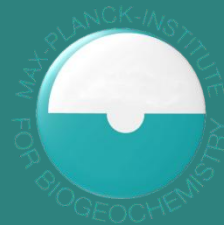




Management of environmental data collections - experiences of long-term, interdisciplinary research projects

J. Nieschulze MPI-BGC, Jena



Heterogeneous data sources, challenges

■ Many data types

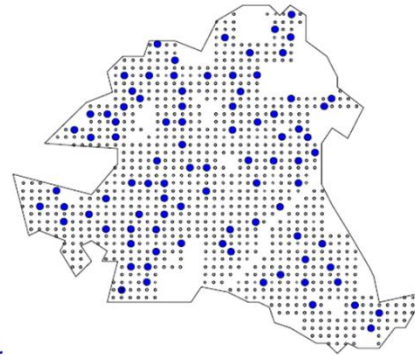
- space, time
 - » resolution
- observations
- models,...

■ Many data formats

- Excel, Access, DBF, ASCII,...
- orientation, matrices,....
- syntax, encoding

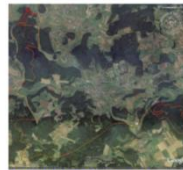
■ Individual Concepts / Semantics

- synonyms, homonyms
- acronyms, abbreviations

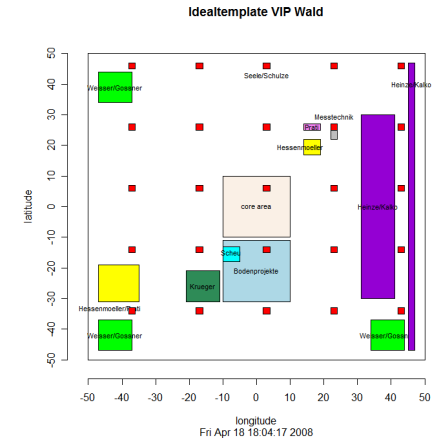
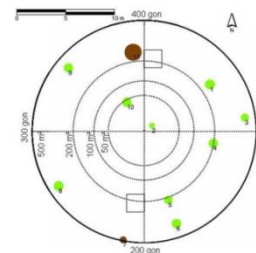
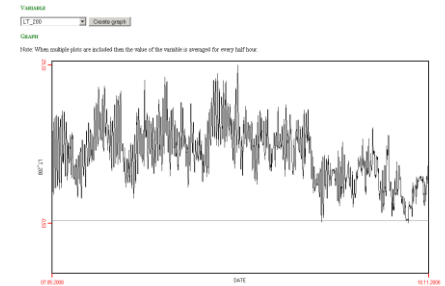


Gr

• Scheinwerferzählung:
standardisierte Routen entlang von Waldwegen



Transekt-Route Schwäbische Alb



Heterogeneous data sources, challenges

Two major challenges:

1. Capabilities of humans to interpret within context
 - » human readable \neq machine readable
 - » spreadsheets used for data entry

