

Big Data, organisation and analysis

Steffen M. Noe, Spring 2023

What is data?

- In a way, almost everything!
- Data can be in *analog* and digital form



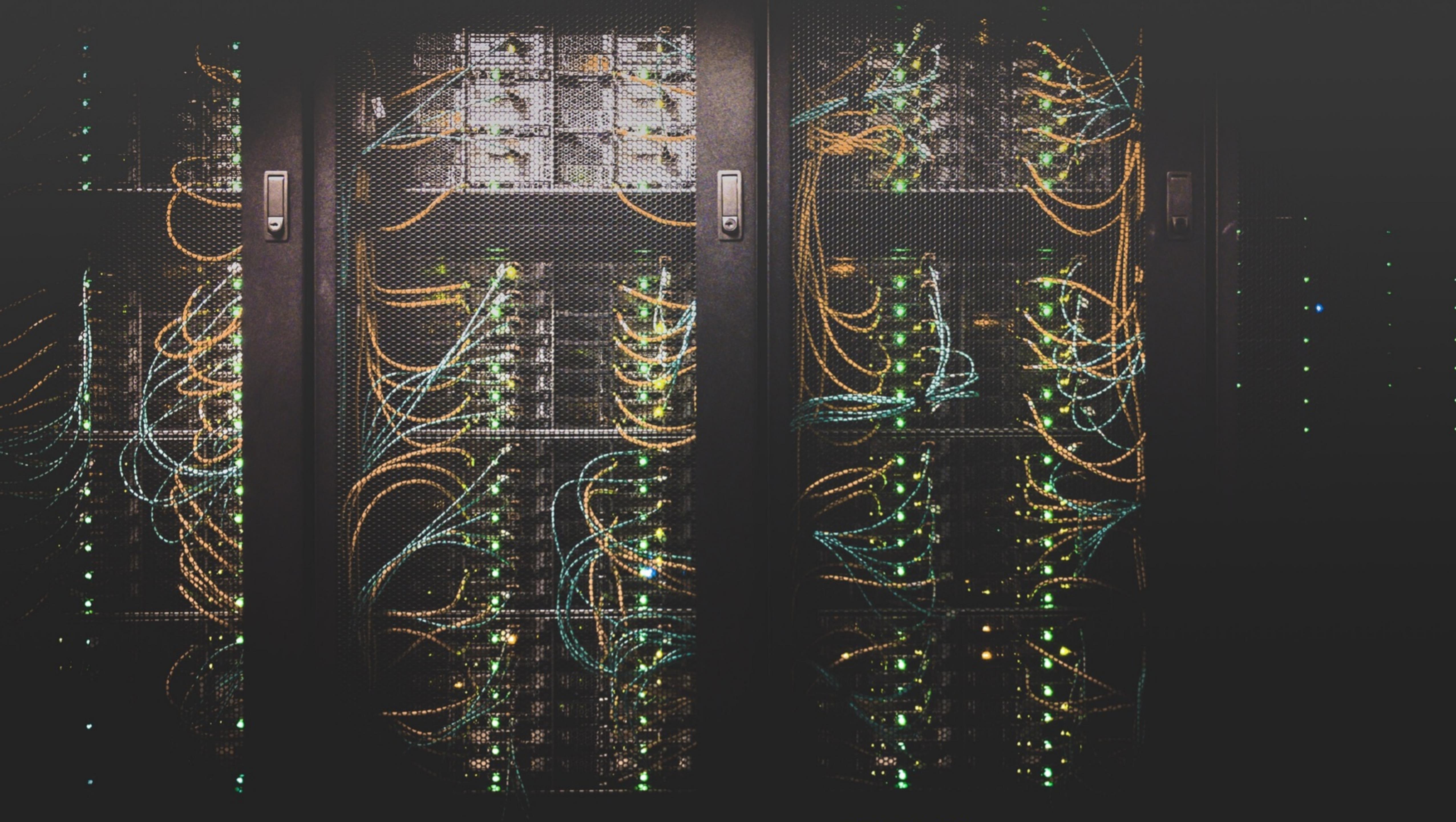
Proovi nr. ja suur- kus	Vahikond kvarv. nr. nr. Proovi täiki asetuse skem Muld Maapinnal kate	Puu seis. Bon- teet	1926-aasta mõõtmine						1929-aasta mõõtmine						9. VI			1930-aasta mõõtmine 17. 10. 30.												
			Puu nr. E. H. B.	Läbimõõt sentim.	Läbilõike piind cm²	Tähendused			Puu nr. nr.	Läbimõõt sentim.	Tähendused			Puu nr. nr.	Läbimõõt sentim.	Tähendused			Puu nr. NS WO Kesk.	Puu nr. NS WO Kesk.	Puu nr. NS WO Kesk.	Läbilõike piind cm²	Puu nr. NS WO Kesk.							
						NS	WO	Kesk.			NS	WO	Kesk.			NS	WO	Kesk.												
1. Kivisaare vltk.	Mai	1. 11.3 11.4 11.3 100.3	4	16.0	13.5	1	11.9 11.9 11.9	Mä	Puude arv	1	12.2 12.2 12.2	Mä	Puude arv	1	12.2 12.2 12.2	Mä	Puude arv	1	12.2	12.2	12.2	12.2	Läbilõike piind cm²	12.2						
0.06ha KV. 280.	--	2. 11.5 11.2 11.3 100.3	8	16.5	12.5	2	11.3 11.6 11.4	7.0	1	2	11.8 12.0 11.9	7.5	1	0.0444	3	0.0151	3	0.0151	3	0.0151	3	0.0151	3	0.0151	3					
	--	3. 7.5 7.5 7.5 44.18	13	15.0	13.0	3	1.0-0.3 0.0310	8.5	7	3	Kuivakud	9.0	12	3	Kuivakud	8.5	5	0.0284	9.0	13	0.0827	9.0	13	0.0827	9.0	13				
	--	4. H 13.2 13.8 13.5 143.1	25	16.0	12.5	4	13.6 14.0 13.8	10.0	12	4	13.9 14.4 14.1	9.5	5	4	13.9 14.4 14.1	9.5	5	0.0354	10.0	11	0.0864	10.0	11	0.0864	10.0	11				
	--	5. 19.6 20.6 20.1 317.3	29	15.5	13.7	5	20.8 21.4 21.1	11.0	5	5	21.4 22.3 21.9	10.5	9	5	21.4 22.3 21.9	10.5	9	0.0779	11.0	5	0.0475	11.0	5	0.0475	11.0	5				
	--	6. 15.1 14.8 14.9 174.4	41	17.0	14.0	6	15.9 15.3 15.6	12.0	11	6	15.7 16.7 15.7	11.5	5	6	15.7 16.7 15.7	11.5	5	0.0519	12.0	11	0.1244	12.0	11	0.1244	12.0	11				
