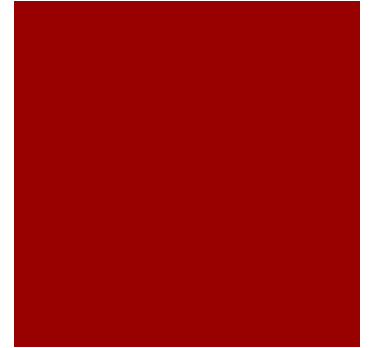




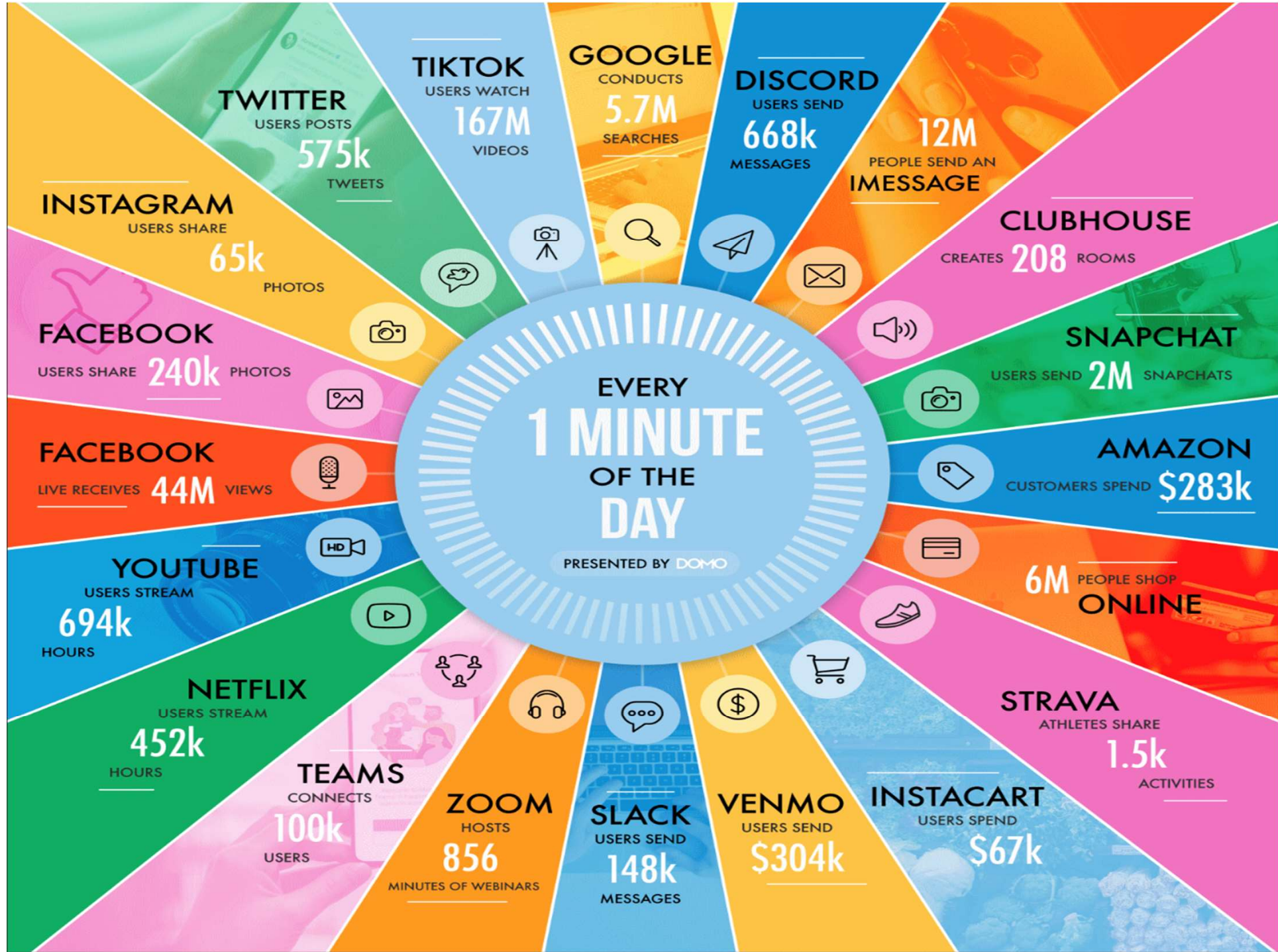
De Informatiemaatschappij van 2025

Dr. G.J. Van Busseel

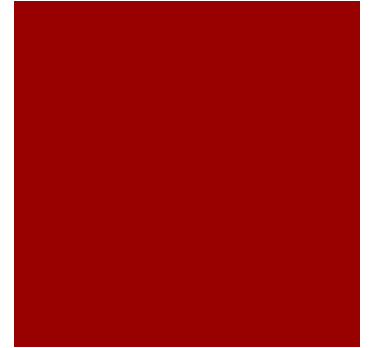


- Archivaris, Bedrijfskundige, Bestuurlijk Informatiekundige
- Directeur Van Busseel Document Services
- Adviseur Strategisch Informatiemanagement
- Auditor NEN 2082, NEN-ISO 15489, ISO 16175, RODIN, ED3
- Oud-lector Digital Archiving & Compliance (HvA)
- Visiting Professor bij universiteiten in België, Finland, VS
- Auteur van 250 boeken en artikelen





Zettabyte scale



- IDC (2021)
- *Worldwide Global Data-sphere Forecast, 2021–2025. The World Keeps Creating More Data. Now, What Do We Do With It All?*
- 2018: 33 zettabytes; 2025: 181 zettabytes

Unit	Value	Example
Kilobytes (KB)	1,000 bytes	a paragraph of a text document
Megabytes (MB)	1,000 Kilobytes	a small novel
Gigabytes (GB)	1,000 Megabytes	Beethoven's 5th Symphony
Terabytes (TB)	1,000 Gigabytes	all the X-rays in a large hospital
Petabytes (PB)	1,000 Terabytes	half the contents of all US academic research libraries
Exabytes (EB)	1,000 Petabytes	about one fifth of the words people have ever spoken
Zettabytes (ZB)	1,000 Exabytes	as much information as there are grains of sand on all the world's beaches
Yottabytes (YB)	1,000 Zettabytes	as much information as there are atoms in 7,000 human bodies

What is Big Data?

Big Data often indicates large data sets that are too complex for traditional data analysis but may reveal insights, trends, and patterns when analyzed.

Big Data was initially characterized by its three Vs, namely Volume, Velocity, and Variety.

Volume

Simply refers to the amount of data available. Big Data comes in large sizes, global big data is somewhere around 60 Zettabytes.

Velocity

Refers to the speed at which data is produced, made available, can be processed, and used. Big Data is produced at staggering speeds and requires novel techniques to be processed.

Variety

Big Data comes in all shapes and sizes and can be structured, unstructured, and semi-structured. It also comes in various formats and forms which can be extremely difficult to organize.

Later on more Vs were added to offer a clearer description of Big Data.

Variability

Variability refers to the degree upto which the data changes. It focuses on understanding the true meaning of the data.

Veracity

Refers to the data's trustworthiness. Not all data is legit or accurate and with Big Data one must be careful if the data can be trusted.




Visualization

Can the data be visually represented in understandable formats?
Visualization is all about presenting data in ways that are most meaningful.