2. Semantics
 - » concepts are individual
 - » one man's meat is another man's poison

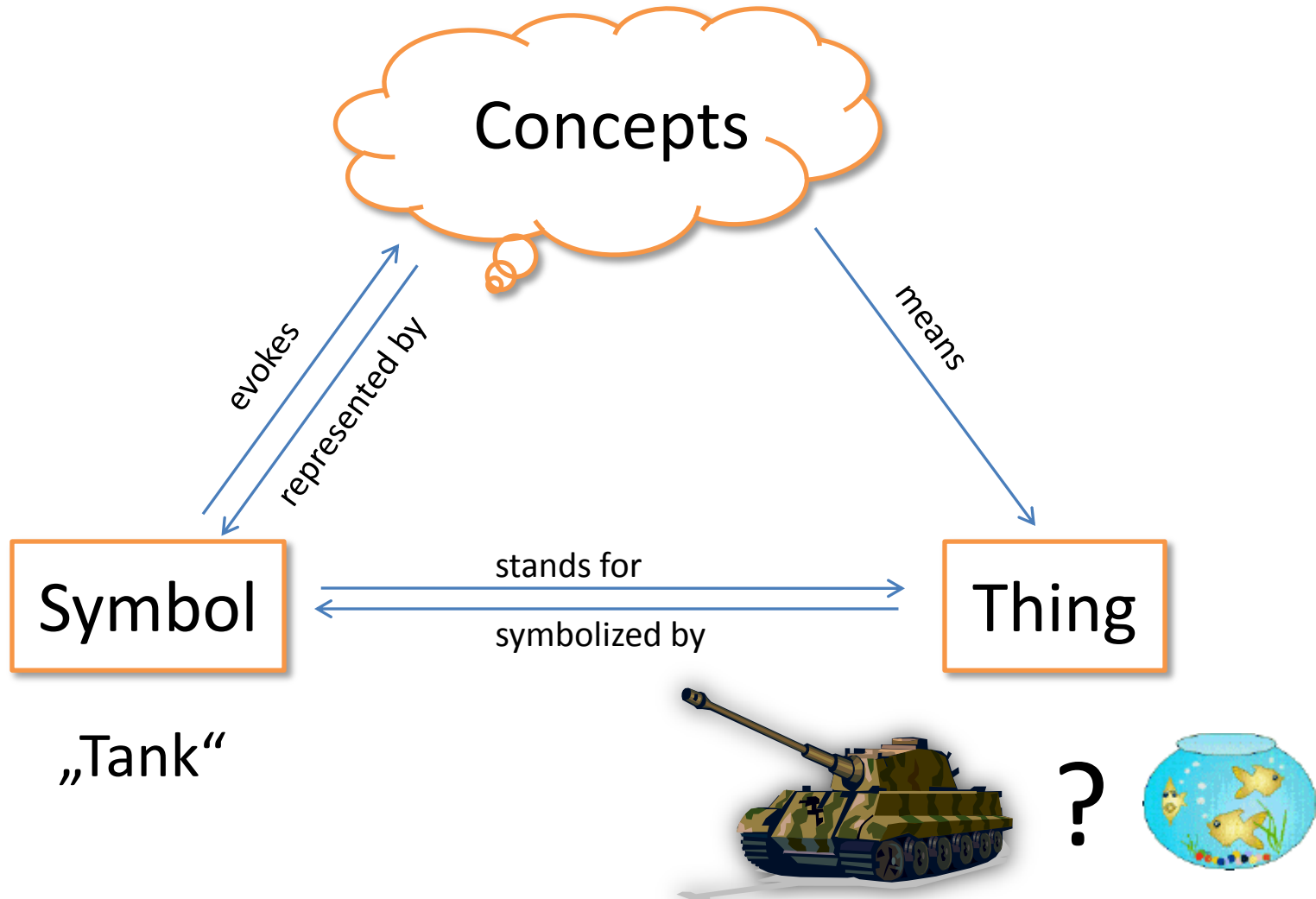
Heterogeneous data sources, challenges: human "API", intermingling of data types

Easy for a human to realize data within context; example shows several relations on one spreadsheet

P219		fx 6,40608695652174		C	D	E	F	G	H	I	J	K			
1	2	3	4	5	A	B	C	D	E	F	G	H	I	J	K
1							Automatic								
2					BHD(cm)		56,6591597	93,58311	57	67	78	55,70423	48	54	42
3					g (m²) Beginning of 2002		0,25213326	0,687836	0,255176	0,352565	0,477836243	0,243706	0,180956	0,229022	0,138544
4					Tree height (m)		34,5	36,4	34,6	34,2	34,5	34,8	26,4	33	33,4
5					Messumfg (cm)		178	294	179,0708	210,4867	245,044227	175	150,7964	169,646	131,9469
6					Circumferences (m)		1,78	2,94	1,790708	2,104867	2,45044227	1,75	1,507964	1,69646	1,319469
7	Date				Time										
8					tree number		1	2	3	4	5	6	7	8	9
9					other numbers										
	•	663	30.08.2005		38594		14,02	17,07	13,02	12,89	15,2	14,01	9,42	8,47	11,74
	•	664	13.09.2005		38608		14,03	17,09	13,02	12,91	15,22	14,01	9,42	8,49	11,75
	•	665	10.10.2005		38635		14,06	17,14	13,07	12,95	15,21	14,06	9,46	8,52	11,8
	•	666	25.10.2005		38650		14,075	17,17	13,07	12,96	15,24	14,055	9,65	8,55	11,83
	•	667	01.11.2005		38657		14,06	17,17	13,06	12,95	15,23	14,06	9,65	8,55	11,83
	•	668	21.11.2005		38677		14,06	17,25	13,07	12,96	15,25	14,09	9,63		11,87
	•	669													
	•	670	22.03.2005												
	•	671	29.03.2005												
	•	672	31.05.2005												
	•	673	14.07.2005												
	•	674	25.07.2005												
	•	675	10.08.2005												
	•	676	22.08.2005												
	•	677	30.08.2005												
	•	678	13.09.2005												
	•	679	11.10.2005												
	•	680	25.10.2005												
	•	681	01.11.2005												
	•	682													
	•	683	30.04.2002				6,03278	7,11667	6,65167	6,34595	6,72443	6,87833	6,75833	3,65833	4,02978
	•	684	01.11.2002				8,12785039	9,488739	8,336359	7,13688	8,505584321	8,455759	6,955215	5,13923	6,970874
	•	685	30.04.2003				8,61014493	10,092	8,7665	8,07066	9,097918782	8,818261	7,1095	5,4885	7,295178
	•	686	01.11.2003				10,3505263	11,36257	10,12261	9,555907	10,68877324	10,55566	7,973438	6,307463	8,635
	•	687	01.05.2004				10,6572743	12,51591	10,06895	9,694043	10,84305506	10,56494	8,044061	6,340424	8,586491
	•	688	01.11.2004				12,3452941	14,56412	11,58588	11,10941	13,97117647	12,27353	8,665294	7,144706	10,40765
	•	689	01.05.2005				12,52	14,81	11,645	11,42	13,84	12,43	8,895	7,17	10,43
	•	690	01.11.2005				14,06	17,21	13,065	12,955	15,24	14,075	9,64	8,55	11,85
	•	691													
	•	692	g End of 2002	(m)			0,25810344	0,69898	0,26	0,35522	0,484807996	0,248119	0,181429	0,233038	0,144789
	•	693	g End of 2003	(m)			0,263116	0,704986	0,263916	0,360232	0,491077566	0,253026	0,183511	0,235274	0,14768
	•	694	g End of 2004	(m)			0,26802381	0,714722	0,268337	0,36504	0,503522984	0,257898	0,185015	0,23748	0,151655
	•	695	g End of 2005	(m)			0,27254082	0,726215	0,272503	0,370292	0,509143415	0,262633	0,186828	0,241289	0,154791

