

# Med både sejl og motor

---

Nye muligheder og udfordringer for  
færøsk sprogteknologi

Iben Nyholm Debess

6. nov 2024

Trondheim



# Færøsk sprogteknologi

---

- Status og udvikling
- Udfordringer og muligheder
- Strategier for navigation
  - Lokale initiativer
  - Inklusion af færøsk på store platforme
  - Nordisk samarbejde



# Færøsk sprog

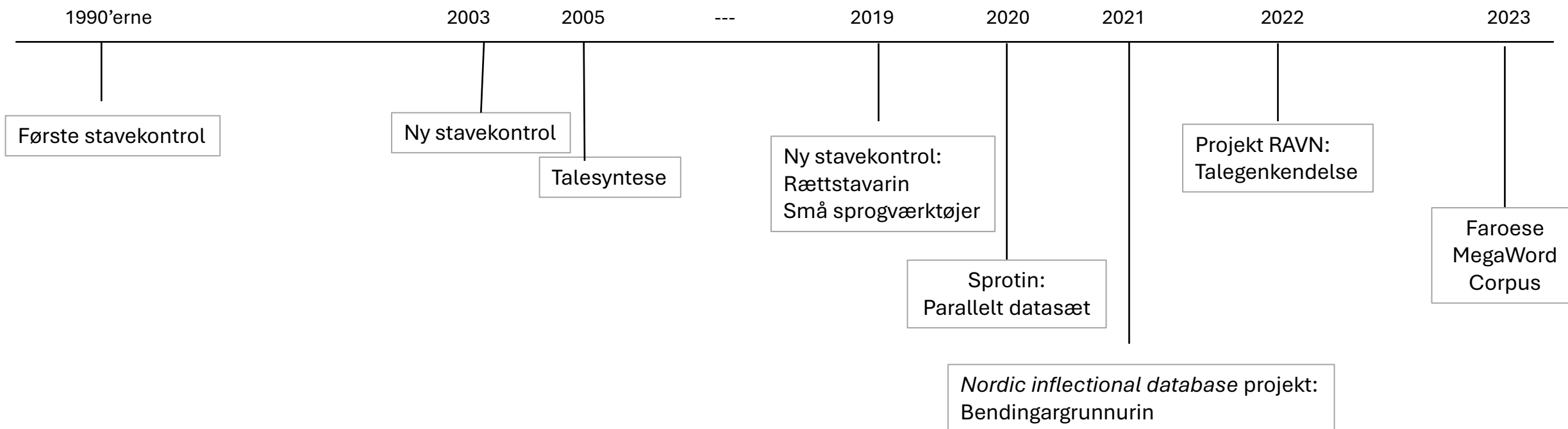
---

- Generelt god udvikling i sprogteknologi
- Små sprog: få talere, få data, manglende infrastruktur, manglende ekspertise, manglende processeringskraft.....
- Styrker:
  - Nationalsprog / hovedsprog
  - Solidt sprogligt fundament: leksikografisk tradition, etableret grammatik, normeringsorgan
  - Høj sproglig bevidsthed og stolthed
  - Engagement fra sprogbrugere
- Geografisk og typologisk midt i Norden