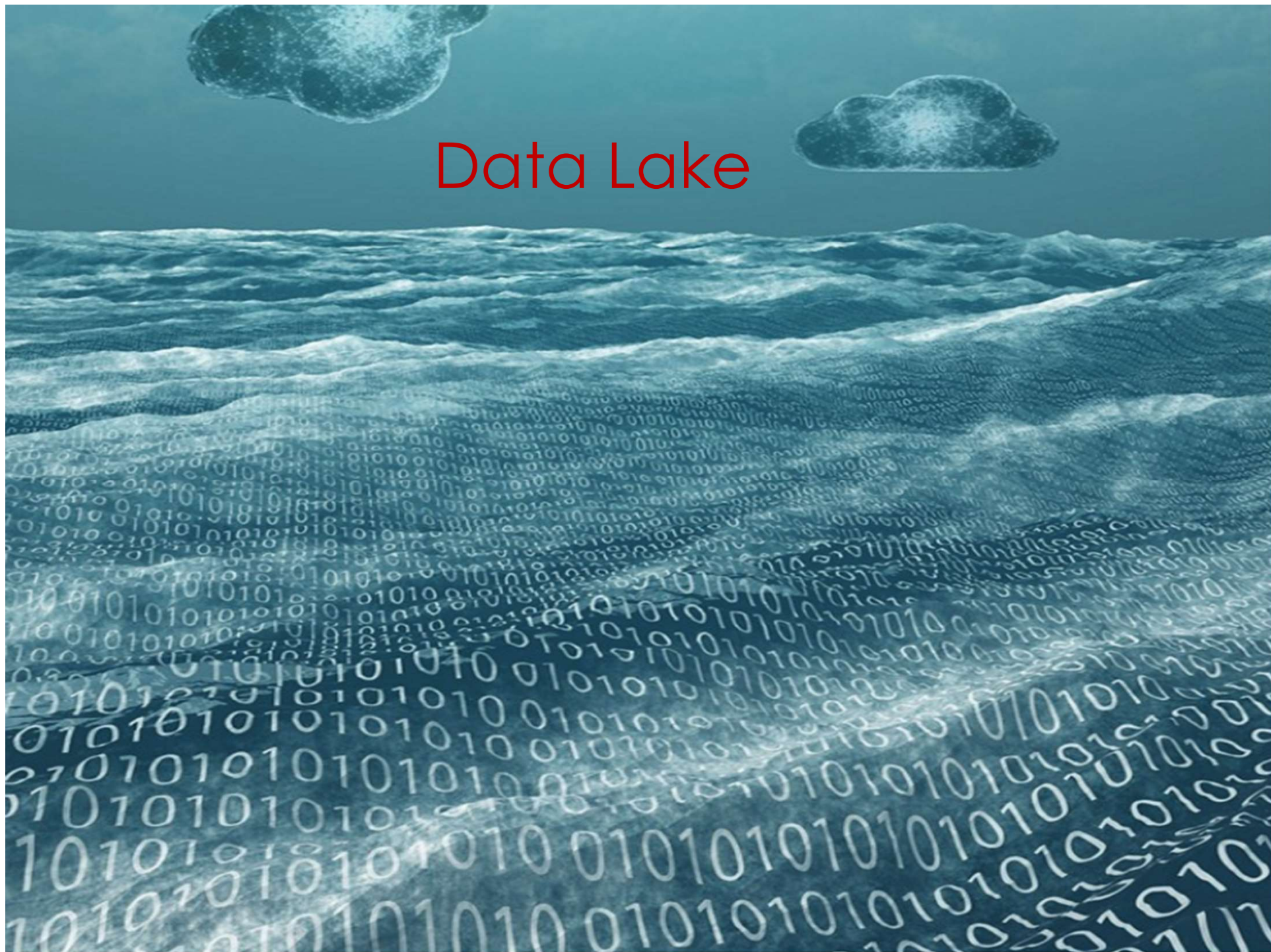
Value

Simply indicates the business value of the data in question.

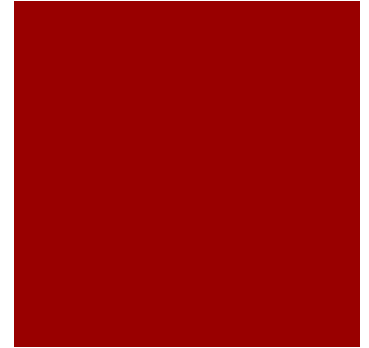
Data Sizes Comparison Table

	Examples	Characteristics	Typical tools	Analytical methods
 Small Data (megabytes)	Sales records, Customers database (small and medium companies)	Hundreds – thousands of records	Personal computer, Excel, R, other basic statistics software	Simple statistics
 Large Data (gigabytes - terabytes)	Customer databases (big companies)	Millions of records, mostly structured data	Server workstation computer, Relational database systems, data warehouses	Advanced statistics, business intelligence, data mining,
 Big Data (terabytes – petabytes)	Customer interactions (social media, mobile), multimedia (video, images, free text), location-based data, RFIM	Over millions of records, distributed, unstructured	Cloud, data centers, Distributed databases, NoSQL, Hadoop	MapReduce, Distributed File Systems

