

Den vanskelige reisen til datadrevet virksomhet

Pål de Vibe

Fagsjef for Data Engineering and AI, Knowit Objectnet

knowit

—



“for var det noko han ikkje likte, så var det store ord,
dei løynde og tildekte berre, dei stora orda,
dei lét ikkje det som var
få vere og leve,
men tok det bort inn i noko som stort ville vere”

— *Jon Fosse (Det er ales)*

- Data-drevet ved dekret?
- Trippelreisen
- Juridisk labyrint
- Flaskehalsen
- Demokratisere data-arbeidet
- Få med alle fra ende-til-ende, tidlig
- Data- og erfaringsdeling
- AI: revolusjon, tyranni, idiokrati?

Data-drevet ved dekret?

knowit



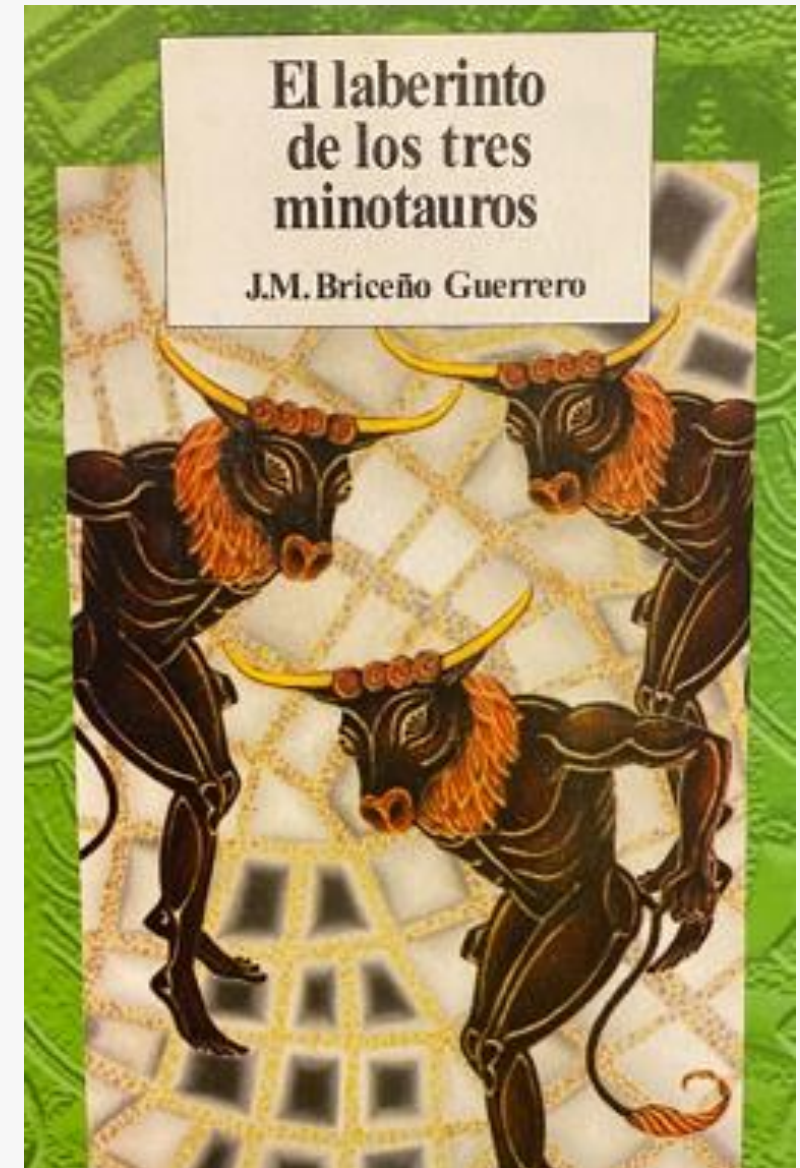
The commune is like a gigantic dragon, production is noticeable awe-inspiring. Wu Shaoyun (吴少云), Zhang Yuqing (章育青), Lu Zezhi (陆泽之)

Trippelreisen

“Det er lett å se at disse tre diskursene interpenetrerer, parasitterer og hindrer hverandre, i en tragisk kamp der ingen vinner.”