# FÆRØSKE SPROGTEKNOLOGISKE VÆRKTØJER OG PROJEKTER

## HISTORIE OG UDVIKLING - HIGHLIGHTS

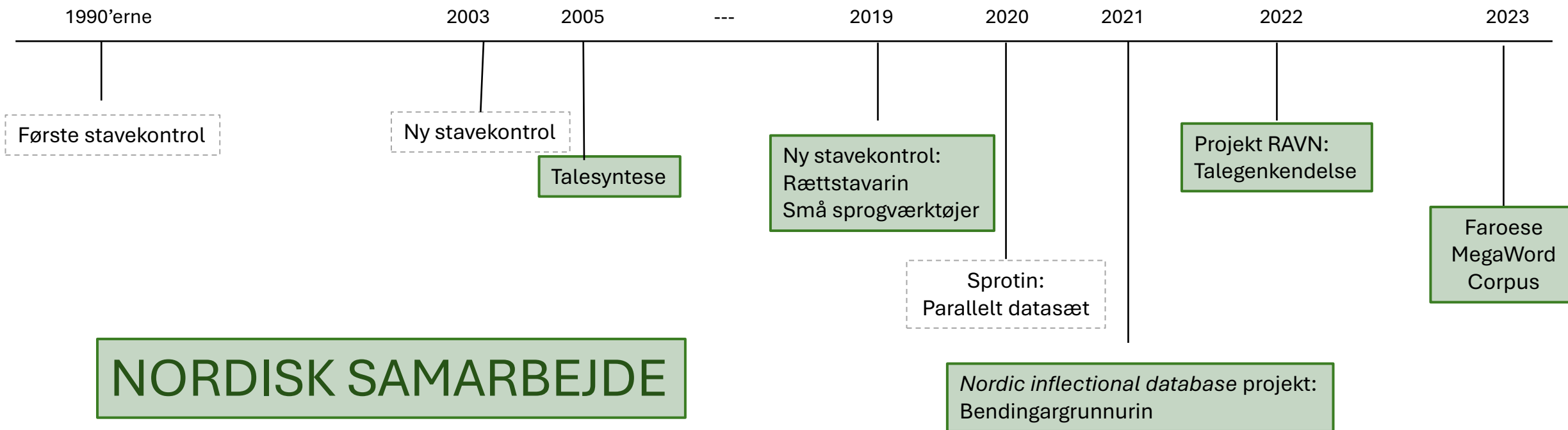
1990-2023



# FÆRØSKE SPROGTEKNOLOGISKE VÆRKTØJER OG PROJEKTER

## HISTORIE OG UDVIKLING - HIGHLIGHTS

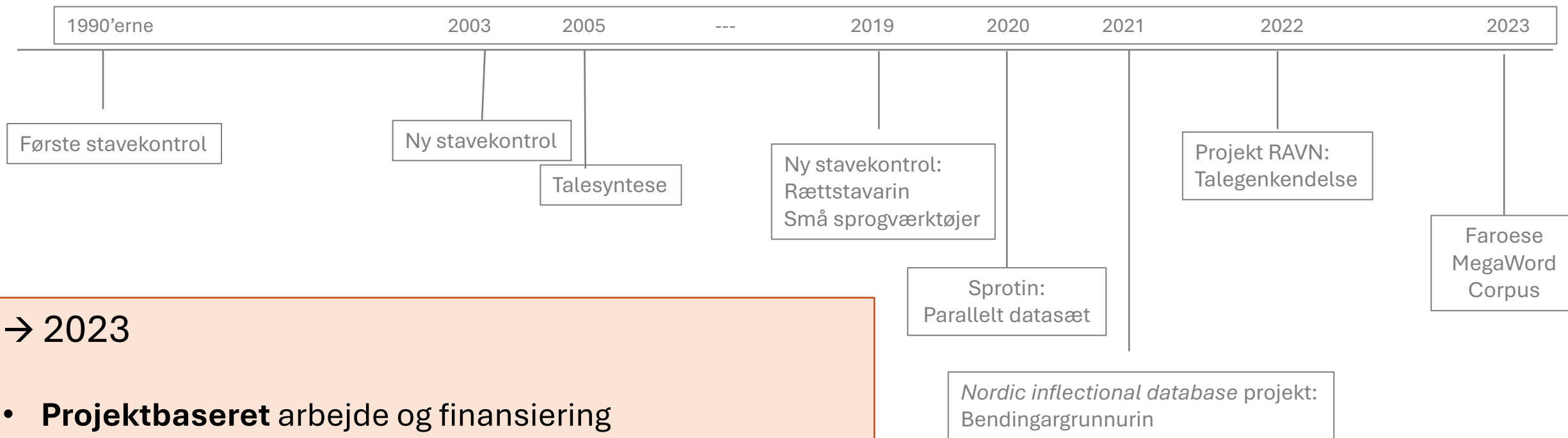
1990-2023