Data Lake

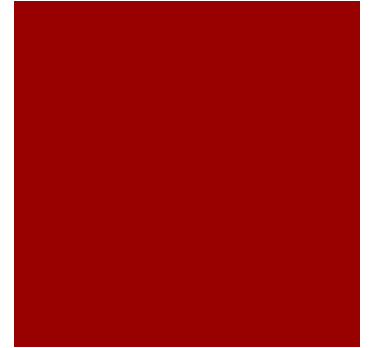


Een paar voorbeelden



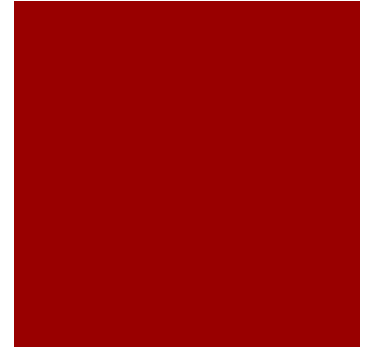
- In dertig minuten vliegtijd produceert een enkele vliegtuigmotor tien terabyte data. Meerdere motoren per vliegtuig en duizenden vluchten per dag levert meerdere petabytes aan data op.
- Concreet: Transavia heeft vier retourvluchten per dag naar Faro (Portugal) Totale datageneratie per retourvlucht is 480 terabyte. Vier retourvluchten per dag leveren meer dan anderhalve petabyte aan data op.

Een paar voorbeelden



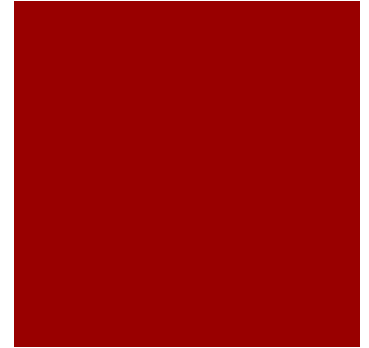
- Het Rubin Observatorium maakt foto's van de zuidelijke sterrenhemel via een grote ruimtetelescoop in Chili. Het produceert vijftien terabyte aan data per nacht. In de tien jaar die het project duurt wordt in totaal tweehonderd petabyte geproduceerd.

Een paar voorbeelden



- Een jaar na de start van Google Photo's (in 2016 dus) had deze dienst 13,7 petabytes aan fotomateriaal opgeslagen.
- In 2020 worden vier biljard foto's opgeslagen, waarvan de omvang niet is gespecificeerd, maar waarschijnlijk meer dan een exabyte vergt.

Een paar voorbeelden

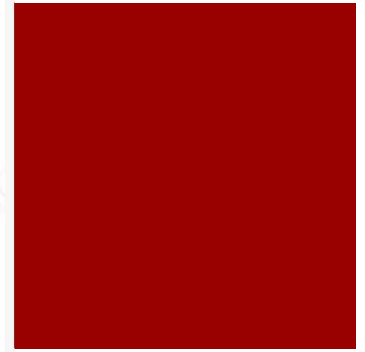


- CERN slaat 340 petabytes data via haar tape-robot-systeem. Dit is gelijk aan tweeduizend jaar HD video. In 2025 is die opslag aan data gegroeid tot een exabyte.
- En zo zijn er nog veel meer voorbeelden van gigantische hoeveelheden data die worden opgeslagen als gevolg van bedrijfsactiviteiten.
- Heel veel data en datavormen dus!
- Het lijkt op Big Data, maar is het dat ook echt?