- JMB Guerrero

- Tre kompliserte tek/org-reiser
 - Sky
 - Dataplattform
 - Data mesh + data drevet
- Ytterligere utfordringer
 - Juridisk
 - AI



knowit

10/10/2016

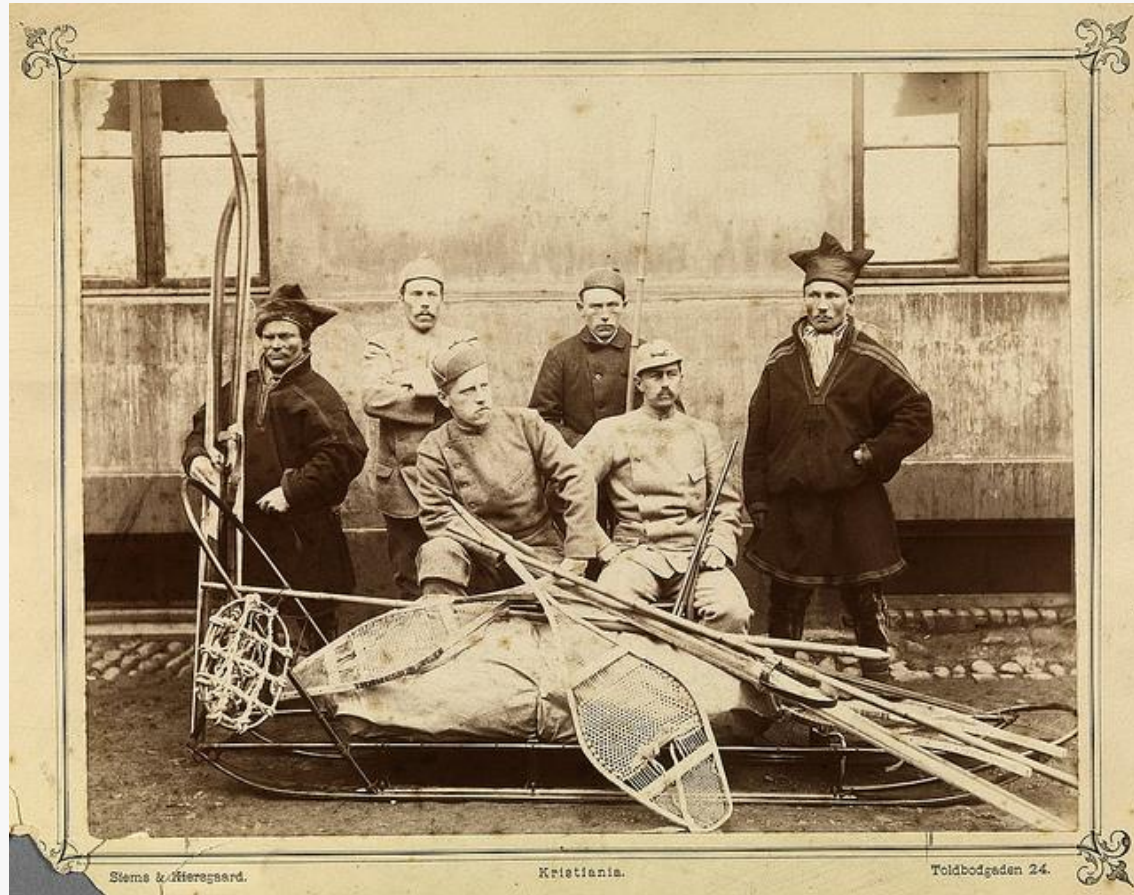
- [illegible]

Kartet laget av kartografen David Crantz i 1765. Foto: Crantz

Inn i det ukjente...

knowit
—

Hvorfor er det så vanskelig å bli data-drevet?



