

# Big Data, organisation and analysis

Steffen M. Noe, Spring 2025

# What is data?

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



111p

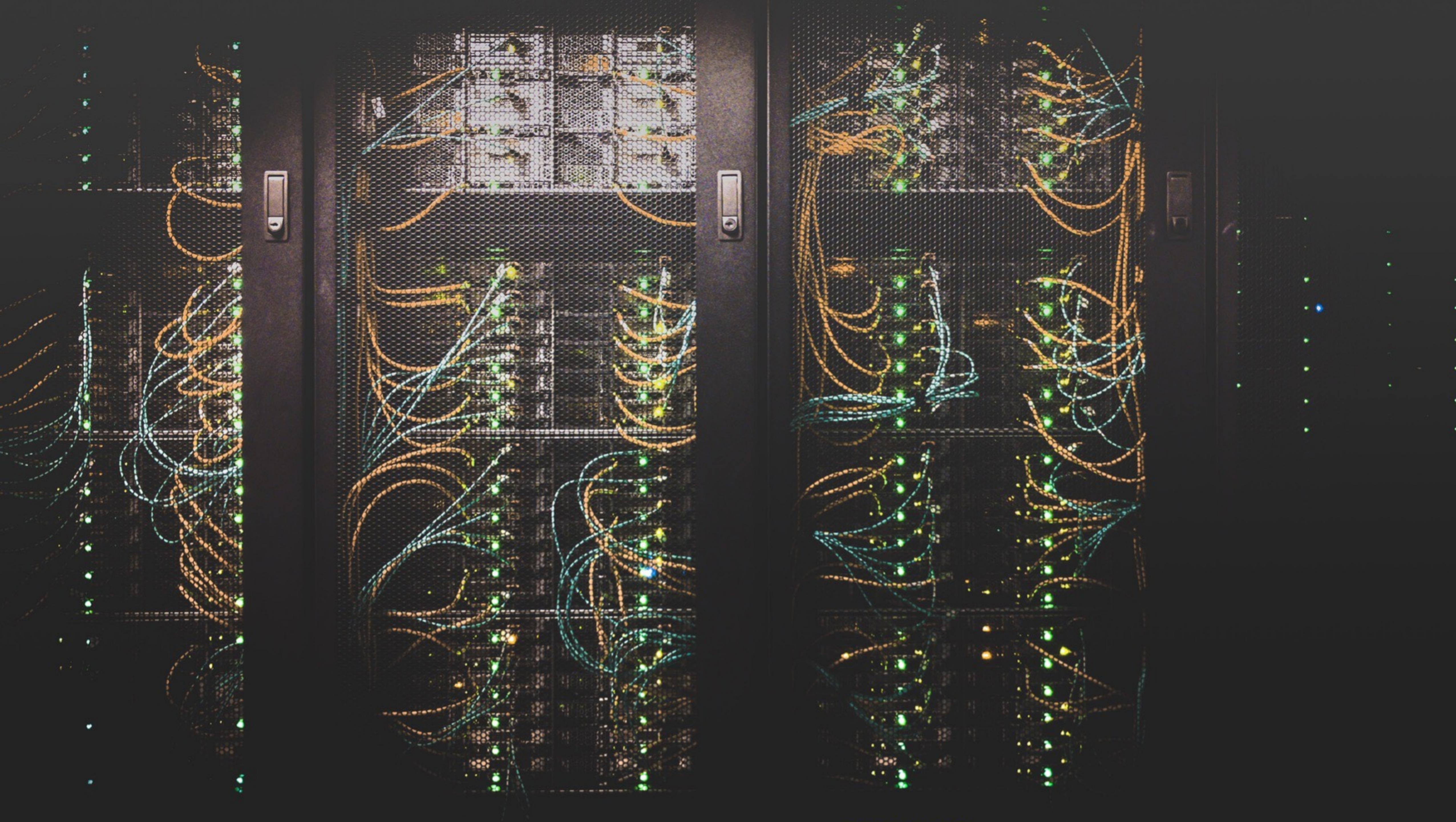
Proovi nr. nr. lo asu- rus	Vahikond kvart. nr. nr. Proovi tiki asenuse skeem Muld Maapilma kate	Puu seis. Boniteet	1926 aasta mõõtmine			1929 aasta mõõtmine 9.11			1930 aasta mõõtmine 17.10.30						
			Lähimöör pind cm²			Tähendused			Lähimöör sentim.			Mistänu sel koor natusarane nurkne Tähendused			
			Puu nr. nr.	NS	WO	Kesk.	NS	WO	Kesk.	NS	WO	Kesk.			
1. Kuraasare vhtk.	Nä 1	11.3.11.4.11.3.100.3	H	16.0	13.5	1	119	119	11.9	Mä	Puude avv	1. 12.2.12.2.12.2. Mä	Lähimöör pind		
0.06ha	NV. 280.	--	2. 11.5.11.2.11.3.100.3	8	16.5	12.5	2. 113	116	11.4	7.0	1	2. 11.8.12.0.11.9	7.5	1. 0.0444	
	--	3. 7.5.7.5.7.5.44.18	13	15.0	13.0	3. 1-13.3.0.0.510	8.5	7	8.0	3	0.0151	3. keivavund	8.5	5. 0.0284	
	--	4. 13.2.13.8.13.5.143.1	26	16.0	12.5	4. 13.6.14.0.13.8	10.0	12	9.0	13	0.0827	5. 0.0354	9.0	11. 0.0864	
	--	5. 19.6.20.6.20.1.347.3	29	15.5	13.7	5. 20.8.21.4.21.1	11.0	9	10.0	11	0.0779	5. 0.0519	10.0	11. 0.0779	
	--	6. 15.1.14.8.14.9.174.4	41	17.0	14.0	6. 15.9.15.3.15.6	12.0	11	11.0	9	0.0779	6. 0.1244	12.0	11. 0.1244	
	--	7. 11.8.11.3.11.5.103.9	56	16.5	13.5	7. 11.8.11.5.11.7	13.0	7	12.0	4	0.0491	7. 13.3.12.8.13.0	13.0	4. 0.0531	
	--	8. 12.8.12.3.12.5.122.7	69	15.0	13.0	8. 12.3.13.1.13.0	14.0	4	13.0	7	0.0402	9. 22.0.22.6.22.3	14.5	5. 0.0769	
	--	9. 20.0.20.2.20.1.317.3	82	16.5	13.5	9. 21.4.21.8.21.6	15.0	6	14.0	1	0.0165	10. 10.3.10.1.10.2	15.0	3. 0.0530	
	--	10. 10.1.9.9.10.0.78.54	91	16.0	13.3	10. 10.3.10.2.10.3	16.0	5	15.0	6	0.1132	11. 15.6.15.5.15.6	16.5	5. 0.1005	
	--	11. 14.4.14.0.14.2.158.4	106	15.5	13.8	11. 15.6.14.7.15.1	17.0	5	17.0	4	0.0855	12. 13.0.14.2.13.6	17.5	5. 0.1135	
	--	12. 12.1.13.1.12.6.124.7	127	16.5	14.0	13. 12.6.13.9.13.3	18.0	1	19.0	2	0.0589	13. 13.0.14.2.13.8	18.5	3. 0.0862	
	--	13. 12.4.13.6.13.0.132.7	135	16.0	13.5	13. 13.0.13.9.13.4	19.0	1	19.0	2	0.0567	14. 7.0.7.0.7.1.7.1	20.0	4. 0.0567	