Heterogeneous data sources, challenges: different semantics

Interplay of things, concepts, and symbols



Data Management → formats & semantics

Interpretation?

» RW?

Coordinate?

» 4382600?

What kind of projection?

DHDN4, ETRS89, PD83?

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1			Koordinaten neuer Mittelpunkt			
2						
3						
4			Plot	RW	HW	Bedeutung
5	1		20	4382600	5674600	neuer Mittelpunkt
6	2		32	4383921	5675121	neuer Mittelpunkt
7	3		68	4386710	5672800	neuer Mittelpunkt
8	4		274	4382900	5673600	neuer Mittelpunkt
9	5		1620	4386200	5676270	neuer Mittelpunkt
10	6		87	4387600	5675690	neuer Mittelpunkt
11	7		104	4375100	5684120	neuer Mittelpunkt
12	8		130	4376400	5684600	neuer Mittelpunkt
13	9		150	4377230	5682900	neuer Mittelpunkt
14	10		44604	4376200	5684800	neuer Mittelpunkt

Data Management → formats & semantics

Example: four rows of meta-information before primary data start

```

Notepad++ - H:\explos\Data\LoggerGKorte\CR800 Teststation_Daten.dat
File Edit Search View Format Language Settings Macro Run TextFX Plugins Window ?
swap.sql listDir.vbs cn.vbs ex.vbs verzeichnis.vbs viewWaldinventur.sql CR800 Teststation_Daten.dat
1 "TOA5", "CR800 Teststation", "CR800", "1638", "CR800.Std.01", "CPU:MESSTECHNISCHE_EINRICHTUNG"
2 "TIMESTAMP", "RECORD", "Temp_L_Avg", "Temp_100_Avg", "Temp_05B_Avg", "Temp_10B_Avg", "Temp_20B_Avg"
3 "TS", "RN", "Deg C", "Deg C", "Deg C", "Deg C", "Deg C", "Deg C", "", "", "mSec"
4 "", "", "Avg", "Avg", "Avg", "Avg", "Avg", "Avg", "Avg", "Avg", "Avg"
5 "2007-03-13 12:00:00", 0, 20.28, 19.91, 20.19, 20.29, 20.14, 19.99, 0, 0, 0
6 "2007-03-13 13:00:00", 1, 20.32, 19.96, 20.19, 20.29, 20.18, 19.99, 0.017, -0.002, 0.39
7 "2007-03-13 14:00:00", 2, 20.46, 20.11, 20.31, 20.42, 20.32, 20.1, 0.032, -0.003, 0.755
8 "2007-04-02 16:00:00", 3, 22.25, 21.7, 21.98, 21.95, 21.86, 21.83, 0, 0, 0
9 "2007-04-02 17:00:00", 4, 22.45, 21.83, 22.11, 22.14, 22.07, 21.89, 0.017, -0.002, 0.39
10 "2007-04-02 18:00:00", 5, 22.97, 22.1, 22.69, 22.27, 23.7, 22.87, 0.032, -0.004, 0.756
11 "2007-04-23 20:00:00", 6, 20.15, 19.25, 19.59, 19.38, 19.35, 19.63, 0, 0, 0
12 "2007-04-23 21:00:00", 7, 20.03, 19.24, 19.56, 19.36, 19.35, 19.58, 0.02, -0.002, 0.396
13 "2007-04-24 12:00:00", 8, 20.29, 26.49, 27.94, 30.96, 25.91, 29.34, 0, 0, 0
14 "2007-04-24 13:00:00", 9, 21.14, 27.18, 21.75, 21.03, 19.79, 20.08, 0.038, 0.025, 0.428
15 "2007-04-24 14:00:00", 10, 21.64, 29.88, 16.98, 14.25, 11.12, 9.78, 0.08, 0.051, 0.839
16 "2007-04-24 15:00:00", 11, 21.76, 29.51, 19.13, 15.62, 11.68, 9.79, 0.085, 0.054, 0.848
17 "2007-04-24 16:00:00", 12, 21.97, 30.27, 20.34, 16.71, 12.18, 9.8, 0.086, 0.054, 0.849
18 "2007-04-24 17:00:00", 13, 22.33, 30.43, 21.25, 17.56, 12.62, 9.85, 0.086, 0.054, 0.849
19 "2007-04-24 18:00:00", 14, 21.64, 27.15, 21.49, 18.19, 13.02, 9.89, 0.086, 0.054, 0.849
20 "2007-04-24 19:00:00", 15, 20.56, 22.85, 20.87, 18.42, 13.38, 9.95, 0.086, 0.054, 0.849
21 "2007-04-24 20:00:00", 16, 18.1, 18.65, 19.8, 18.2, 13.62, 10.02, 0.086, 0.054, 0.849
    
```