- Ydmykhet - Tidligere erfaring er ikke nok, snakk med folk som er lokalkjent
 - Fagkunnskap trengs på alle i nivåer for å redusere risiko
- Tidligere transformasjons- og digitaliseringsprosjekter ikke nødvendigvis helrelevante

Juridisk labyrint

- GDPR
- Personvern
- Er data egentlig anonymisert?
- Samtykke eller hjemmel?
- Kan vi bruke Sky?
- Enorm innsats er nødvendig!



HELSEDATA: Mariann Hornnes, direktør i E-helsedirektoratet, ser ikke noen annen mulighet enn å sette Helseanalyseplattformen på pause inntil de finner en løsning som sikrer at ikke nordmenns helsedata kommer på avveie.

Foto: Vidar Sandnes

Pauser Helseanalyseplattformen: – En svært vanskelig beslutning



Franz Kafka. Anon.
Lisens: Begrenset gjenbruk

Organisasjonens selvbedrag

- Er man egentlig i en tekk-organisasjon?
 - Man er sjelden Google/Spotify/Facebook
- Hvor mye tid/penger er man egentlig villig til å satse for å bli data-drevet?
- Vil vi få ressursene til å ha data-arbeidere i alle avdelinger?
- Kommuniser skånsomt, og bruk den luftige visjonen pragmatisk



Hjalmer og Hedvig Ekdal i Vildanden (1963). Foto: Norsk filminstitutt

- Data-drevet ved dekret?
- Trippelreisen
- Juridisk labyrint
- **Flaskehalsen**
- Demokratisere data-arbeidet
- Få med alle fra ende-til-ende, tidlig
- Data- og erfaringsdeling
- AI: revolusjon, tyranni, idiokrati?