# FÆRØSKE SPROGTEKNOLOGISKE VÆRKTØJER OG PROJEKTER

## HISTORIE OG UDVIKLING

1990-2023



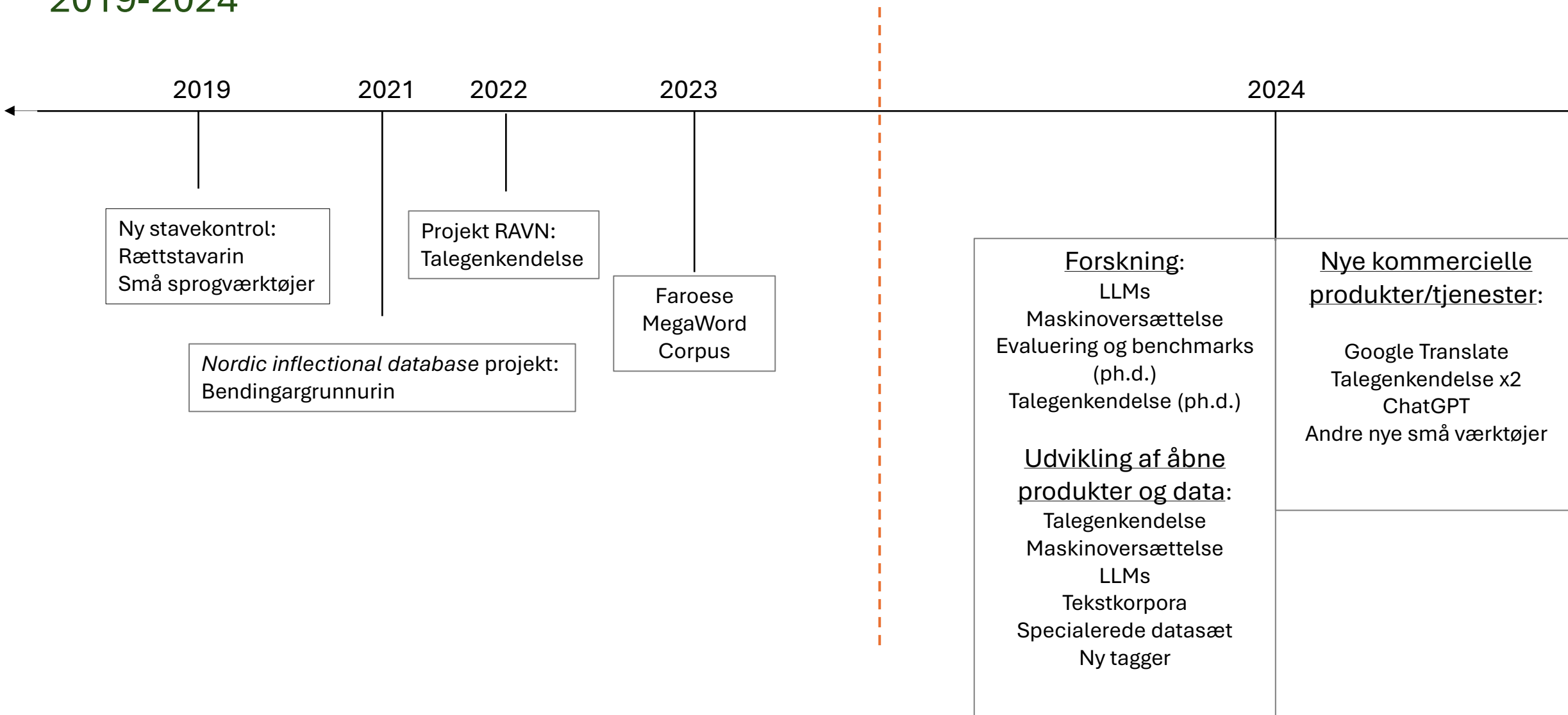
→ 2023

- **Projektbaseret** arbejde og finansiering
- Drevet af **enkeltpersoner**
- Fokus på **dataindsamling** og specifik **produktudvikling**