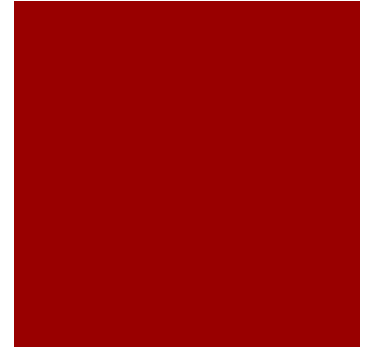


Adobe Stock | #74249044

**“Let’s shrink Big Data into Small Data ...
and hope it magically becomes Great Data.”**



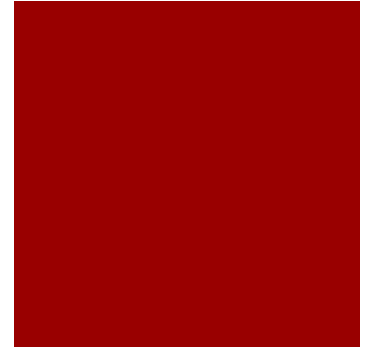
Trend naar large en small data



- LEGO: rond 2000 concludeerde LEGO op basis van big data analyses dat de jongere generatie snel wilde worden bevredigd en niet lang ergens mee bezig wilde zijn. Dus: vergroot de blokken. De generatie bleek echter de blokken niet te willen hebben en vroeg om de oude, kleinere blokken. LEGO was aan de rand van faillissement. Die informatie had LEGO beschikbaar uit klantonderzoeken, maar ze vertrouwde volledig op de voorspellingen uit Big Data.
- Google Flu: de poging van Google om vanaf plm. 2003 tot 2015 griepgolven te voorspellen op basis van zoekgedrag. In geen enkele van de jaren bleken de algoritmes van Google het goed te voorspellen.