Flaskehalsen - data breadline

knowit
—



Origin unknown

Flaskehalsen - data engineer overstretch

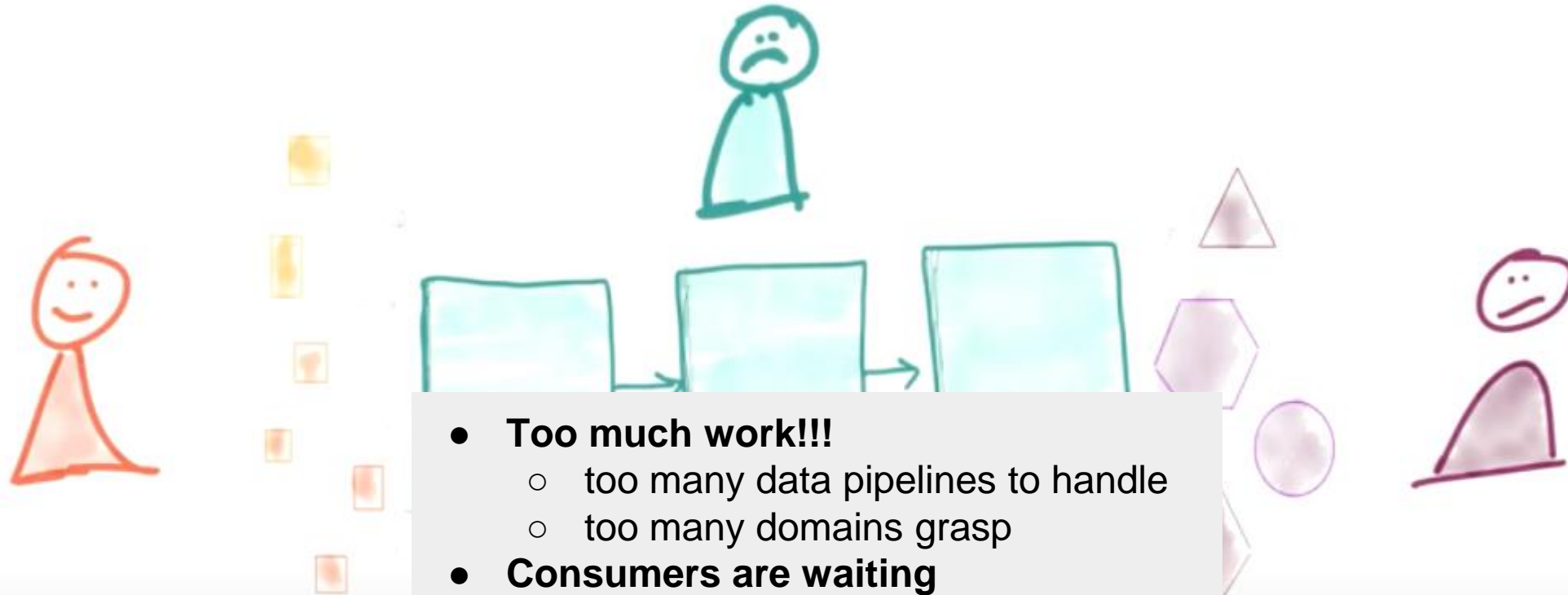
knowit



Licensed image from Alamy

Flaskehalsen

Data Platform Engineers



Domains' Operational systems Teams

Data Scientists, BI teams, ...