Data Management → formats & semantics

Example: orientation of variables; heterogeneous variable length

	A	B	C	D	E	F	G
13	Bemerkung						
14	Neigung	°					
15	Exposition		NW				
16	Wölbung_horizontal		X				
17	Wölbung_vertikal		G				
18	Reliefformtyp		KH				
19	Lage im Relief		Z				
20	Mikrorelief		RE				
21	Bodenabtrag/-auftrag						
22	Vegetation		LW				
23	Witterung		WT2				
24	Entfernung Baum	cm					
25	Entfernung Stubben	cm					
26	Bemerkung						
27	Bedeckungsgrad	%					
28	Beginn Steine	cm	15				
29	Anzahl Horizonte		4				
30	Horizontgrenze	cm	+4	7	20	35+	
31	Horizontschärfe		sc	de	sc		
32	Black Carbon						
33	Carbonatgehalt						
34	oxidative Merkmale						
35	reduzierte Merkmale						
36	sonstige pedogene Merkmale						
37	Bodenart			Ut4	Tu4		
38	Horizontsymbol		L	Ah	Bv	cCv	
39	Kernverlust		nein	nein	nein	nein	
40	Bemerkung						
41	Bodensystematische Einheit		BB				
42	Bodensubtyp		BBn				
43	Humusform						

Challenges, examples

- ≈ Inconsistent data structure []
- ≈ Several variables per column []
- ≈ Several units per variable []

Date	Horizon	Slope
...	10-40-B	10% S
...	1030B	11 W
...	B_10_20	0

Challenges , examples → synonyms

- ✓ Corresponding data sets with different identifiers

ID	X_Coord	Y_Coord
1	3514000	5355700
2	3514000	5355800
3	3514000	5355900

Site	RW	HW
1	351400	5355700
2	351400	5355800
3	351400	5355900

Challenges , examples

- ✓ Data sets same variable different content (Gauss-Krueger DHDN5 & ETRS89)

RW	HW
5432900	5886600
5433000	5886500
5432900	5886501

RW	HW
3432792,4	5884690,2
3432892,4	5884590,3
3432792,4	5884591,3

Challenges

✓ Data sets with different scales

A	Temp C/min
...	13,2
...	13,1
...	13,1

A	Temp C/h
...	13,2
...	14,1
...	14,6

Data Management → challenge of encoding when using Umlauts

Gesamttabelle Hainich sortiert_02.xlsx - M

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Entwicklertools

AO43

	A	AK	AL	AM	AN	AO
1	Vegetations-/Nutzungstyp	Mittelwerte				
14	Moose bearbeitet	0.6	0.6	0.6	1.0	0.0
15	Flechten bearbeitet	0.3	0.1	0.4	0.9	0.0
	Pflegemaßnahmen	0.0		0.0		
	Phänologie					
18	Bemerkung					
19	Deckung vegetation ges. [%]	94	94	90	92	92
20	Deckung b1 (>5-10m)	1.9	34	7.3		
	Mittlere Höhe b1 (m)	7.5	6.5	7.2		
22	Deckung s (verholzt) [>0-5m] [%]	7.2	3.5	3.9	38	38
	mittlere Höhe s [m]	0.7	0.4	0.7	1.8	1.8
24	Deckung krautschicht [%]	92	89	85	62	90
	mittlere Höhe kraut	0.4	0.4	0.3	0.2	0.3
26	Deckung moosschicht [%]	8.2	9.3	15	36	36
27	Deckung flechten	0.1	1.1	0.4	0.9	0.9
28	Algen (z.b.nostoc)	0.0		0.2	0.5	0.5
29	Streu	28	57	30	34	34
30	Totholz	0.1	1.0	0.3	0.5	0.5
31	Gestein (<5cm)	3.5		0.8	3.5	3.5
32	Gestein (>5cm)	1.9		2.9	1.8	1.8
33	Offene erde	4.8	3.9	5.0	5.3	5.3