	--	14. 7.0.7.2.7.1.39.59	140	15.5	13.4	14. 7.0.7.1.7.1	20.0	1	19.5	4	0.1194	15. 10.4.10.6.10.6	20.5	3. 0.0943	
	--	15. 10.2.11.5.16.3.83.32	152	16.0	13.5	15. 10.4.10.7.10.5	21.0	1	20.0	3	0.0943	16. 19.8.20.0.20.0	20.4	4. 0.1732	
	--	16. 17.7.18.3.18.0.254.5				16. 18.8.19.1.19.0	22.5	1	21.0	4	0.1732	17. 19.0.20.1.19.7	21.5	5. 0.1140	
	--	17. 8.9.9.0.8.9.62.31	Mä			17. 9.1.9.1.9.1	23.0	2	22.0	3	0.0348	18. 9.1.9.1.9.1	24.0	1. 0.0416	
	--	18. 17.4.17.9.17.6.243.3				18. 18.3.18.7.18.5	24.0	1	23.0	1	0.0434	19. 9.0.19.8.19.7	24.5	1. 0.0452	
	N. rihm. Õ-hanadu	--	19. 11.5.11.4.11.4.102.1			19. 12.1.12.0.12.0			19. 12.3.12.1.12.1	25.0	1	20.0	1	0.0471	
	Klar. Murometja	--	20. 9.4.9.2.9.3.67.93			20. 9.5.9.5.9.5			20. 9.5.9.7.9.5	140	2	22.374 m²	21. 1.13.0.0.426		
	järvelkassi puudutaval	--	21. 9.1.9.1.9.1.65.04			21. keivavund			21. 1.13.0.0.426			22. 17.0.17.0.17.0	- d = 14.38m.		
	Pindlate Hypnum	--	22. 16.1.16.4.16.2.206.1			22. 16.6.16.9.16.8			23. 17.6.17.9.18.2			24. 10.5.16.9.10.7			
	Lycopodium annotatum	--	23. 16.3.17.4.16.8.221.7			23. 17.0.18.3.17.6			24. 10.5.16.9.10.7			25. 13.3.13.7.13.5			
	Pinola scandula, Pedi-	--	24. 10.3.10.7.10.5.86.59			24. 10.4.10.8.10.6			25. 12.3.13.7.13.5			26. 13.3.13.7.13.5			
	? uniflora, Equisetum	--	25. 12.2.12.8.12.5.122.7			25. 12.9.13.4.13.2			26. 8.4.8.3.8.3			27. 13.5.13.9.13.7			
	pratense, Melica	--	26. 8.4.8.2.8.3.54.11			26. 8.4.8.3.8.3			27. 10.8.11.2.11.0			28. 12.1.12.0.12.0			
	tatud hanalepille	--	27. 15.8.16.2.16.0.208.1			27. 16.3.17.1.16.7			28. 15.6.15.7.15.6			29. 12.6.12.4.12.1			
	juuris lehti lau-	--	28. N.0.14.1.14.0.153.0			28. 14.9.14.9.14.9			29. 14.9.15.0.14.9			30. 11.6.11.7.11.6			
	ante ei oleks, dominante	--	29. 13.8.13.6.13.7.147.4			30. 11.5.11.7.11.6			31. 10.8.11.9.11.0			32. 11.6.11.7.11.6			
	raunakarate Hypnum	--	30. 11.3.11.5.11.4.102.1			31. 10.9.11.3.11.0			31. 10.9.11.3.11.0			32. 11.6.11.7.11.6			
	schrubri, Hypnum ercta	--													
	caeruleum, Hypnum pro-	--													