Data mesh

knowit
—

Data platform(s)
with orderly access
and decentralized
domain responsibility

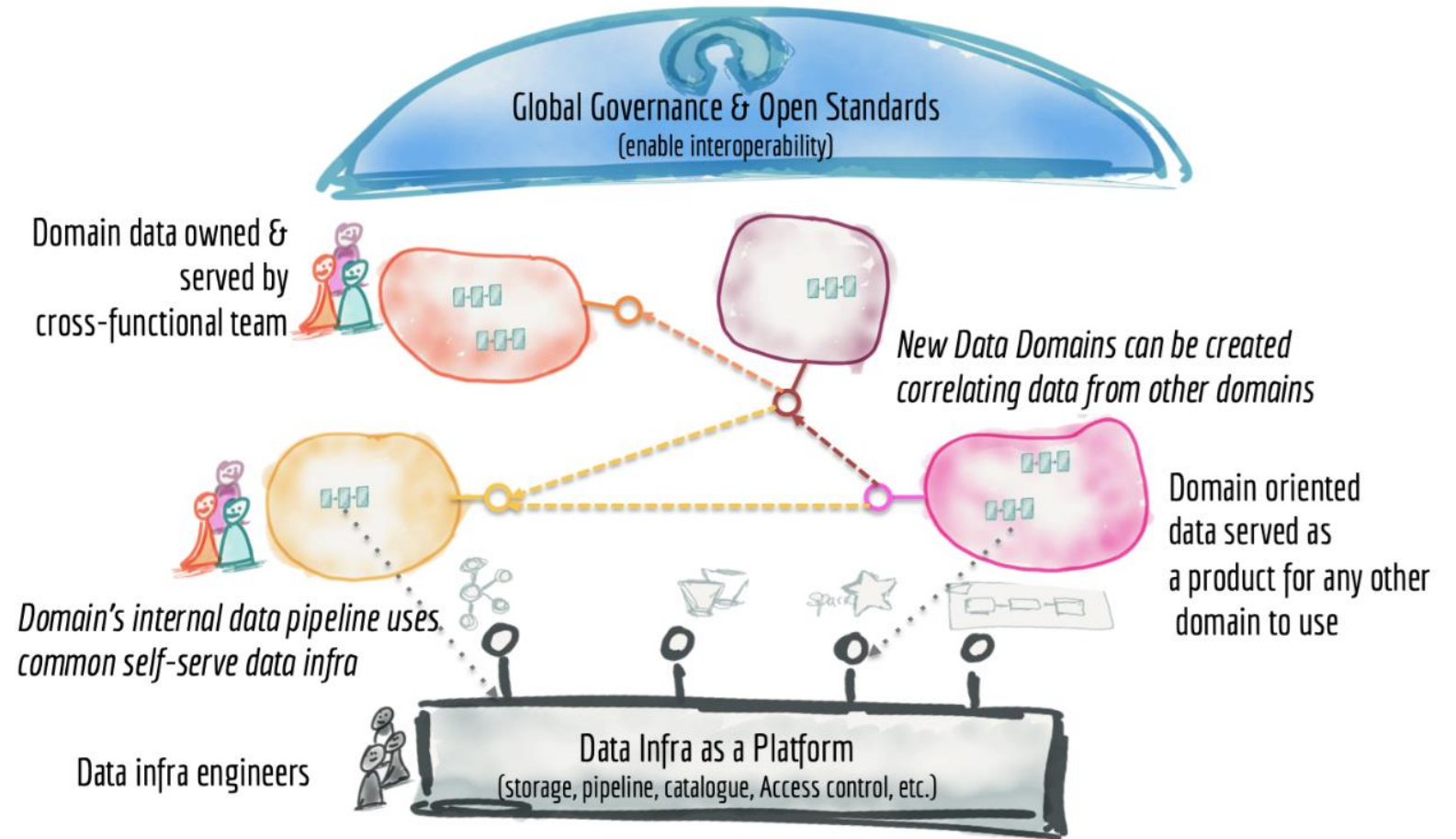


Figure 12: Data mesh architecture from 30,000 foot view

Source: <https://martinfowler.com/articles/data-monolith-to-mesh.html>

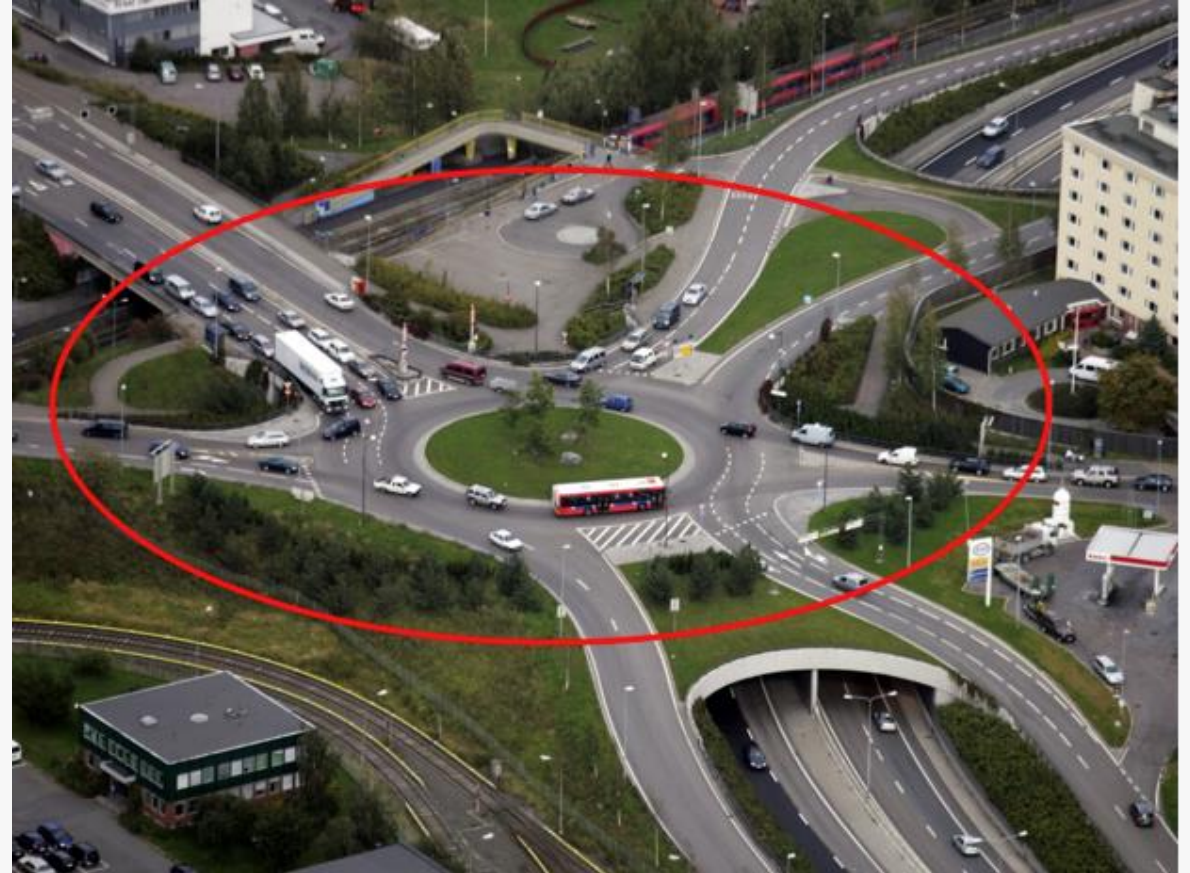
Inspired by micro-service architecture and domain thinking

