



UiT Noregs arktiske universitet

## Organisering, styring og finansiering av infrastruktur for FAIR forskingsdata og særleg relevante forvaltingsdata for forskning:

*Kva er føresetnaden for å lykkast innanfor mitt område, og kva betydning og verdi vil det ha å lykkast for forskning?*

Innspel til Datainfrastrukturutvalet  
Møte 4, 19. november 2021

Philipp Konzett

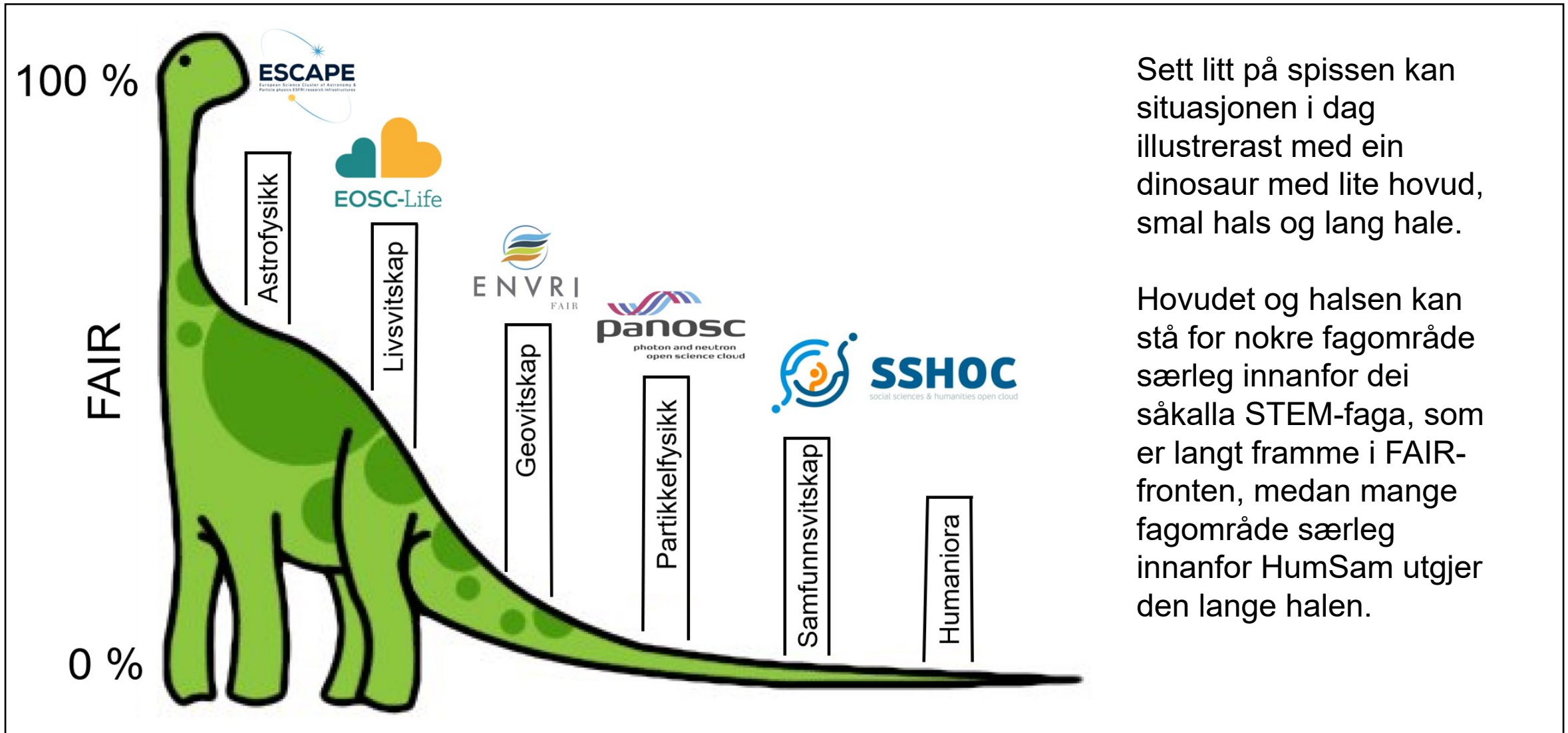
Universitetsbiblioteket  
UiT Noregs arktiske universitet



**2030:**

Eg vil gjerne starta med eit målbilete for FAIR data i Noreg i 2030, og det er at flest mogleg forskarar **har tilgang til** infrastrukturar som gjer det enkelt å gjenbruka andre sine data, og gjera eigne data **så FAIR som mogleg**, også **på tvers av fag**.

