

CL



20. november 2025

Pilotprojekt om byrådsdata: Juridiske afklaringer og mulige løsninger



Sprogteknologi.dk er et **søgbart katalog** over metadata for **sprogrsourcer**, der er tilgængelige for dansk. Det dækker derudover også et **fællesoffentligt initiativ**, der supplerer og understøtter **udviklingen** af **dansk sprogteknologi** gennem en **række indsatser**:

1. **Finde og udstille metadata** om relevante **danske sprogrsourcer**
2. **Udvikle nye sprogrsourcer** og indgå i **samarbejdsprojekter** herom
3. **Vidensdeling, netværk og internationalt samarbejde**

Besøg sprogteknologi.dk, kontakt os ved info@sprogteknologi.dk, eller følg os på [LinkedIn](#).

Regeringens 'Strategisk indsats for kunstig intelligens', 2024

Skabe et solidt fundament og en klar retning for udvikling og anvendelse af kunstig intelligens i Danmark gennem fire initiativer:

1. Den Digitale Taskforce for kunstig intelligens skal gøre den offentlige sektor verdensførende
2. Center for kunstig intelligens i samfundet (CAISA)
3. Sikker platform til udvikling af transparente danske sprogmodeller (Danish Foundation Models)
4. **Danske tekstdata skal gøres frit tilgængelige**



Initiativ 4: Danske tekstdata skal gøres frit tilgængelige

Målsætningen er at tilgængeliggøre danske tekstdata i omegnen af 300 mia. tokens (ca. 200 mia. ord), udgivet i et standardiseret og maskinlæsbart format frem mod udgangen af 2027.

Danske tekstdata er en vigtig forudsætning for, at kunstig intelligens fungerer godt på dansk i fremtiden – både med hensyn til sprog og kulturforståelse.

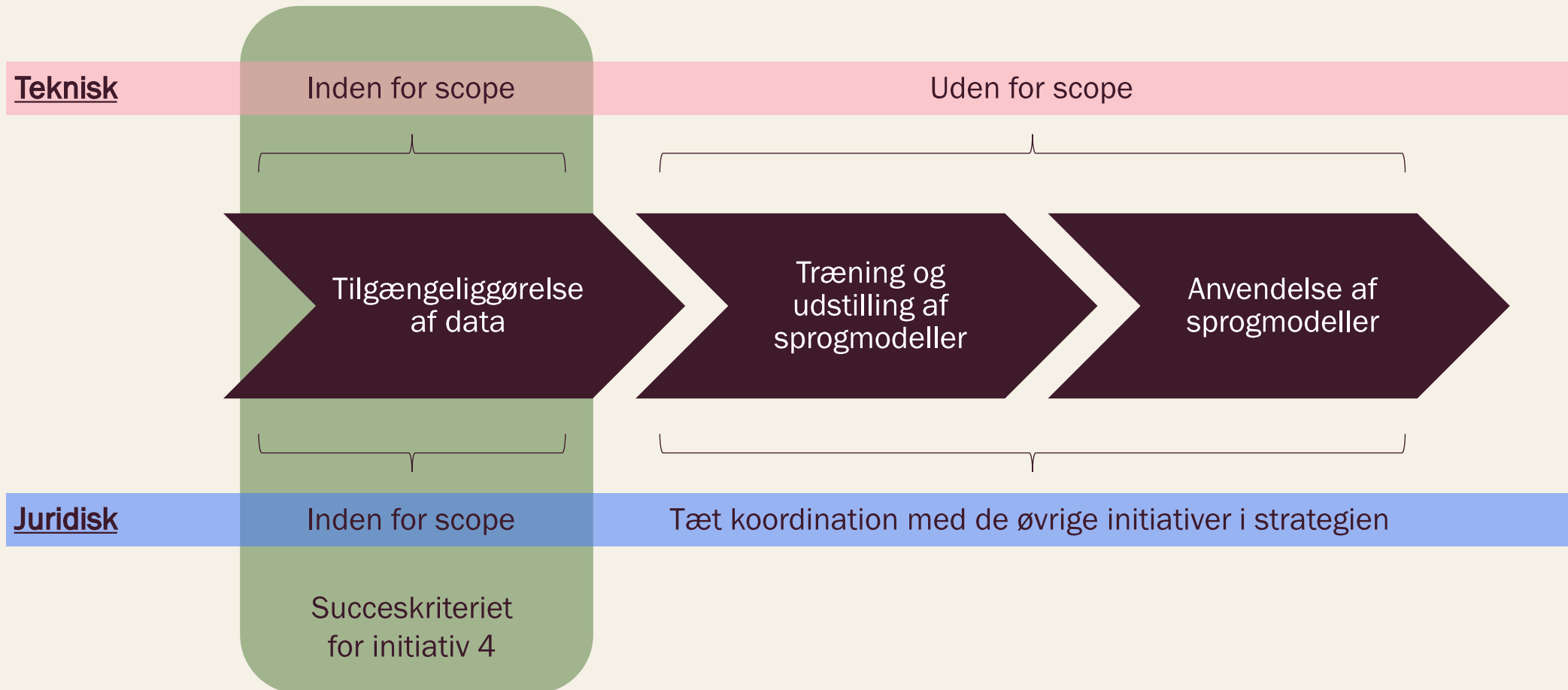
Med initiativet tilgængeliggøres i første omgang tekstdata fra Rigsarkivet, Det Kgl. Bibliotek og Folketinget.