Best Practice → how it ended up being designed

data capture independent of form and presentation

GRLD-Nutzungsschema-Datenbank_080326.xls [Kompatibilitätsmodus] - Microsoft Excel

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Entwicklertools

C1 Plot-ID

	C	D	E	F	G	H	I
1	Plot-ID	Nutzungsart	Pflege Trend	Nachsaat Trend	Beweidung Trend	Mahdt Trend	Duengung Trend
62	A8338	Wiese					
63	A8338	Wiese	x	x			
64	A8338	Wiese	x	x			
65	A8338	Wiese	x	x			
66	A8338	Wiese	x	x			
67	A8338	Wiese	x	x			
68	A8338	Wiese	x	x			
69	A8338	Wiese	x	x			
70	A8338	Wiese	x	x			
71	A8338	Wiese					
72	A8338	Wiese				x	
73	A8338	Wiese				x	
74	A8338	Wiese				x	
75	A8338	Wiese					
76	A8338	Wiese					
77	A8338	Wiese					
78	A8338	Wiese				x	
79	A8338	Wiese				x	
80	A8338	Wiese				x	
81	A8338	Wiese					
82	A8338	Wiese					

Best Practice → how it was originally designed

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1.Registration, Location:				2.Land information, factors of soil formation											
2	No.:	Location:			distance orientation		Major Landform: SE		Local landform: SE		Landuse: FN1		Parent material: SO1			
3	Dat	22-Apr			next tree:	1.25	310°		Position: UP		Gradient (%): 05		Crops or veg.: FD			
4	Author: EW				stubben:	/	/		Slope form: SS		Orientation: NE		Humaninfluence: VM		Age of land surface:	
5	Weather conditions: OV WC4					Erosion: N	Rock outcrops: N				Hn					
6	3.Observations	Depth	Horizon designation	Boundary dist. Top.	Munsell Colour (moist)	Texture fi. earth KA5	Fragm. (vol %)	Mottling (class)	Structure Grade, Type, Size		Cutans	Concre-tions	Roots	other biological features		
7		6.5	L	AW												
8			Of	AW												
9		2.5														
10		7	Ah	CS	2,5Y 4/2	Lu	N	N	WE, GR, ME		N	N	MV, FV	B		
11		21	BEw	CS	10YR 5/4	Tu4	N	N	WE, SB, ME		N	N	FV, MV, CV	B		
12		38	Btg		2,5Y 5/6 10YR 5/4	Tl	CS	CFDC	ST, SB, FI		N	FM, F	MV, CV	/		
13		38+	BCtg		2,5Y 5/6 10YR 5/4	Tl	CS	CFDC	ST, SB, FI		N	FM, F	N	/		
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30	Depth to C or R (cm) 38			Main rooting space (cm) 37		Groundwater act. (cm): not known				4. Soil Classification						
31	Comments: core/borehole 38/38											Stagnic Luvisol				
32																
33																
34																
35																
36																