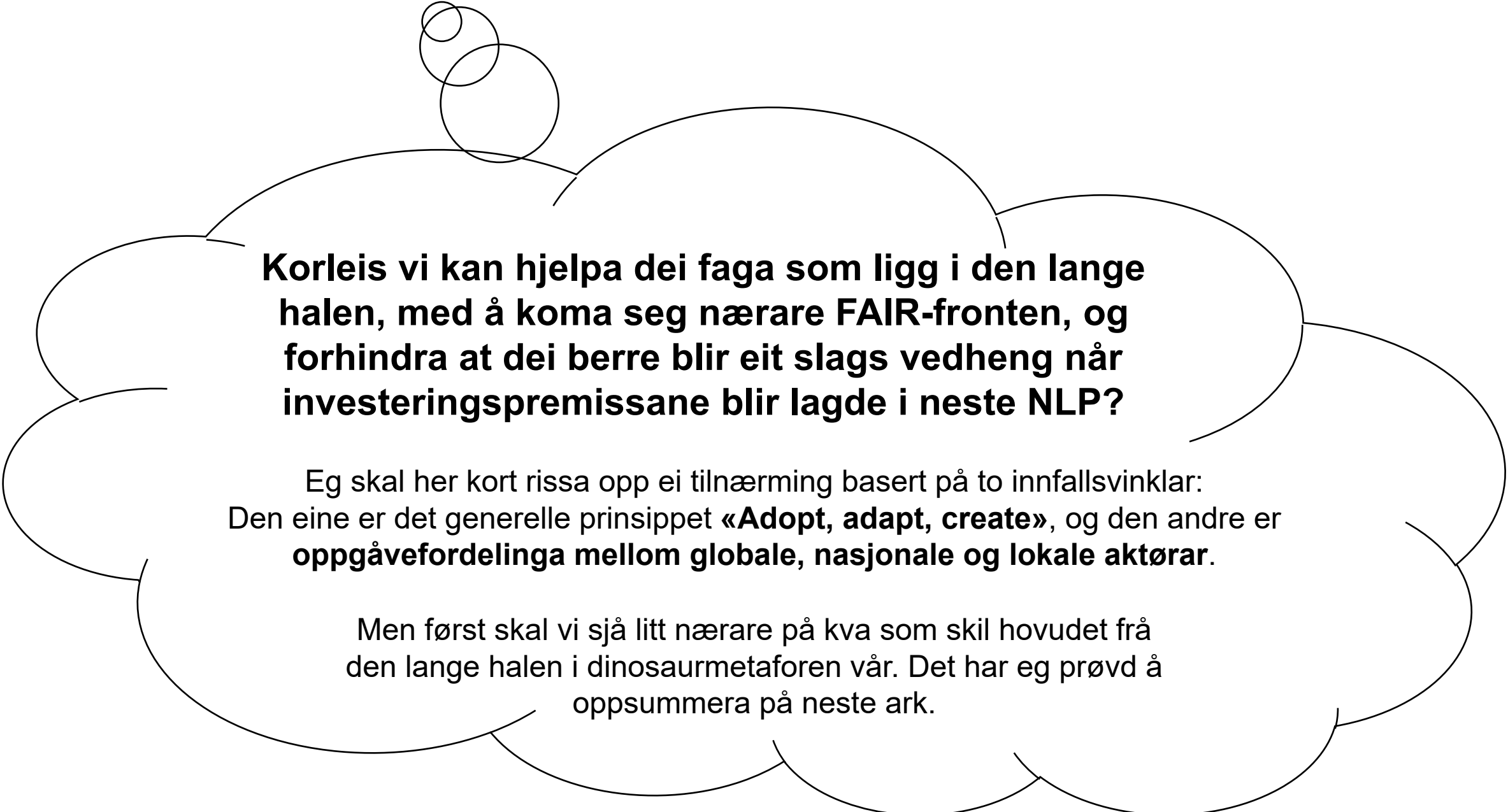
# I DAG:



Sett litt på spissen kan situasjonen i dag illustrerast med ein dinosaur med lite hovud, smal hals og lang hale.