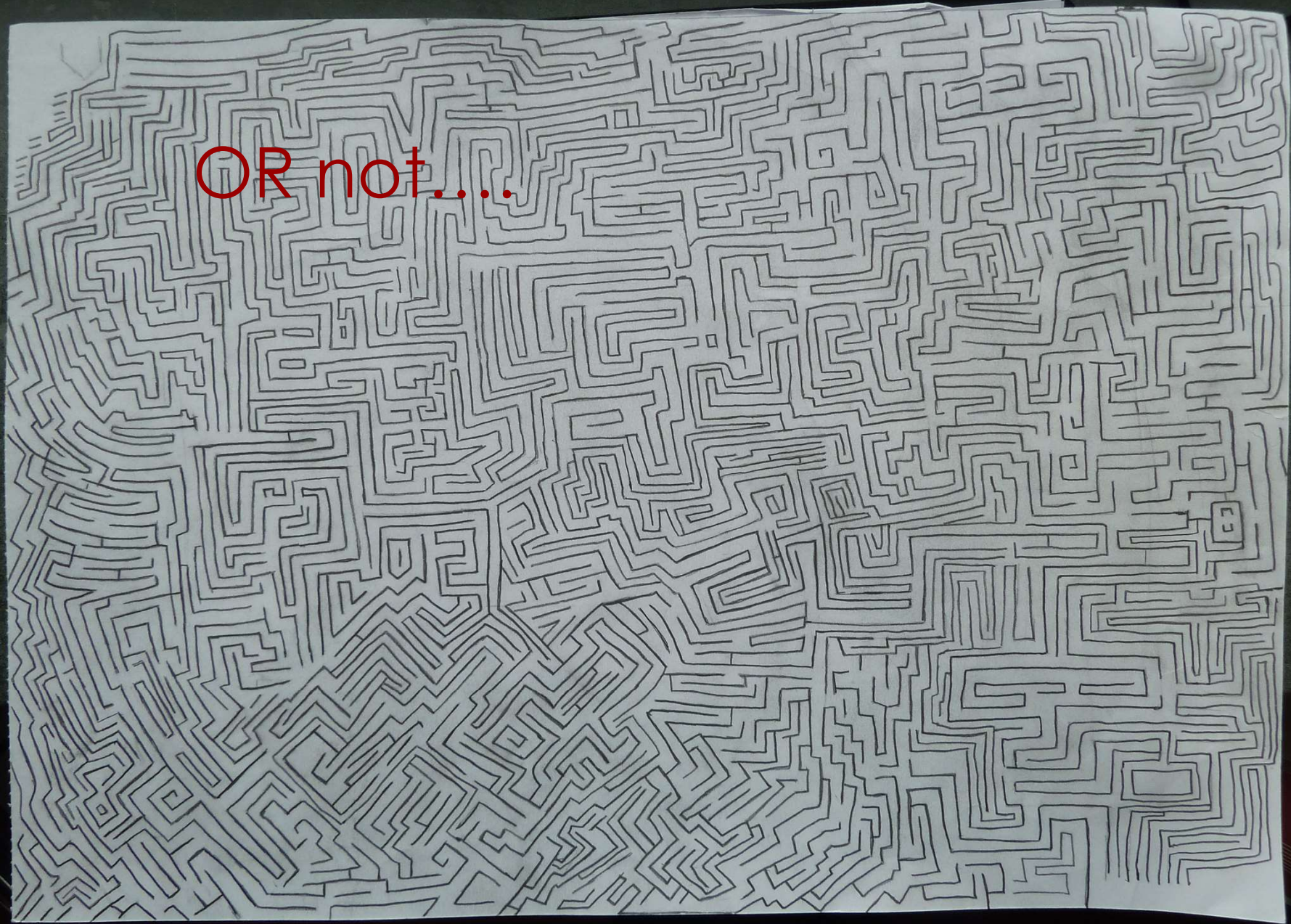
Searching the information haystack



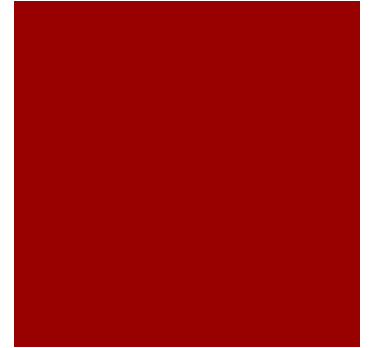
To find the relevant
needle...



OR not....

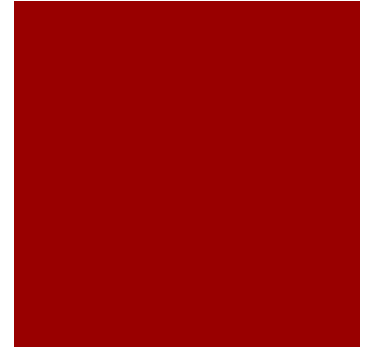


Informatiegedrag



- De effecten van gedrag in organisaties op informatiemanagement zijn al lang bekend.
- Enkele aangetoonde voorbeelden:
- De tendens 'to screen, withhold, modify, or otherwise manipulate upward communication flows, based on employee needs.'
- Het manipuleren van de toegankelijkheid door informatie te 'verbergen' of verkeerd te metadateren.

Informatiegedrag



- Doelbewust vertragen van het overdragen of doorzetten van informatie.
- Doelbewust verkeerde informatie over een collega of manager creëren.
- 'They keep quiet when asked questions, keep knowledge and records confidential, manipulate shared records, do not file records within the organizational archive but use personal storage instead, 'steal' records when leaving, process new records carelessly, or/and destroy records unauthorized.'



OPGERICHT IN HET OORLOGSJAAR 1940 - 70STE JAARGANG NR 20.212

ICT-chaos bedreigt uitkeringen

■ Wethouder Wiebes waarschuwt voor gevolgen gammele automatisering ■ D66 dringt aan op raadsonderzoek

HUGO LOCTENBERG
BAS SOETENHORST

AMSTERDAM - De automatisering van de gemeente Amsterdam is zo gammel, dat chaos dreigt bij het verstrekken van uitkeringen door de sociale dienst en bij aangiftes bij de belastingdienst.

Wethouder Eric Wiebes heeft de ge-

meenteraad vorige week in een besloten vergadering ingelicht over de problemen.

Zo kan het overmaken van uitkeringen in het honderd lopen, waardoor duizenden mensen in problemen komen. Ook bij de belastingdienst en afgifte van vergunningen kan het misgaan. Bij storingen kunnen gegevens verdwijnen door het ontbreken van backups, met administratieve chaos tot gevolg.