	--	7. 11.8 11.3 11.5 103.9	56	16.5	13.5	7	11.8 11.5 11.7	13.0	7	7	11.8 11.5 11.6	12.5	4	7	11.8 11.5 11.6	12.5	4	0.0491	13.0	4	0.0531	13.0	4	0.0531	13.0	4				
	--	8. 12.8 12.3 12.5 122.7	69	15.0	13.0	8	13.3 12.8 13.0	14.0	4	8	13.3 13.1 13.2	13.5	7	8	13.3 13.1 13.2	13.5	7	0.1002	14.0	5	0.0769	14.0	5	0.0769	14.0	5				
	--	9. 20.0 20.2 20.1 317.3	82	16.5	13.5	9	21.4 21.8 21.6	15.0	6	9	22.0 22.6 22.3	14.5	1	9	22.0 22.6 22.3	14.5	1	0.0165	15.0	3	0.0530	15.0	3	0.0530	15.0	3				
	--	10. 10.1 9.9 10.0 78.54	91	16.0	13.3	10	10.3 10.2 10.3	16.0	5	10	10.3 10.1 10.2	15.5	5	10	10.3 10.1 10.2	15.5	5	0.1132	16.0	5	0.1005	16.0	5	0.1005	16.0	5				
	--	11. 14.4 14.0 14.2 158.4	106	15.5	13.8	11	15.6 14.7 15.1	17.0	5	11	15.6 15.5 15.6	16.5	4	11	15.6 15.5 15.6	16.5	4	0.0855	17.0	5	0.1135	17.0	5	0.1135	17.0	5				
	--	12. 12.1 13.1 12.6 124.7	127	16.5	14.0	12	12.6 13.9 13.3	18.0	1	12	13.0 13.2 13.5	18.5	2	12	13.0 13.2 13.5	18.5	2	0.0569	19.0	2	0.0962	19.0	2	0.0962	19.0	2				
	--	13. 12.4 13.6 13.0 132.7	135	16.0	13.5	13	13.0 13.9 13.4	19.0	1	13	13.0 13.9 13.4	19.5	4	13	13.0 13.9 13.4	19.5	4	0.0567	19.0	2	0.0567	19.0	2	0.0567	19.0	2				
	--	14. 7.0 7.2 7.1 39.59	140	15.5	13.4	14	7.0 7.1 7.1	20.0	1	14	7.0 7.1 7.1	20.5	4	14	7.0 7.1 7.1	20.5	4	0.0943	20.0	3	0.0943	20.0	3	0.0943	20.0	3				
	--	15. 10.2 11.5 10.3 83.32	152	16.0	12.5	15	10.4 10.7 10.5	21.0	1	15	10.4 10.7 10.5	21.5	3	15	10.6 10.6 10.6	20.5	-	15	10.6 10.6 10.6	20.5	-	0.1732	21.0	5	0.1732	21.0	5	0.1732	21.0	5
	--	16. 11.7 11.3 11.0 254.5				16	9.1 9.1 9.1	24.0	1	16	18.8 19.1 19.0	23.0	2	16	9.1 9.1 9.1	24.0	1	0.0398	23.0	3	0.1140	23.0	3	0.1140	23.0	3				