# FÆRØSKE SPROGTEKNOLOGISKE VÆRKTØJER OG PROJEKTER

NYERE UDVIKLING: Fokus på forskning

2019-2024



# FÆRØSKE SPROGTEKNOLOGISKE VÆRKTØJER OG PROJEKTER

NYERE UDVIKLING: Årsager – rammer omkring sprogteknologi

2023

2024

→ **1) Stabilt grundlag: første offentlige kontinuerlige bevilling til sprogteknologisk arbejde:**

- *Máltøknidepilin* / Center for Sprogteknologi på Færøernes Universitet
- Nytilkommen ekspertise

→ **2) Færøsk, LLMs og AI:**

- Generel accelereret udvikling i AI, GPT4 fra OpenAI, GPTSw3 fra AI Sweden  
→ færøske kompetencer

→ **3) Samarbejder:**

- Samarbejde og forskningsinteresse fra andre lande
- Typologiske “alliancer”



1.

Lokal forskning og udvikling

1.

Lokal  
forskning  
og  
udvikling



MÁLTØKNIDEPILIN  
FRÓÐSKAPARSETUR FØROYA

**Støðuga grundarlagið undir  
menning av føroyskari máltøkni**

“Smáar dátur, stórir dreymar,„

Center for Sprogteknologi  
Universitet på Færøerne

**Stabilt fundament under udvikling  
af færøsk sprogteknologi**

“Små data, store drømme”



MÁLTØKNIDEPILIN  
FRÓÐSKAPARSETUR FØROYA

Støðuga grundarlagið undir  
menning av føroyskari máltøkni

“Smáar dátur, stórir dreymar,”

1.

Lokal  
forskning  
og  
udvikling

## Fokusområder

- Forskning
- Åbne data og ressourcer
  - Produktudvikling
- Rådgivning og støttefunktion
  - Udvikling af ekspertise
  - Eksternt samarbejde
- Oplysning, deling og promovring

## Vi forsker især i...

- Data: indsamle, generere
- Metoder til udvikling, der kræver mindre data
  - Finjustering af LLMs
  - Maskinoversættelse
- Evaluering og benchmarks
  - Talegenkendelse

Se aktiviteter, ressourcer og værktøjer på

[mtd.setur.fo](https://mtd.setur.fo)



MÁLTØKNIDEPILIN  
FRÓÐSKAPARSETUR FØROYA

**Støðuga grundarlagið undir  
menning av føroyskari máltøkni**

“Smáar dátur, stórir dreymar,”

1.

Lokal  
forskning  
og  
udvikling

- Styrker lokal udvikling
- Giver grobund for lokal ekspertise og kompetence
- Samler spredte projekter og effektiviserer
- Giver eksternt vindue ind til færøsk sprogteknologi

2.

Færøsk, LLMs og AI

# Ny udvikling

2.

Færøsk,  
LLMs og AI

Færøske sprogkompetencer i kommercielle applikationer, f.eks

- Google Translate
- ChatGPT, OpenAI
- Claude, Anthropic

- GPT4 kan analysere **sentiment** i færøske tekster omtrent lige så konsekvent som mennesker
- Claude 3.5 Sonnet får højeste score i **maskinoversættelse** i vores egen evaluering
- Begejstrede brugere: virksomheder implementerer egne ChatGPT

Åben modeller på færøsk

- GPTSw3, AI Sweden
  - Llama3, Meta

2.

Færøsk,  
LLMs og AI

Hvordan har store modeller lært færøsk sprog med så lidt data?

- Multilingvale fordele
  - Typologisk fordel som 'midt-nordisk' sprog
- Islandsk politisk indsats og samarbejde med bl.a. OpenAI
  - "Forklædt" færøsk data

→ Færøske sprogfærdigheder er 'ikke med vilje'

Tilfældigheder

Held

Nåde

2.