Best Practice → how it ended up being designed; data capture independent of form

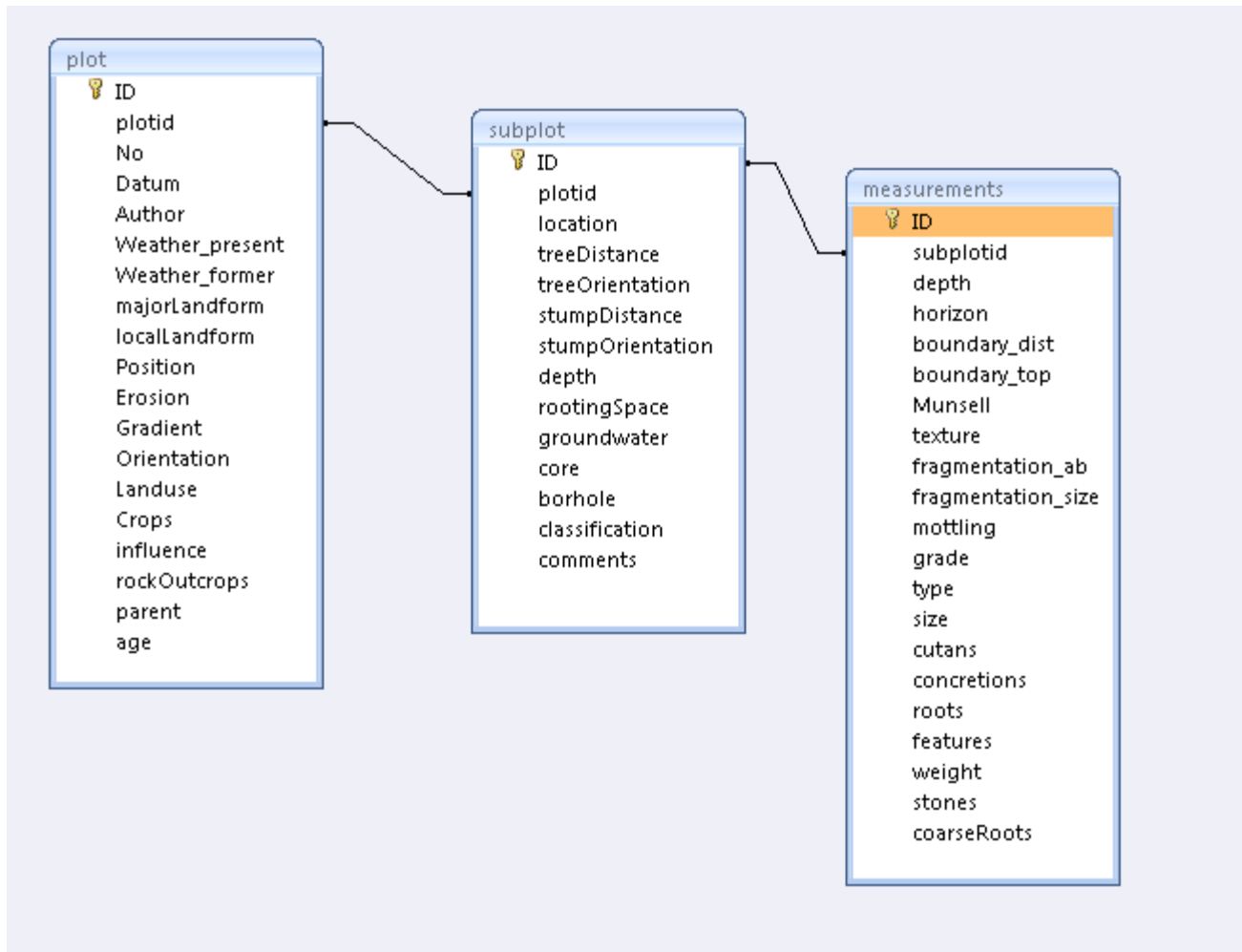
Eingabe

<p>plotid: <input type="text" value="HEG1"/></p> <p>No: <input type="text" value="42"/></p> <p>Date: <input type="text" value="12.12.1994"/></p> <p>Author: <input type="text"/></p> <p>Weather conditions: <input type="text"/></p> <p>Major landform: <input type="text"/></p> <p>Local landform: <input type="text"/></p> <p>Position: <input type="text"/></p> <p>Erosion: <input type="text"/></p> <p>Gradient (%): <input type="text"/></p> <p>Orientation: <input type="text"/></p> <p>Landuse: <input type="text"/></p> <p>Crops or veg.: <input type="text"/></p> <p>Human influence: <input type="text"/></p> <p>Rock outcrops: <input type="text"/></p> <p>Parent material: <input type="text"/></p> <p>Age of land surface: <input type="text"/></p>	<p>Location:</p> <p>Plotid: <input type="text" value="HEG1"/></p> <p>Location: <input type="text" value="A"/></p> <p>Tree distance: <input type="text" value="2"/></p> <p>Tree orientation: <input type="text"/></p> <p>Stump distance: <input type="text"/></p> <p>Stump orientation: <input type="text"/></p> <p>Depth to C or R (m): <input type="text"/></p> <p>Main rooting space: <input type="text"/></p> <p>Groundwater act.: <input type="text"/></p> <p>Core: <input type="text"/></p> <p>Borhole: <input type="text"/></p> <p>Soil classification: <input type="text"/></p> <p>Comments: <input type="text"/></p>	<p>Measurement</p> <p>Depth: <input type="text" value="5"/></p> <p>Horizon designation: <input type="text" value="Ah"/></p> <p>Boundary dist. Top: <input type="text" value="AS"/></p> <p>Munsell color (moist): <input type="text" value="2,5Y 4/2 7,5Y 4/"/></p> <p>Texture fi. earth KA5: <input type="text" value="Lu"/></p> <p>Fragm. (vol%): <input type="text" value="N"/></p> <p>Mottling (class): <input type="text" value="N"/></p> <p>Grade (structure): <input type="text" value="WE"/></p> <p>Type (structure): <input type="text" value="SB"/></p> <p>Size (structure): <input type="text" value="FI"/></p> <p>Cutans: <input type="text" value="N"/></p> <p>Concretions: <input type="text" value="N"/></p> <p>Roots: <input type="text" value="VFC, FF, MV"/></p> <p>Other biol. features: <input type="text" value="B"/></p> <p>Weight: <input type="text" value="349.6"/></p> <p>Stones weight: <input type="text" value="0"/></p> <p>coarse roots weight: <input type="text" value="0"/></p>
--	--	---