Der skal tages hensyn til ophavsret, personoplysninger og relevante særlovgivninger. Med initiativet afsøges muligheden for aftalelicenser med dataejere.

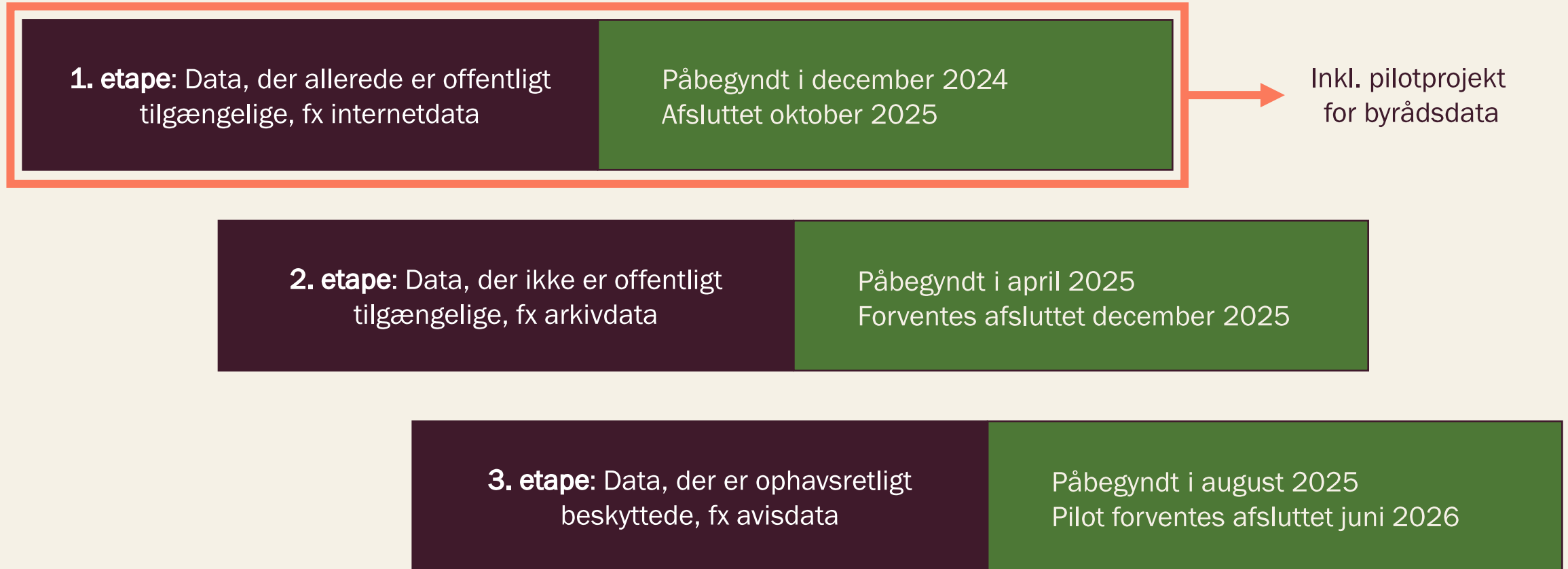
Nem og effektiv adgang til data skal bidrage til at accelerere innovation i eksempelvis små- og mellemstore danske virksomheder.



AI-forsyningskæden



Det juridiske spor for tilgængeliggørelse af sprogdata



Pilotprojekt for byrådsdata

Deltagere

- Digitaliseringsstyrelsen
- Open Data DK
- Aarhus Kommune
- Favrskov Kommune
- Kalundborg Kommune
- Tønder Kommune
- Vejle Kommune

Formål

- PoC for tilgængeliggørelse af byrådsdata
- Dokumentation af juridisk og teknisk protokol med henblik på genbrug, inspiration og opskalering

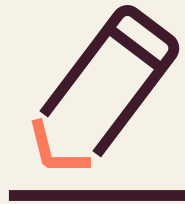
Datas karakteristika

- Offentligt tilgængelige
- Fra den offentlige sektor
- Indeholder sagsfremstillinger
- Indeholder personoplysninger
- Indeholder ophavsretligt beskyttet materiale

Fase 1: Juridiske forudsætninger



Udvælge data til
juridisk analyse



Formulere juridiske
spørgsmål



Overblik over
juridiske rammer



Juridisk protokol
for dataudgivelse



Juridiske
vurderinger



Overblik over juridiske rammer

Generelt notat om internetdata

- Databeskyttelsesretlige og ophavsretlige rammer for videreanvendelse med henblik på genbrug til forskellige slags internetdata, der er lovligt tilgængelige.