Demokratisere data-arbeidet

knowit

Mix arbeidsformer med skyorkesteret

- ... som lar folk gjøre jobben selv
- Rigg for å mixe inn ny teknologi fortløpende
- Fra Data Engineer to Data Platform Engineer
- Kjenn publikumet
- Eksempler på demokratiserende teknologi
 - DBT
 - Notebook-sentrisk data-arbeid
 - LLM-støttehjul
 - Dataspøringer i menneskelig språk



- Data-drevet ved dekret?
- Trippelreisen
- Juridisk labyrint
- Flaskehalsen
- Demokratisere data-arbeidet
- **Få med alle fra ende-til-ende, tidlig**
- Data- og erfaringsdeling
- AI: revolusjon, tyranni, idiokrati?

Få med alle fra ende-til-ende tidlig

Involvér brukerne

- tidlig
- sette produsenter og konsumenter sammen
- ...før plattformen er klar
- gå gjennom bruk på "papir"

slik finner man bedre retning, og reduserer store risikoer



A scientific team conducts a survey of an evergreen forest in Yoleko, DRC, by James P. Blair

Del erfaringer og data

Organisasjoner støter på lignende utfordringer

- Juss, datakatalog, datadeling, data mesh, dataprodukt eierskap, sky
- Tverrsektorielt datasamarbeid for samferdsel
 - <https://samferdselsdata.no/>
- Tilrettelegger for enklere datadeling i samferdselssektoren
- Juridisk samarbeid
- Konkrete datadelingsinitiativer
- Nasjonal infrastruktur for datadeling
- <https://offentlig-paas.no> fagdag/chat for erfaringsdeling - **suksess!**