	--	17. 8.9 9.0 8.9 62.21				17	9.1 9.1 9.1	24.0	1	17	18.3 18.7 18.5	146	2.2 2.165 $d = 13.9 \text{ sm}$	1	17	18.7 19.5 19.2	23.5	1	0.0416	23.0	1	0.0434	23.0	1	0.0434	23.0	1			
	Maapinnalangus	--	18. 11.4 11.9 11.6 243.3	4.0	1	18	18.3 18.7 18.5			18	12.1 12.0 12.0			18	12.1 12.1 12.1	24.0	-	18	12.1 12.1 12.1	24.5	-	0.0452	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	N. rihis. II-wanadra	--	19. 11.5 11.4 11.4 102.1	4.5	3	19	9.5 9.5 9.5			19	9.5 9.5 9.5			19	9.5 9.5 9.5	24.0	-	19	9.5 9.5 9.5	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	Klar. Muromets ja järukasv puudutav	--	20. 9.4 9.2 9.3 67.93	4.5	4	20	9.5 9.5 9.5			20	9.5 9.5 9.5			20	9.5 9.5 9.5	24.0	-	20	9.5 9.5 9.5	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	Pindlate Hyponum,	--	21. 9.1 9.1 9.1 65.04	4.5	9	21	10.4 10.7 10.5			21	10.4 10.7 10.5			21	10.4 10.7 10.5	24.0	-	21	10.4 10.7 10.5	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	Lycopodium annotinum	--	22. 16.1 16.4 16.2 206.1	4.0	6	22	16.6 16.9 16.8			22	17.0 18.3 17.6			22	17.0 18.3 17.6	24.0	-	22	17.0 18.3 17.6	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	Pinola scunda, Prola uniflora, Equisetum	--	23. 16.3 17.4 16.8 221.7	4.5	8	23	10.8 10.8 10.6			23	10.8 10.8 10.6			23	10.8 10.8 10.7	24.0	-	23	10.8 10.8 10.7	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	pratense. Metsoon-	--	24. 10.3 10.7 10.5 86.59	4.5	5	24	12.9 13.4 13.2			24	12.9 13.4 13.2			24	12.9 13.4 13.2	24.0	-	24	12.9 13.4 13.2	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	Tatud vanale põllule	--	25. 12.2 12.8 12.5 122.7	4.5	2	25	8.4 8.4 8.3			25	8.4 8.4 8.3			25	8.4 8.4 8.4	24.0	-	25	8.4 8.4 8.4	24.5	-	0.0471	19.0	-	0.0471	19.0	-	0.0471	19.0	-
	1924. juuni tehtud kaun- 																													