Færøsk,  
LLMs og AI

Hvordan har store modeller lært færøsk sprog med så lidt data?

- Multilingviale fordele
- Typologisk fordel som 'midt-nordisk' sprog
- Islandsk politisk indsats og samarbejde med bl.a. OpenAI
- "Forklædt" færøsk data

→ Færøske sprogfærdigheder er 'ikke med vilje'

Tilfældigheder

Held

Nåde

Færøsk inkluderet i andre projekter og platforme

- TrustLLM
- No Language Left Behind, Meta
  - TTS og ASR, Meta
  - ScandEval
- Danish Foundation Models



Intention

Inklusion





## 2.

### Færøsk, LLMs og AI

- Vigtig og betydningsfuld udvikling for færøsk
  - Ikke en selvfølge for små sprog
  - Snart del af **samfundsinfrastruktur** (skole, offentlig administration)
- En komponent for **sproglig vitalitet** og **overlevelse**: følge sprogbrugeren
- Udviklingen har “vækket” brugeren - vi må følge med

2.

Færøsk,  
LLMs og AI

Men...

- Tjenesterne inkluderer kun færøsk sprog på output-niveau
  - Ingen lokalisering
  - Ofte er inklusion af færøsk ubevidst
- Sproglig kvalitet er overfladisk
  - Data af dårlig kvalitet
  - Kulturelle bias
- Sårbarhed
  - sproget bliver 'styret' udefra – **manglende autonomi**
  - påvirker vores sprogbrug
  - Vi er afhængige af de store virksomheders nåde og 'goodwill'

Hvordan får vi gavn af mulighederne samtidig med, at vi bevarer sproglig autonomi?



## Vi bygger en motor

En lille, men stabil, “motor” med styrmand og maskinmester

- Institutionel forankring
- Udvikler lokal ekspertise og kompetencer → “national resiliens”
- Fokuseret forskning og udvikling
- Arbejder for lokalspecifikke tjenester
- Deler alle data og modeller
- Støtter op om lokalt miljø for produktudvikling



## Vi sætter sejlene

Tilpasse vores aktiviteter og data til fordel for inklusion i større tjenester

- Udnytter samtidens vind
- Er med til at bestemme retningen
  - Kurere data efter forespørgsel
  - Pushe god data
  - Skabe forbindelser
  - Følge med udvikling og giver anbefalinger
  - Tilbyde hjælp til lokalisering – ingen andre gør det, end os



## Vi bygger en motor

En lille, men stabil, "motor" med styrmand og maskinmester

## Vi sætter sejlene

Tilpasse vores aktiviteter og data til fordel for inklusion i større tjenester

## Sproglig kvalitetssikring

- Udvikle benchmarks til evaluering
  - Originalfærøske testdata



3.

Nordisk samarbejde

3.

Nordisk  
samarbejde

Nordisk samarbejde er nødvendigt for os!

- Infrastruktur
- Vidensdeling
  - Netværk
  - Forskning
- Udvikling af tjenester
  - Vejledning (ph.d.)
  - Regional resiliens