Hovudet og halsen kan stå for nokre fagområde særleg innanfor dei såkalla STEM-faga, som er langt framme i FAIR-fronten, medan mange fagområde særleg innanfor HumSam utgjer den lange halen.

Tilpassa frå Shearer (2014), Burgelman (2015), Méndez et al. (2021)

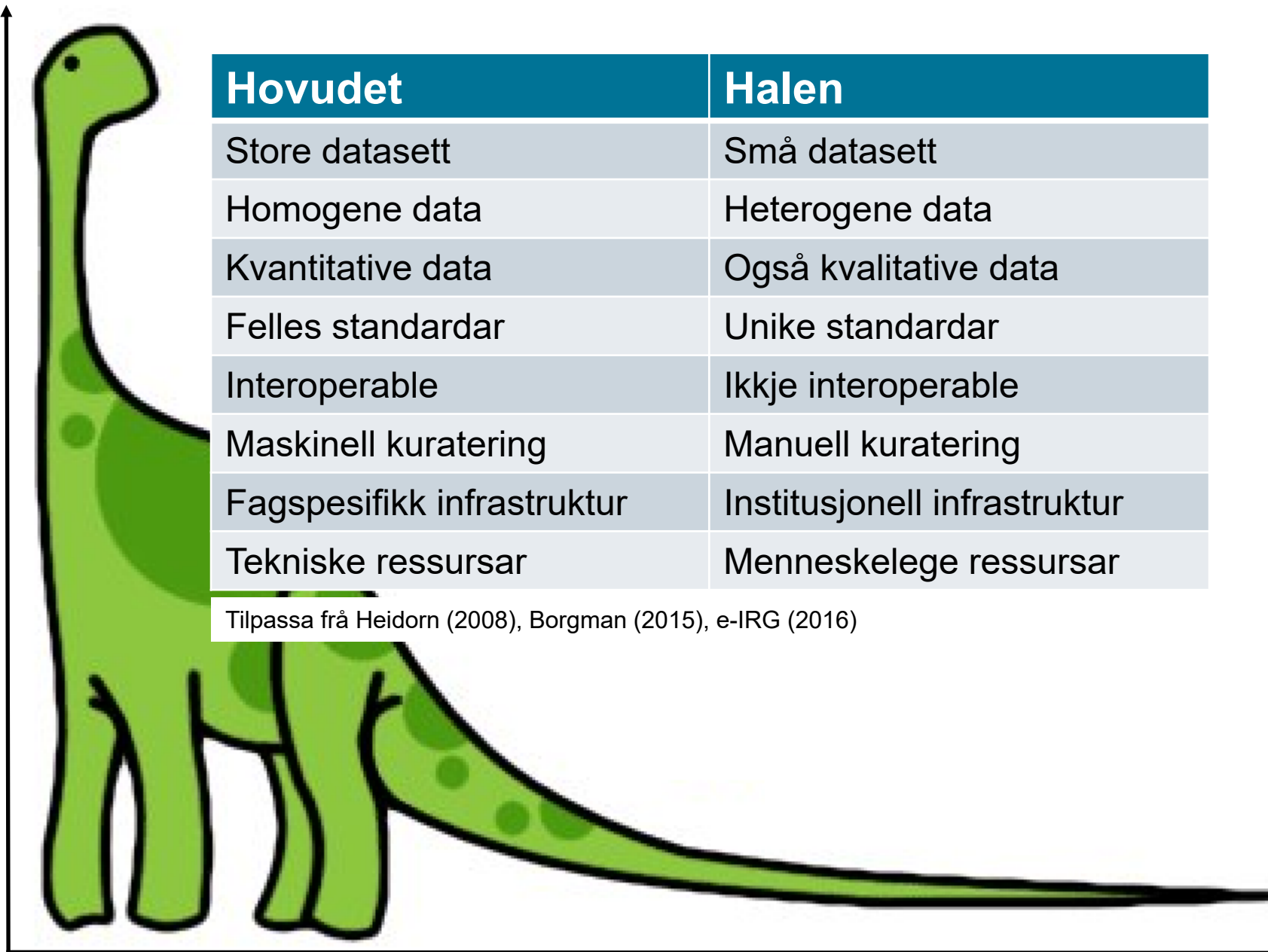


**Korleis vi kan hjelpa dei faga som ligg i den lange halen, med å koma seg nærare FAIR-fronten, og forhindra at dei berre blir eit slags vedheng når investeringspremissane blir lagde i neste NLP?**