Datensatz: 1 von 2 Kein Filter Suchen

Datensatz: 1 von 3 Kein Filter Suchen

Best Practice → how it ended up being designed; structuring of data



Data entry → supply of templates secures common syntax and in part common semantics

eingabeMetkeSchemakonform.xls [Kompatibilitätsmodus] - Microsoft Excel

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Entwicklertools

E7

	B	C	D	E
1			Angaben zur Grünlandnutzung im Biodiversitätsexploratorium Schwäbische Alb	
2			Interviewer	
3			Datum	
4	Allgemeine Angaben zur Fläche	01	Projekt-ID der Untersuchungsfläche	
5		02	Name des Bewirtschafters	
6		03	Anschrift des Bewirtschafters	
7		04	Bewirtschaftungseinheit	
8		04.1	Angabe der Flurstücknummern	<ul style="list-style-type: none"> ein Flurstück mehrere Flurstücke Teil des Flurstücks (bitte in Karte skizzieren)
9		05	Größe der Bewirtschaftungseinheit [ha]	
10		06	Die Fläche liegt in folgender Gemarkung	
11	Angaben zur Nutzungsart	07	Nutzungsart im Jahr 2006 - bitte Erläuterung zu Mischform/Sonstige	
12		07.1	Erläuterung	
13		08	Wurde diese Nutzung auch in den vorherigen 10 Jahren durchgeführt? - wenn nein, bitte Erläuterung	
14		08.1	Erläuterung	
15		09	Ist die Art der Nutzung mit gleichbleibender Intensität auch für die nächsten Jahre geplant / absehbar?	
16		09.1	Erläuterung	

~~Best Practice~~

→ when you have lots of time at hand

functionality versus human readability

	A	B	C	D	E	F
1	Plot		20			
2						
3						
4		RW	HW			RW (Trimble) f
5	Mittelpunkt	4382600	5674600			
6						
7		RW	HW			eingemessen mit:
8	Inventur Ecke NW	4382590	5674610			Suunto, Vertex, Fluchtstangen
9	Inventur Ecke NO	4382610	5674610			Suunto, Vertex, Fluchtstangen
10	Inventur Ecke SW	4382590	5674590			Suunto, Vertex, Fluchtstangen
11	Inventur Ecke SO	4382610	5674590			Suunto, Vertex, Fluchtstangen
12						
13						
14		Richtung	Meter			
15	Verschiebung					
16						
17		RW	HW			eingemessen mit:
18	neuer Mittelpunkt	4382600	5674600			
19						
20		RW	HW			eingemessen mit:
21	EP Ecke NW	4382550	5674650			4382550
22	EP Ecke NO	4382650	5674650			4382651
23	EP Ecke SW	4382550	5674550			4382546
24	EP Ecke SO	4382650	5674550			4382645
25						
26	Zaun Quadrant	NO				
27	Zaun Korrektur nach S					
28	Zaun Korrektur nach N					
29	Zaun Korrektur nach W					
30	Zaun Korrektur nach O					
31						
32		RW	HW			eingemessen mit:
33	Zaun NW	4382618	5674630			
34	Zaun NO	4382630	5674630			
35	Zaun SW	4382618	5674618			
36	Zaun SO	4382630	5674618			
37						
38		RW	HW			eingemessen mit:
39	Startpunkt	4382623	5674550			4382620
40						
41						

System development → general settings (technical and administrative)

- What kind of demands has the system to fulfill?
 - » Ressources (project administration)
 - » Maps
 - » Data management
- Objectives of the system/database?
 - » Awareness of IT potential?
 - » Awareness of heterogeneous data sources?
- Data exchange/interchange regulation needed
 - » Protocol of data exchange
 - » Copy rights
- Data access
 - » Compliance with international standards (GML?)
 - » Public access?

Example target goals data access to be specified by steering committee

- visualization
 - » maps
 - » graphical data analysis
- query by metadata
- query by maps
- form based queries featuring
 - » filter
 - » aggregation (group by)
 - » merge/joins over several data sets
- quantitative analysis capabilities
- Downloads (API, syntax)

Objectives

Importance of meta data

- » only that what is documented can be interpreted, analyzed, merged, and re-used
- » mandatory supply of meta information **prior** to the field work
- » optimized effort, support by data management necessary
- » added value integration; synergies; re-use

Hardwired semantical registration → XML schema driven forms

Location

Exploratory

Key explo:

Value explo:

Plots

Key plot:

Value plot:

SRS

EPSG:

Coordinates

Longitude:

Latitude:

The reference to the exploratory:

- HAI, ALB, SCH: The data belongs to Hainich, Schwäbische Alb or Schorfheide.
- asVariable: The exploratory is stored as a variable in the data (provide variable below)
- none: The data does not come from an exploratory or the exploratory is already included in the plot number indirectly

The reference to the plot:

- asVariable: The plot is stored as a variable in the data (provide variable below)
- asPlotname: The data comes from a single plot (provide plot name below)
- none: The data does not come from a plot

The Spatial Reference System of the location of the data (optional). Provide the EPSG coordinate system and the longitude (x) and latitude (y) accordingly.