1934 aasta mõõtmine 5/5.31.			1932 aasta mõõtmine 26. sept 1932						1933 aasta mõõtmine 11. sept 1933						1934 aasta mõõtmine 2.8.37					
Läbiröötlusid sentim.	Tähendused		Puu nr. nr.	Läbiröötlusid sentim.		Tähendused		Puu nr. nr.	Läbiröötlusid sentim.		Tähendused		Puu nr. nr.	Läbiröötlusid sentim.		Tähendused				
				NS	WO				NS	WO				NS	WO					
12.2	12.2	12.2	1	12.2	12.5	12.4	1.6	Paude avv	Lehtlae pind	1	12.4	12.6	12.5	4	- 18.2 m.	1	12.7	12.8	12.7	
2.0	2	3.8	2	2.0	2	1.0038		Mä	- 20.0 m.					2						
8.0	2	10.1	2	8.0	2	0.0101	2.							2	Mä 1 15	15.2				
8.5	4	22.7	3	8.5	4	0.0227								3			- 25	19.3		
9.0	13	8.27	3	9.0	11	0.0700	3.							4	14.3 14.8 14.5	-	30	15.7		
9.5	5	3.57	4	9.5	4	0.0284	4	11.2	14.5	14.3	- 41	- 19.6 m.		5	23.1 24.1 23.6	-	34	20.0		
10.0	11	8.64	5	10.0	11	0.0864	5	22.6	23.8	23.2	56.	- 18.1 m.		6		-	43	17.8		
10.5	9	7.79	5	10.5	9	0.0779								7	11.9	11.4	11.7	-	52	19.6
11.0	3	2.95	6	11.0	3	0.0285	6.							8		-	56	19.2		
11.5	5	5.20		11.5	5	0.0520								9	22.9	23.1	23.0	-	57	14.5
12.0	7	7.92		12.0	4	0.0452								10		-	95	20.0		
12.5	7	8.59		12.5	9	0.1104	7.	11.9	11.4	11.7	82	- 19.1 m.								
13.0	5	6.64	7	13.0	4	0.0531														
13.5	6	8.59	8	13.5	7	0.1001														
14.0	5	7.70	8	14.0	4	0.0816	8.	0.500												
14.5	2	3.30	9	14.5	2	0.0330	9.	22.8	22.3	22.3	106.	- 18.1 m.								
15.0	3	5.30	9	15.0	3	0.0530														
15.5	1	1.79		15.5	1	0.0189	10.													
16.0	7	14.08	10	16.0	6	0.1202														
16.5	5	10.69		16.5	6	0.1283														
17.0	7	15.89	11	17.0	1	0.0227	11.	16.3	15.8	16.1	140	- 18.2 m.								
17.5	3	7.22		17.5	2	0.1684														
18.0	1	2.55	12	18.0	2	0.0504	12.	13.3	14.6	13.9										
18.5	1	2.69		18.5	3	0.0806														
19.0	4	11.94	13	19.5	2	0.0591	13.	13.9	14.5	14.2	Mä	Paude avv	Lehtlae pind	13	14.6	14.8	14.7	14.0	1.0050	
19.5	4	12.67		19.5	2	0.0591								14						
20.0	1	3.30	14	20.0	5	0.1621	14.	0.500												
20.5	1	7.1	17.1	20.5	1	0.0330														
21.0	4	13.86		21.0	2	0.0693														
21.5	1	3.63	15	21.5	4	0.1452	15.	10.4	10.4	10.4	9.5	1	0.0071	15	10.6	10.4	10.5	10.0	0.0142	
22.0	1	3.80	16	22.5	4	0.1090														
22.5	4	15.90	16	22.5	1	0.0434	16.	20.4	20.6	20.5	10.5	5	0.0433	16	20.6	21.2	20.9	11.0	0.0095	
23.0	2	9.05	17	23.0	/	0.0452														
23.5	2	9.01	17	23.5	/	0.0474	17.	0.475												
24.0	1	4.91	17	24.5	/	0.0474														
24.5	1	2.55	18	25.5	/	0.0511														
25.0	1	1.35	2.2158	19	19.3	20.1	19.7	131	2.2368	19.	19.6	20.2	19.9	13.5	5	0.0715	13.0	2	0.0265	