Eg skal her kort rissa opp ei tilnærming basert på to innfallsvinklar: Den eine er det generelle prinsippet «**Adopt, adapt, create**», og den andre er **oppgåvefordelinga mellom globale, nasjonale og lokale aktørar.**

Men først skal vi sjå litt nærare på kva som skil hovudet frå den lange halen i dinosaurmetaforen vår. Det har eg prøvd å oppsummera på neste ark.

Dataintensitet



Tilpassa frå Heidorn (2008), Borgman (2015), e-IRG (2016)

Tal på forskingsprosjekt

Tendensen er at forskarar i FAIR-fronten gjerne opererer med store, homogene data som er basert på felles standardar, medan den lange halen består av små, heterogene datasett som er basert på unike eller ad hoc standardar.

Medan dataa i FAIR-fronten gjerne får støtte frå store internasjonale infrastrukturar, så er det som regel institusjonane som «sit att» med ansvaret for dataa frå den lange halen, og dei krev i mykje større grad menneskelege ressursar for å bli lagde til rette for FAIR deling og gjenbruk.

Spørsmålet er då korleis vi kan rigga til investeringane og organiseringa av infrastrukturar på ein måte som gjer at situasjonen i 2030 liknar meir på ein gris enn ein dinosaur. Og då er vi over på det siste lysarket.

Hovudbodskapen min – som illustrert på neste ark – er

1. at tekniske infrastrukturar, standardar og beste praksis i all hovudsak bør utviklast i internasjonalt samarbeid, og
2. at desse ressursane bør takast i bruk og dersom nødvendig tilpassast og driftast nasjonalt, og
3. at ein på lokalt plan først og fremst bør fokusera på forskarnære støttetjenester rundt desse infrastrukturane.

Det betyr at nasjonale investeringar i størst mogleg grad bør stø opp om internasjonale løysingar. Framfor å utvikla eigne system, bør ein gå inn i samarbeid med internasjonale, community-drivne miljø om å gjera eksisterande verktøy endå betre enn dei allereie er. Det vil frigjera ressursar til å gje lokale støttetjenester eit ordentleg løft. Det er gjennom dialog med forskingsmiljøa lokalt på institusjonane at ein kan skapa endringar for den lange halen. Derfor er lokalt støttepersonell ein heilt essensiell del av den nasjonale forskingsinfrastrukturen. Takk for merksemda!