Ambtenaren gebruiken gmail om gemeentelijke server te ontwijken

Een ander gevaar vormen de vele datacenters die niet voldoen aan normen voor klimaatbeheer en veiligheid. Een stad als Amsterdam zou kunnen volstaan met twee tot drie datacenters, maar heeft er liefst

vijftig. In een groot aantal loopt de temperatuur in de zomer soms op tot veertig graden en gaan systemen plat. Ambtenaren sturen hun mail vaak via gmail of hotmail om de trage gemeentelijke servers te ontwijken, die er soms uren over doen.

De ict-kosten van de gemeente bellopen 130 miljoen euro per jaar, negenduizend euro per ambtenaar. Andere grote gemeenten doen het met vijfduizend euro. Om de kosten

omlaag te brengen, zijn eerst forse investeringen nodig. De grootste oppositiepartij, D66, wil een raadsonderzoek om boven tafel te krijgen hoe de ict zo kon ontsporen. Raadslid Carlien Roodink: "Er ligt een plan om liefst 100 miljoen te investeren, alleen om puin te ruimen. Voordat we dat doen, moeten we lessen trekken uit het verleden."

■ Pagina 5: Erger dan de NZ-lijn

FOKKE & SUKKE

GAAN ELKE ZONDAG DE KROEG IN

VROEGER WAREN WE OP
MAANDAG ALTIJD "ZIEK"....

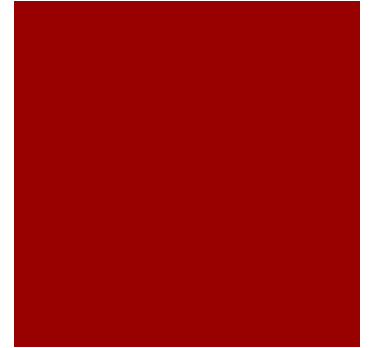
MAAR NU HEET HET
"HET NIEUWE WERKEN"



