhen-Koordinate relativ zum Referenzpunkt	Double	m
<Rand>	Baum liegt außerhalb der Fläche (0 = falsch, 1= wahr)	Integer	
<Flaechenfaktor>	Faktor des Baumes bei repräsentativer Aufnahme um den Flächenwert herzuleiten, im Normalfall = 1.0	Double	
<Bemerkung>	Freie Anmerkung zum Baum	Text	

Hilfsprogramm für das Datenaustauschformat

Zum besseren Verständnis, für eine einfache Bearbeitung und als Programmierbeispiel wurde das Hilfsprogramm DVFFA_Data.jar geschrieben. Es handelt sich um ein einfaches Java Programm mit zwei Funktionen:

- Schreiben einer XML-Datei im DVFFA-Austauschformat in eine SQLite-Datenbank
- Erzeugen einer XML-Datei im DVFFA-Austauschformat aus einer SQLite-Datenbank

Bei der SQLite-Datenbank handelt es sich um eine freie Datenbank, in der sämtliche Daten ähnlich wie bei Microsoft Access in einer Datei gehalten werden. SQLite kann auf fast allen Betriebssystemen eingesetzt werden. Für die Arbeit mit einer SQLite Datenbank empfiehlt sich das Firefox-Addon **SQLite Manager**, mit dem die Datenbank im Browser bearbeitet und angesehen werden kann. Es besteht die Möglichkeit, die Datenbanktabellen in Text-Dateien (z.B. CSV) zu exportieren. Beim Schreiben der XML-Datei in die SQLite-Datenbank werden die Abhängigkeiten, die sich aus der Struktur der XML-Datei ergeben, über zwei zusätzliche Variablen idParzelle und eine idAufnahme abgebildet. Das Programm erzeugt die in Tabelle 8 aufgeführten Tabellen. Die Struktur und die vorhandenen Tabellen der SQLite Datenbank sollten für eine Rücktransformation in die XML-Datei nicht geändert werden. Fehlende Werte werden in der SQLite-Datenbank mit dem Wert „NULL“ gekennzeichnet. Das Datum wird entsprechend der oben genannten Vereinbarung eingeben. Im Java Sourcecode der Klassen DVFFAdataexchange.java und XmlTool.java findet man die notwendigen Programmierzeilen, um mit einer eigenen SQLite-Datenbank das Austauschformat zu erzeugen bzw. zu lesen.

Tabelle 8: Erzeugen der SQLite-Tabellen

Tabelle	SQL – Befehl für SQLite
Parzelle	CREATE TABLE Parzelle (_id INTEGER NOT NULL PRIMARY KEY)
Metadaten	CREATE TABLE Metadaten (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, Name TEXT, JahrBegrueundung INTEGER, AlterBeiBegrueundung INTEGER, JahrAnlage INTEGER, JahrAufgabe INTEGER, GeoReferenz TEXT, Referenzpunkt TEXT, Latitude REAL, Longitude REAL, HoeheNN_m REAL, Versuchstyp TEXT, Versuchsziel TEXT, ErstelltAm DATE, ErstelltMit TEXT, Datenbesitz TEXT, DVFFADa- taFormat TEXT)
Baumarten	CREATE TABLE Baumarten (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, Code TEXT, CodeBWI INTEGER, DtName TEXT, LatName TEXT)
Aufnahmen	CREATE TABLE Aufnahmen (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, Datum DATE, Typ INTEGER, Kluppschwelle_cm REAL, Flaechengroesse_ha REAL, Polygon INTEGER)
Eckpunkte	CREATE TABLE Eckpunkte (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, idAufnahme INTEGER, Name TEXT, XEckpunkt_m REAL, YEckpunkt_m REAL, Z_Eckpunkt_m REAL)
Bestandeswerte	CREATE TABLE Bestandeswerte (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, idAuf- nahme INTEGER, Code TEXT, Schicht INTEGER, Alt INTEGER, NVerb_st_ha INTEGER, N7Verb_st_ha INTEGER, GVerb_qm_ha REAL, VVerb_cbm_ha REAL, DgVerb_cm REAL, HgVerb_m REAL, D100Verb_cm REAL, H100Verb_m REAL, NAus_st_ha INTEGER, N7Aus_st_ha INTEGER, GAus_qm_ha REAL, VAus_cbm_ha REAL, DgAus_cm REAL, HgAus_m REAL, Desernorm TEXT, Bemerkung TEXT)
Baumwerte	CREATE TABLE Baumwerte (_id INTEGER NOT NULL PRIMARY KEY , idParzelle INTEGER, idAufnahme INTEGER, Code TEXT, Kennung TEXT, Schicht INTEGER, Alt INTEGER, AusGrund INTEGER, AusDatum Date, DMesshoehe_m REAL, D_cm REAL, DK1_cm REAL, DK2_cm REAL, Hoehe_m REAL, KAnsatz_m REAL, KBreite_m REAL, KBruch_m REAL, Kraft INTEGER, ZBaum INTEGER, DVFFA TEXT, IUFR0 TEXT, X_m REAL, Y_m REAL, Z_m REAL, Rand INTEGER, Flaechenfaktor REAL, Bemerkung TEXT)

Danksagung

An dieser Stelle möchten wir uns bei all denen bedanken, die den ersten und zweiten Vorschlag durchgesehen und wichtige Anregungen für das Datenaustauschformat geliefert haben. Bitte haben Sie Verständnis, wenn Ihre Vorschläge am Ende vielleicht nicht aufgenommen wurden. Auch während der Programmierung des Hilfsprogramms haben sich noch einige Änderungen ergeben. Wir hoffen, dass dieses Austauschformat in der vorgeschlagenen Version hilfreich für die Zusammenarbeit sein wird und sind uns darüber klar, dass sicher noch Änderungsbedarf herrscht.