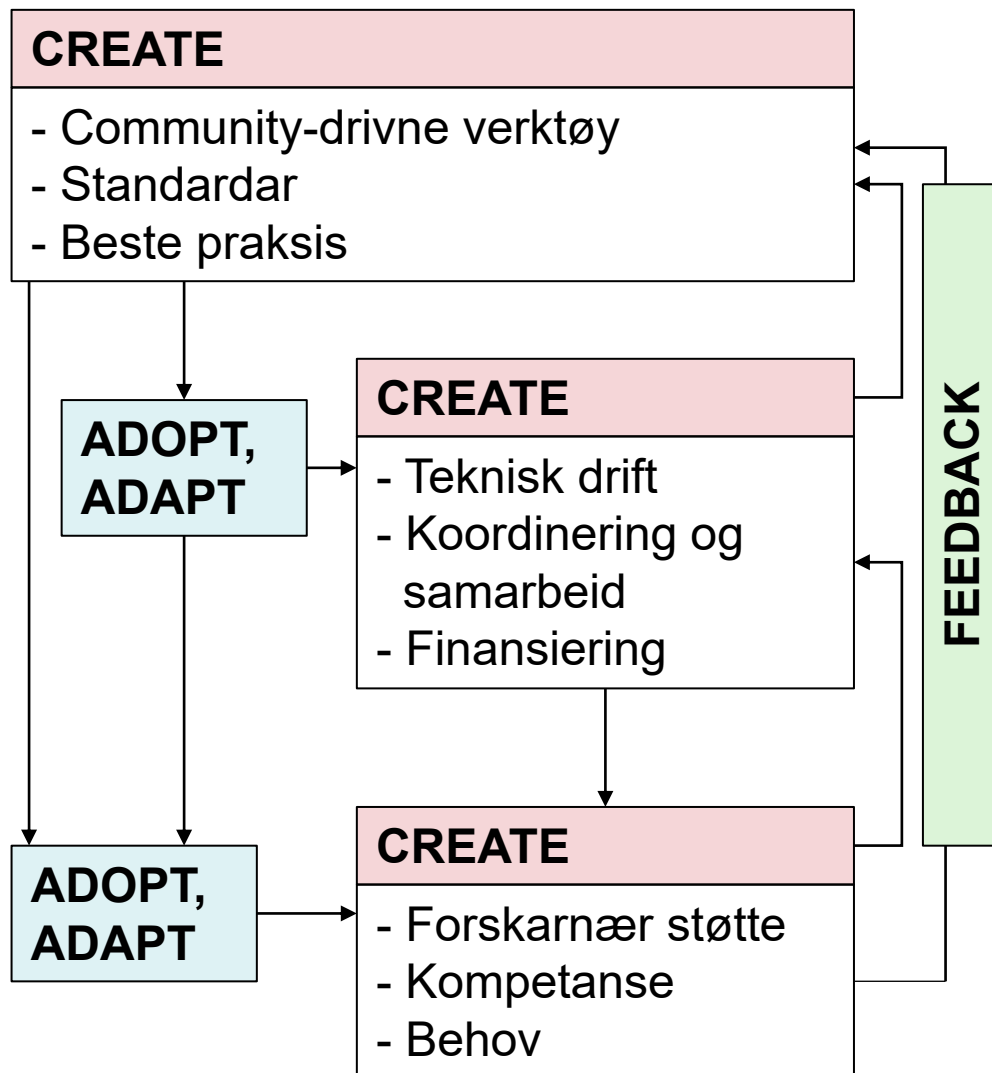
# GLOBALT



# NASJONALT

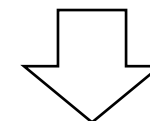


# LOKALT



## Not invented here:

«The panel also issued a plea for more convergence and a move away from the ‘not invented here’ syndrome which leads to much replication and redevelopment of similar services in the Open Science field.» (Jones et al. 2021)



## Adopt, adapt, create:

- **Adopt** existing, community-driven resources
- **Adapt** where necessary
- As a last resort, **create** new resources

+

<https://openscholarlyinfrastructure.org/>  
<https://investinopen.org/>  
<https://scoss.org/>

# Referansar

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press.

Burgelman, Jean-Claude. 2015. «Open Science Policy: Results of the Consultation on "Science 2.0: Science in Transition" and Possible Follow up». Presentert på e-IRG Workshop «Open Science and e-Infrastructures», Riga, 3. juni 2015. <https://e-irg.eu/documents/10920/288074/1.+J.C.+Burgelman.pdf>.

e-IRG. 2016. «Long Tail of Data». E-IRG Task Force Document. The Hague: e-IRG. <http://e-irg.eu/catalogue/eirg-1001>.

Heidorn, P. Bryan. 2008. «Shedding Light on the Dark Data in the Long Tail of Science». *Library Trends* 57, nr. 2 (2008): 280–99. <https://doi.org/10.1353/lib.0.0036>.

Jones, Sarah, Rudolf Dimper, Ingrid Dillo, Hilary Hanahoe, og Shalini Kurapati. 2021. «Making the European Open Science Cloud (EOSC) Work: Where to Go from Here?», 28. oktober 2021. <https://doi.org/10.5281/zenodo.5607692>.

Méndez, Eva, Joy Davidson, Zsolt Silberer, og Shelley Stall. 2021. «The Data You Document are the Data We Love». 5. november 2021. <https://doi.org/10.5281/zenodo.5651257>.

Shearer, Kathleen. 2014. «Wagging the Long Tail. Current Metadata Practices for Long Tail Research Data». Presentert på EUDAT 3rd Conference, Amsterdam, 25. september 2014. <https://www.eudat.eu/sites/default/files/KathleenShearer.pdf>.