## Motor

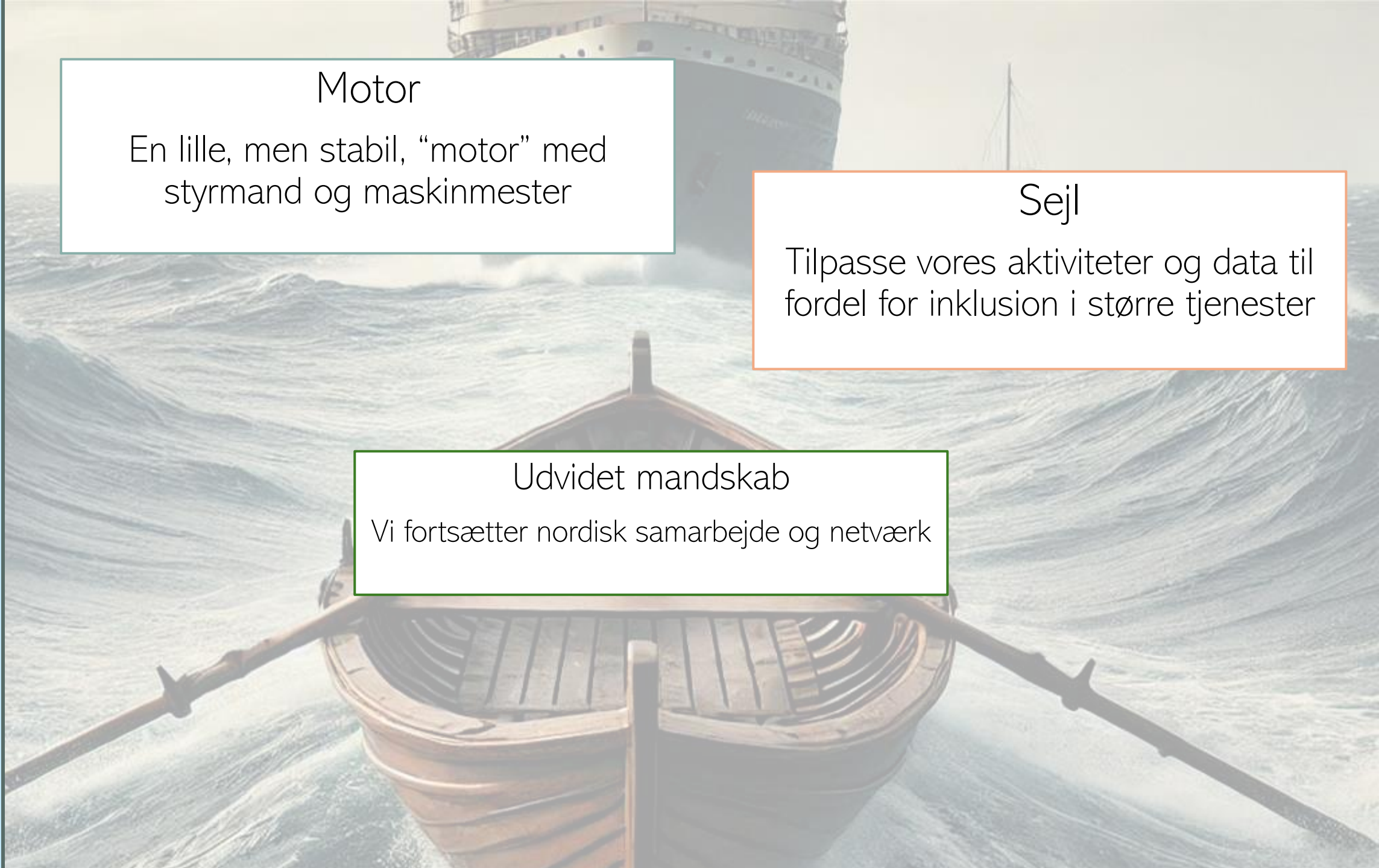
En lille, men stabil, "motor" med  
styrmand og maskinmester

## Sejl

Tilpasse vores aktiviteter og data til  
fordel for inklusion i større tjenester

## Udvidet mandskab

Vi fortsætter nordisk samarbejde og netværk





# Mod destinationen

---

Navigation mellem udfordringer og muligheder: vi kan ikke gøre alt selv, men vi vil være med!

Trods positiv udvikling, er færøsk stadig i en sårbar position

- Balance mellem lokal autonomi og fleksibilitet mod store modeller, platforme og projekter
- Øge lokalt politisk ansvar
- Fortsat nordisk samarbejde, dataindsamling og kvalitetsbevidst udvikling for færøsk sprogs overlevelse og vitalitet



Tak!

---

DREAM  
**BIG**