KYSTVERKET



Statens vegvesen



Jernbane-
direktoratet



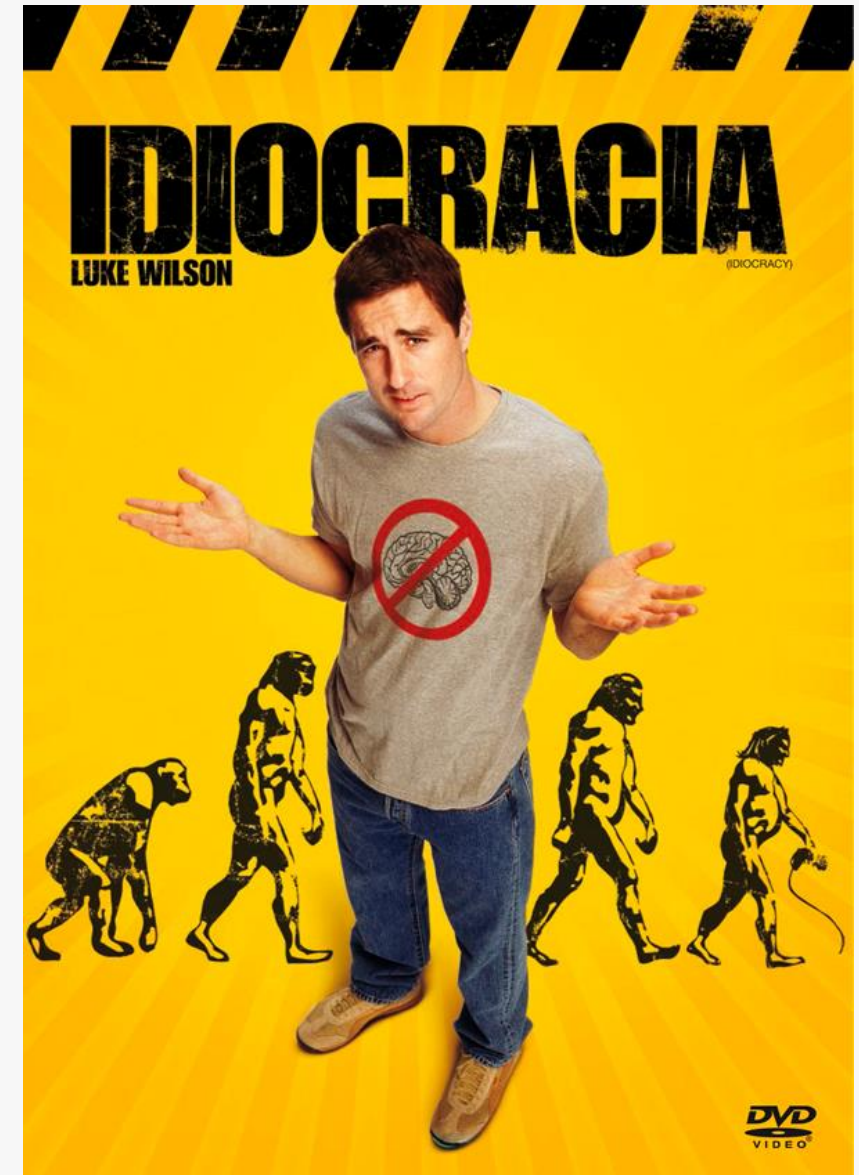
Pilot med DigDir

- Gjøre data.norge.no til interaktiv datakatalog
- Skyporten (Maskinporten) - Del skyressurser til et orgnr
https://docs.digdir.no/docs/Maskinporten/maskinporten_skyporten
- Forenklet onboarding til Maskinporten

AI: revolusjon, tyranni, idiokrati?

knowit

- Digital desensitization
 - Prøv å fortell et eventyr til dagens barn
- "GPT har svaret"
 - Stol på Boten, ikke mennesket
- "Idiotisering", ref politikk i USA
- Disintermediation
 - Video-media blir lavterskel
 - AI "stammifiserer" oss mer
 - American Culture Wars
 - China velger å avkreve samhold, f.eks. aktivt målrettet sensur på Tiktok, som optimiserer for antatt bærekraftige verdier
 - Tiktok gjør det motsatte i vesten
- Demokrati eller pøbelvelde?



AI: revolusjon, tyranni, idiokrati?

knowit

Vi vet hva vi frigjøres fra...

men ikke hvem sin makt vi faller i.

Ingen kjenner utfallet.

Lovgivning av design-intensjoner er fåfengt for verktøy som essensielt er fleksible og åpent rettet.



Karmelitt-martyrene i Compiègne

AI: revolusjon, tyranni, idiokrati?

Objektiv menneskelig moral er nødvendig

"This type of AI, called a large language model, has no single intended use" *ChatGPT broke the EU plan to regulate AI*

Is it possible to define risky with a **flexible intelligence** without talking **objective morals and values**?