RGvT

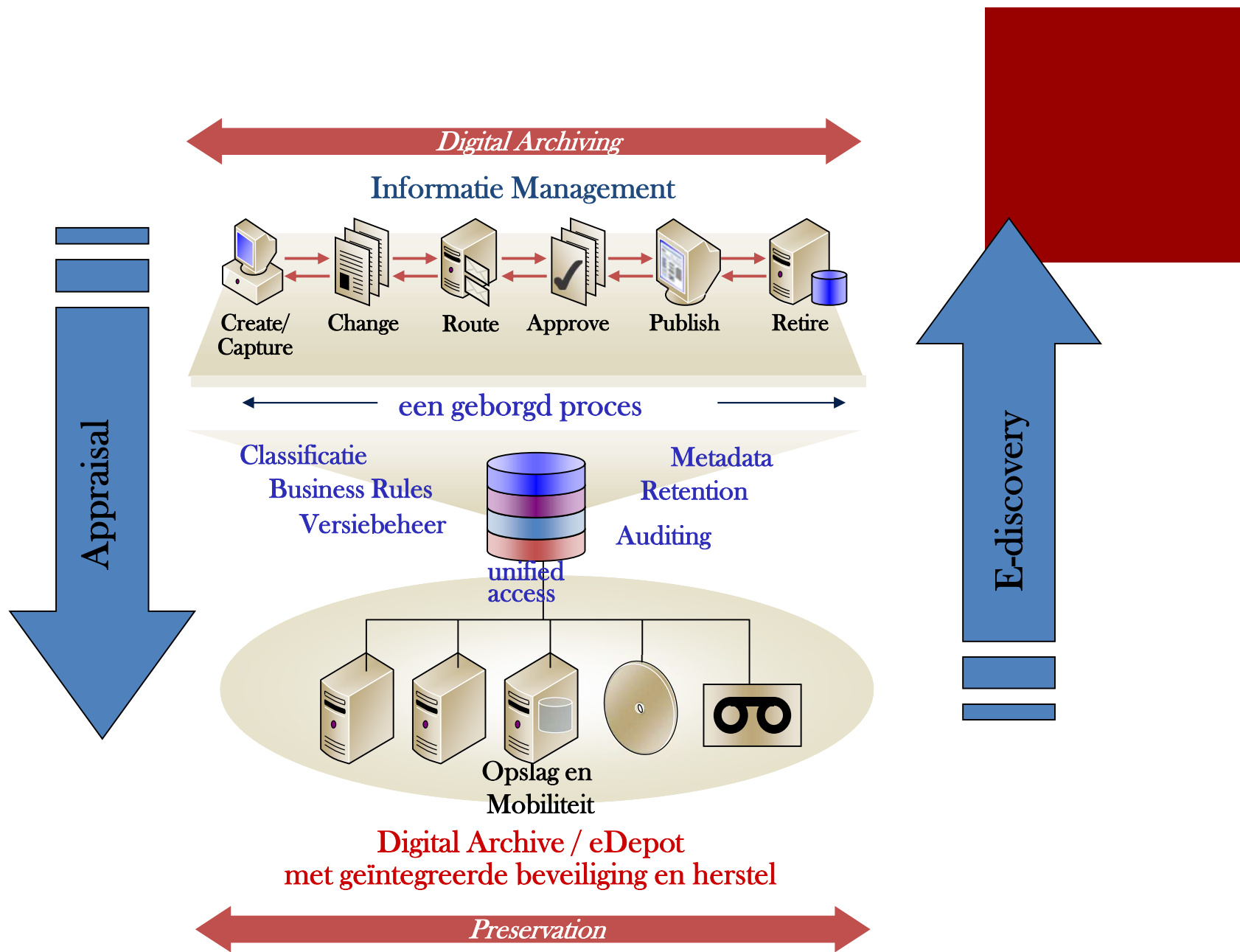
www.foksuk.nl

Naast gedrag



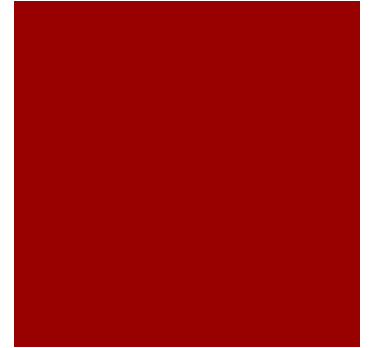
- Hebben we nog wat andere uitdagingen:
 - PAAC
 - AIAAC
 - MLAAC
 - AAAC
 - BCAAC
 - IAAC
 - CGAAC
 - DPAAC







Drew McDermott

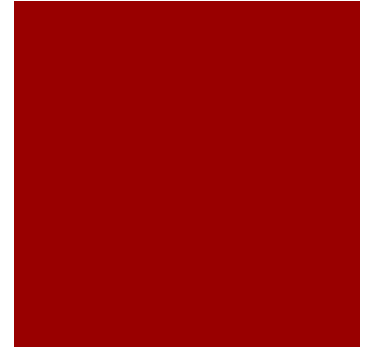


Artificial Intelligence

Doesn't mean you are allowed
to show

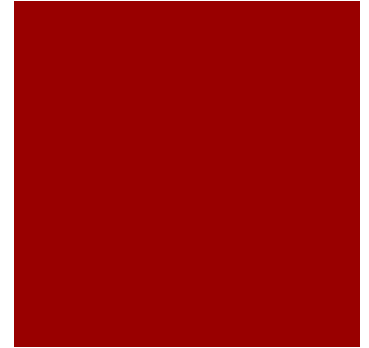
Natural Stupidity!

Janelle Shane



The danger of AI
is not that it will revolt
against its creators,
but that it exactly does
what we want it to do

Leesvoer



- G.J. van Bussel (2020). *A Sound of Silence. Organizational Behaviour and Enterprise Information Management. Papers on Information and Archival Studies, I*, Van Bussel Document Services, Helmond. xii + 109 pp. ISBN/EAN 978-90-831078-0-6.
- G.J. van Bussel (2021). *An Accountability Puzzle. Organizations, Organizational Governance and Accountability. Papers on Information and Archival Studies, II*, Van Bussel Document Services, Helmond, x + 168 pp. ISBN/EAN 978-90-831078-1-3.
- G.J. van Bussel (2023). *The Accessibility Paradoxes. Accessibility of Information and its Requirements. Papers on Information and Archival Studies, III*, Van Bussel Document Services: Helmond.



Dank U

Van Busseel Document
Services

Hof Bruheze 421

5704 NR Helmond

www.vbds.nl