Specifikt notat om byrådsdata

- Kommunernes spørgsmål til forekomster i byrådsdata med konkrete eksempler, fx forekomst af kommunalpolitikeres navne eller hhv. borgeres og virksomheders adresser.

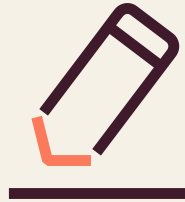
Case-analyse

- Indholdsmæssig gennemgang af dagsorden, referat og samtlige bilag fra to tilfældigt udvalgte møder med henblik på at identificere forekomster af almindelige og følsomme personoplysninger samt ophavsretligt beskyttet materiale.

Fase 1: Juridiske forudsætninger



Udvælge data til
juridisk analyse



Formulere juridiske
spørgsmål



Overblik over
juridiske rammer



Juridisk protokol
for dataudgivelse



Juridiske
vurderinger



Juridisk protokol for dataudgivelse (1)

Ophavsretligt materiale

Frasortering af dokumenter fra private organisationer og alt grafisk materiale.

Almindelige personoplysninger

Af hensyn til udførelse af en opgave i samfundets interesse, som beskrevet i FODS 2022-2025 og Strategisk indsats for kunstig intelligens.

Følsomme personoplysninger

Frasortering af dokumenter indeholdende følsomme personoplysninger, der ikke er offentliggjort af de registrerede selv, af hensyn til væsentlige samfundsinteresser.

Formålsforenelighed

Afsæt i Datatilsynets vejledning om Offentlige myndigheders brug af kunstig intelligens med særlig vægt på, at oplysningerne allerede er offentligt tilgængelige.

Mulig videregivelse?

Foreløbigt ikke nogen sikker videregivelseshjemmel, men der arbejdes videre med dette spor til en eventuel opskalering.

Juridisk protokol for dataudgivelse (2)

DPIA mv.

Gennemførelse af risikovurdering, tærskelanalyse og konsekvensanalyse.

Oplysningspligt

Massekommunikerer via styrelsens hjemmeside, da det er vanskeligt, hvis ikke umuligt, at udpege og finde kontaktoplysninger på de registrerede.

Dataminimering

Direkte personidentifikatorer pseudonymiseres, da internetdata ikke kan anonymiseres, og det ikke kan forsvares at videregive autentiske identifikatorer. Quasi-identifikatorer bevares af hensyn til semantik.

Videreanvendelse

Fordi data kun er pseudonymiserede, skal anvendere af datasættet supplere med juridiske vurderinger.

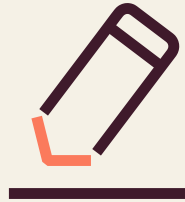
Rigtighed

Der indføres tidsopmærkning i metadata, så anvendere kan vurdere, om data er aktuelle eller forældede til formålet med videreanvendelsen.

Fase 1: Juridiske forudsætninger



Udvælge data til
juridisk analyse



Formulere juridiske
spørgsmål



Overblik over
juridiske rammer



Juridisk protokol
for dataudgivelse



Juridiske
vurderinger



Juridiske vurderinger og risici



At der i de udstillede data kan indgå materiale, der ved en konkret vurdering kan vurderes at opnå værkshøjde og dermed være ophavsretligt beskyttet.



At der i de indsamlede data kan indgå følsomme personoplysninger, som ikke er dækket af hjemmelsvurderingen i konsekvensanalysen. Der lægges derfor op til, at der screenes for sådanne oplysninger med henblik på frasortering.



At der er behandlingsgrundlag for sådan screening og frasortering.



At der er formålsforenelighed mellem behandlingen og den behandling, som oplysningerne oprindeligt er indsamlet til.

Fase 2: Teknisk udførelse



Udvalgte data



Udtræk af data



Efterbehandling til
tekstdatasæt



Udstilling på
sprogteknologi.dk
og opendata.dk



Eventuel
opskalering



December
2025

Hovedbudskaber



At data er offentligt tilgængeligt er ikke ensbetydende med, at de uden videre kan videreanvendes til brug for udvikling af sprogmodeller.



Der er både ophavsretlige og databeskyttelsesretlige hensyn, og det er praktisk umuligt at gennemføre uden at påtage sig visse risici.



Hele dataansvaret kan ikke absorberes af ét led i "forsyningskæden" – alle er nødt til at spille en rolle, hvis det skal lade sig gøre.



Tak for jeres
opmærksomhed!

Casper Frohn, Digitaliseringsstyrelsen
casfro@digst.dk