- China legislates with explicit values:

"content generated...shall reflect the **Socialist Core Values**, and may not contain: **subversion** of state power; overturning of the socialist system; incitement of separatism; harm to national unity; propagation of terrorism or extremism; propagation of **ethnic hatred** or ethnic discrimination; **violent, obscene, or sexual information**; false information; as well as content that may **upset economic order or social order**."

- West will drift towards explicit values enforcement, even if veiled:

"build AI for the **public good**,... what surrounding legal infrastructure, economic relationships, and social norms are necessary to build (and **enforce**) such containers?" *AI Institute Whitepaper*
...but the **challenge** for the West is **worse** because **dismantling objective values is a core value**, ref Karl Popper's Paradox of Tolerance.

knowit



Den vanskelige reisen til datadrevet virksomhet

Sammendrag

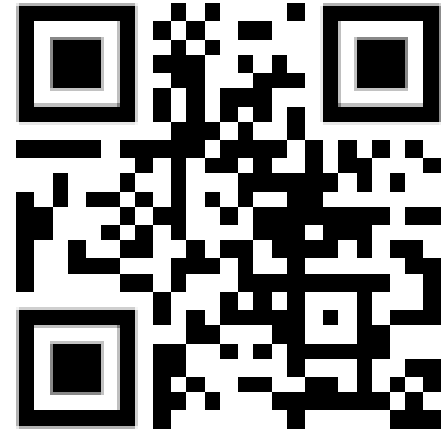
- Data-drevethet kan ikke oppnås ved dekret
- Søk råd på reisen, ta med deg lokalkjente
- Juss trenger enorm innsats
- Unngå data engineer-flaskehalsen
- Demokratisér dataarbeidet
 - Ansvarlig
 - Effektivt
 - Rigg for å ta del i kontinuerlig innovasjon
- Få med alle fra ende-til-ende tidlig
- Søk og del erfaringer med trippelreisen
- AI: Tribalism, dehumanisering. Objektiv moral nødvendig.

“Jeg hadde mindre og dårligere jord enn andre menn, men ble fortalt at etter at kornskatten var betalt, ville vi alle dele innhøstingen likt. Slik ville ingen tape noe, men alle ville vinne. Jeg syntes det høstes greit ut.”

- *En kinesisk bonde reflekterer over landbrukskooperativer*

This slidedeck

<https://tinyurl.com/datadrevet>



Technical resources for Data Mesh and Platform

- Databricks workshop material <https://github.com/microsoft/Azure-Databricks-NYC-Taxi-Workshop>
- Default domain repo for databricks https://github.com/paalvibe/databricks_domain_repo
- Unit-testing framework for py spark code <https://github.com/telia-oss/birgitta>
- Databricks Repos <https://docs.databricks.com/repos/index.html>
- DBT <https://www.getdbt.com/>

Articles

- Data platform urbanism <https://www.linkedin.com/pulse/data-platform-urbanism-sustainable-plans-your-work-p%25C3%25A5I-de-vibe/>
- You can't buy integration <https://martinfowler.com/articles/cant-buy-integration.html>
- Notebooks are The Missing Piece of The Big Data Revolution <https://medium.com/swlh/notebooks-are-the-missing-piece-of-the-big-data-revolution-2e86995446dc>
- Beyond Interactive: Notebook Innovation at Netflix <https://netflixtechblog.com/notebook-innovation-591ee3221233>
- <https://martinfowler.com/articles/data-monolith-to-mesh.html> Data monolith to mesh
- <https://martinfowler.com/articles/data-mesh-principles.html> Data Mesh Principles and Logical Architecture
- <https://maithraraghu.com/blog/2023/does-one-model-rule-them-all/> Does One Large Model Rule Them All? Predictions on the Future AI Ecosystem, Maithra Raghu (Samaya AI), Matei Zaharia (Databricks), Eric Schmidt (Schmidt Futures)
- <https://ai.objectives.institute/whitepaper> AI Institute Whitepaper
- <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/> ChatGPT broke the EU plan to regulate AI



Pål de Vibe
pal.de.vibe@knowit.no