Variable17

Name17: + - ↑ ↓ □

Type17:

Block17:

Unit17:

Description17:

Metadaten-Schema

```

1  <?xml version="1.0"?>
2  <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
3    targetNamespace="http://www.bgc-jena.mpg.de"
4    xmlns="http://www.bgc-jena.mpg.de" elementFormDefault="qualified">
5
6    <!-- ===== dataType/locationType -->
7    <xsd:complexType name="locationType">
12     <xsd:sequence>
13       <xsd:element name="exploratory">
14         <xsd:complexType>
15           <xsd:sequence>
16             <xsd:element name="keyExplo" type="exploType" />
17             <xsd:element name="valueExplo" type="stringMeta" />
18           </xsd:sequence>
19           <xsd:attribute name="meta" type="xsd:string" />
20           <xsd:attribute ref="bgc:source" />
21         </xsd:complexType>
22       </xsd:element>
23     </xsd:sequence>
24     <xsd:element name="plots">
25       <xsd:complexType>
26         <xsd:sequence>
27           <xsd:element name="keyPlots" type="plotType" />
28           <xsd:element name="valuePlots" type="stringMeta" />
29         </xsd:sequence>
30         <xsd:attribute name="meta" type="xsd:string" />
31         <xsd:attribute ref="bgc:source" />
32       </xsd:complexType>
33     </xsd:element>
34     <xsd:element name="srs" minOccurs="0">
35

```

XML based formalization

- » machine readable
- » connection to standards

Hardwired semantical registration

RESEARCH OBJECTS

Taxa

Taxon1 +

Processes

Process1 +

Services

Service1 +

The taxa covered by the dataset. Use the + and - buttons to add and remove taxa. Use the up and down buttons to change the sequence of taxa.

The processes covered by the dataset. Use the + and - buttons to add and remove processes. Use the up and down buttons to change the sequence of processes.

The services covered by the dataset. Use the + and - buttons to add and remove services. Use the up and down buttons to change the sequence of services.

Context search

RESEARCH OBJECTS

- ▼ Taxa (Clear selection)
 - ▷ arthropods
- ▼ Processes
 - ▷ primary producer abundance
 - ▷ primary consumer abundance
 - ▷ co2 storage
 - ▷ water retention
 - ▷ decomposer activity
 - ▷ decomposer diversity
 - ▷ soil nutrient supply
 - ▷ nutrient concentration in plants
 - ▷ nutrient loss
 - ▷ invader abundance
 - ▷ diversity of other trophic levels
 - ▷ stability of various processes
 - ▷ regulation of hydrological flow
 - ▷ recovery from mobile nutrients
 - ▷ removal and breakdown of excess compounds
 - ▷ removal and breakdown of xenic compounds
 - ▷ pollinator abundance
 - ▷ pollinator diversity
 - ▷ dispersal activity
 - ▷ predation
 - ▷ none
 - ▷ other
- ▼ Services (Clear selection)
 - ▷ nutrient cycling

ALL DATASETS

Soil Arthropods in Grasslands

Dataset Id: **2944**

Version: **1.0.0**

File type: **structuredData**

Date last modified: **17.04.2008**

Authorized: [Yes, show data](#)

[<< 1 >>](#)



Example of semantic agreement

- Plotid of regular grid points:
 - » regular expression $(A|H|S)\d{1,5}$
 - » example: **A49542** or **H1**

- EP:
 - » regular expression $(A|H|S)E(W|G)\d{1,2}$
 - » example: **SEG1**, **HEW42**

Best Practice

- » Structure your data
 - » columns/rows
- » NO empty columns or rows
- » NO umlauts, ß, or whitespace in labels
- » Think of/determine the data type BEFORE you enter data (>80cm; ca. 40)
 - » Numerical data such as integers or real numbers are the most meaningful
- » NO compound variable
 - » 1 column/row per variable

Best Practice

- » Provide structured metadata
- » Provide UNITS (data without units are useless!)
- » Agree on core semantics
- » Formalize data structure (syntax)

```
<!-- ===== dataType/dataStructureType-->
<xsd:complexType name="dataStructureType">
  <xsd:sequence>
    <xsd:element name="format">
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="fileFormat" type="fileFormatType"/>
          <xsd:element name="fileFormatVersion" type="stringMeta"
            minOccurs="0"/>
          <xsd:element name="delimiter" type="delimiterType"/>
          <xsd:element name="decimalCharacter" type="decimalType"/>
          <xsd:element name="quote" type="stringMeta"/>
          <xsd:element name="missingValues" type="stringMeta"/>
          <xsd:element name="orientation" type="orientationType"/>
          <xsd:element name="offset">
```