1927 aasta mõõtmine 5. XI. 31.			1928 aasta mõõtmine 26. sept. 1928			1929 aasta mõõtmine 26. sept. 1929			1930 aasta mõõtmine 11. okt. 1930			Lähimöör sentim.			Lähimöör sentim.		
Lähimöör sentim.			Tähendused			Lähimöör sentim.			Tähendused			Lähimöör sentim.			Lähimöör sentim.		
Puu nr. nr.	NS	WO	Kesk.	Puu nr. nr.	NS	WO	Kesk.	Puu nr. nr.	NS	WO	Kesk.	Puu nr. nr.	NS	WO	Kesk.		
1.2.12.2.12.2.	12.5	Puude avv	Lähimöör pind	1. 12.2.12.2.12.2. Mä	4. - 18.2 m.	1. 12.4.12.6.12.5	4. - 18.2 m.	1. 12.7.12.8.12.7	4. - 18.2 m.	1. 12.4.12.6.12.5	4. - 18.2 m.	1. 12.7.12.8.12.7	4. - 18.2 m.	1. 12.4.12.6.12.5	4. - 18.2 m.	1. 12.7.12.8.12.7	4. - 18.2 m.
1.2.12.2.12.2.	12.0	1. 12.2.12.2.12.2. Mä	4. - 18.2 m.	2. 12.0	1. 12.0	1. 12.0											

# What is data?

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





# Practical information for the course

- I have set up a Github repository where I share info, slides and enable also the discussions to have interactive questions and answers. (<https://github.com/steno/BDOA/tree/main>)
- 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 visualization of data
- Data and machine learning (ML/AI)
- 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, or at least for very long time.

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

# Big data

- From the search in internet (Wikipedia, LLMs, 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!