25.5	1	0.0164-14.5	19	12.4	12.2	12.3	14.8	19.	12.4	12.2	12.3	14.0	3	0.0462	19	12.5	12.3	12.4	13.5	0.0429
26.0	1	10.95	19.5	26.0	0.403	-		14.5	4	0.0660				20		14.0	5	0.0769		
26.5	9	9.5	9.5	26.5	21			15.0	1	0.0177						14.5	4	0.0661		
27.0	1	7.2	17.3	27.0	22	17.5	17.4	17.5	17.0	3	0.0651				21		15.0	2	0.0353	
27.5	1	17.9	19.2	27.5	23	18.3	19.4	18.8	18.0	4	0.1018				22	17.6	17.5	17.5	16.5	0.0641
28.0	1	10.3	10.7	28.0	24	0.556	-	19.0	2	0.0567	24				23		16.0	2	0.0402	
28.5	1	15.15	13.9	28.5	25	13.5	13.9	13.7	20.5	4	0.1320	25	13.5	14.1	13.8	19.5	1	0.0298		
29.0	1	2.1	17.3	29.0	26			21.0	2	0.0693					26		20.0	2	0.0628	
29.5	1	17.4	17.1	29.5	27	18.4	17.5	17.2	23.0	4	0.1520	27	17.3	17.8	17.5	21.5	1	0.1650		
30.0	1	15.9	16.1	30.0	28	16.0	16.3	16.2	24.0	2	0.0905	28	16.4	16.6	16.5	23.0	2	0.0538		
30.5	1	15.1	15.1	30.5	29	15.3	15.4	15.3	25.0	1	0.0491	29	15.8	15.6	15.7	24.0	2	0.0567		
31.0	1	11.6	11.7	31.0	30	11.6	11.8	11.7	26.0	1	0.0531	30	11.7	11.8	11.8	25.5	1	0.0511		
31.5	1	11.1	11.1	31.5	31	0.585	-							31				95	2.0078	

What is data?

- In a way, almost everything!
- Data can be in analog and digital form





Practical information for the course

- I will set up a Github repository where I share info, slides and enable also the discussions to have interactive questions and answers.
- Please sign up for a Github account, if you don't already have one.



What we're gonna learn

- What is Big Data?
 - Basic concepts, definitions
- Origins of data
- Edge computing – Cloud computing
- Data storage and data formats
- Analysis and visualisation of data
- Data and machine learning
- Data and FAIR principles

What is Big Data?

Some definitions

- Data that are too large or complex to be dealt with by **traditional** data analysis software! (source Wikipedia)

Does this definition makes sense?



What is the **traditional** data analysis (software)?

- If you start to search for that, you do not find too much information's
- Some refer to the type of storage, like relational databases.
- The storage is in most cases thought to be forever, at least for very long time.

In a way, there is not a very clear distinction!

Big data

- From the search in internet (Wikipedia, different companies,...) its mostly defined via the size of data!
- But there are other topics either, the difference between **structured — unstructured** data
- The speed of data production
- The time of data storage (data lifetime)

The better definition may be that these data need some large infrastructure!