
Teknisk løsning for nasjonalt beregningsverktøy

Dataflyt, systemdesign og krav til infrastruktur

Morgan Kjølervakken
Islen Vallejo



Oppdragsrapport

Innhold

Side

Innhold	1
1 Innledning.....	3
1.1 Mål med rapporten	3
1.2 Arbeidspakke 1 - Teknisk løsning	3
2 Krav til teknisk løsning.....	4
2.1 Krav til løsning per i dag (fase 1)	4
2.2 Krav til fremtidig løsning (fase 2).....	4
3 Systemoversikt.....	5
3.1 Oversikt	6
4 Datafangst	7
4.1 Data fra målestasjoner	7
4.2 Utslippsdata.....	8
4.2.1 Utslippsdata fra trafikk.	8
4.2.2 Utslippsdata fra vedfyring	9
4.2.3 Utslipps data fra industri	9
4.2.4 Utslippsdata fra skip og havn	10
4.3 Meteorologidata.....	10
4.4 Bakgrunnsdata.....	10
4.5 Helsedata og befolkningsdata	11
4.6 Preprosessering av data	11
5 Beregninger	12
5.1 Oversikt	12
5.2 Utslippsmodellen.....	12
5.3 Spredningsmodellen.....	12
5.4 Statistiske beregninger	12
5.5 Eksponering	13
5.6 Datautvekslingsformat NetCDF	13
5.7 Eksport av resultater til databasene	13
6 Visualisering.....	14
6.1 Definerte visninger	14
6.1.1 Visninger for hele Norge.....	15
6.1.2 Visninger for norske byer	15
6.2 Webløsning.....	15
6.3 Web design og prototyp.....	15
6.3.1 Navigasjon og menysystem	16
6.3.2 Forurensningskart.....	17
6.3.3 Befolkningseksponering	18
6.3.4 Luftsonkart.....	19
6.3.5 Kilder.....	20
6.3.6 Fremskrivninger	21
7 Infrastruktur	22

7.1	Krav til infrastruktur	22
7.2	Prosesseringsbehov	22
7.3	System og infrastruktur for visualiseringer.....	23
7.3.1	PostGIS	23
7.3.2	GeoServer.....	23
7.4	Lagringsbehov	23
7.5	Skyløsning	24
8	Åpen kildekode	25
8.1	Utviklingspråk og åpen kildekode	25

1 Innledning

1.1 Mål med rapporten

Målet med rapporten er å dokumentere identifiserte krav og tenkt dataflyt for nasjonalt beregningsverktøy (NBV). Rapporten skal vise veien for første fase av utviklingen av NBV samtidig som den tar høyde for hvordan verktøyet ønskes utviklet videre etter denne første fasen. Rapporten er inndelt i tre hoveddeler som beskriver de tre hovedelementene i verktøyet, datafangst av inngangsdata, beregninger og visualiseringer. I tillegg er det gjort vurderinger i forhold til krav til infrastruktur i denne første fasen av Nasjonalt Beregningsverktøy.

1.2 Arbeidspakke 1 - Teknisk løsning

Hovedmålet med arbeidet i denne arbeidspakken er å etablere den tekniske løsningen som vil være basis for NBV. Det nye beregningsverktøyet må kunne løse beregningsoppgaver i henhold til eksisterende regelverk og benytte seg av eksisterende operative løsninger og tilgjengelige data. Ved å bli koblet opp til eksisterende systemer, vil beregningsverktøyet gi en betydelig merverdi i forhold til om det ble utviklet og driftet isolert. Dette setter viktige føringer til den praktiske gjennomføringen av prosjektet slik at arbeidsplanen bør ta hensyn til behov for kobling- og integrasjonsarbeid av eksisterende løsninger og tilgjengelige data. Denne arbeidspakken har derfor jobbet tett sammen med AP2- utslippsdata for å identifisere relevante data og dataflyt. I tillegg er det vurdert behov for teknisk utviklingsarbeid for å bearbeide data slik at de kan brukes av beregningsverktøyet.

Målet med rapporten er å dokumentere identifiserte krav og planlagt dataflyt for nasjonalt beregningsverktøy. Rapporten skal vise veien for første fase av utviklingen av nasjonalt beregningsverktøy samtidig som den tar høyde for hvordan verktøyet ønskes utviklet videre etter denne første fasen.

2 Krav til teknisk løsning

Denne rapporten tar i hovedsak for seg krav til løsning i denne første utviklingsfasen, men det er i hele systemdesignet tenkt og lagt til grunn at dette også vil være rammeverket for en fremtidig interaktiv løsning. Dette innebærer at mange av designvalgene er gjort med tanke på videreutvikling og oppskalering.

2.1 Krav til løsning per i dag (fase 1)

Løsningen skal være et effektivt verktøy for forskere og konsulenter som jobber med analyse og modellering av luftkvalitet. Det er viktig å ivareta en modelleringsflyt som gjør at forskere og konsulenter kan benytte de ulike modelleringsmodulene også uavhengig av NBV og uten å måtte være koblet opp mot en database. Samtidig skal verktøyet tilby en online datafangst- og publiseringsløsning.

Publisering av resultater skal i stor grad basere seg på å visuell kommunikasjon fra de ulike modellene gjennom et kartbasert grensesnitt hvor man navigerer og analyserer resultatene basert på geografisk lokasjon. Løsningen skal ha et responsivt design slik at den skal kunne brukes på alle typer enheter, mobil, nettbrett og PC.

Systemet skal være fleksibelt i forhold til hva slags modeller som skal kunne integreres, men det vil settes krav til hvilket dataformat og datastruktur modellene benytter.

Det vil også være et krav at systemet har en god ytelse. Særlig viktig blir det at visualiseringsdelen blir rask og brukervennlig.

Det er et krav at løsningen er åpen. Dette vil innebære at all kildekode utviklet for innhenting av data, kildekode for modeller brukt i løsningen og kildekode til visualisering blir gjort fritt tilgjengelig og kan benyttes og videre utvikles i henhold til lisens for åpen kildekode. Dette er også viktig at data blir gjort fritt og lett tilgjengelig gjennom åpne APIer.

2.2 Krav til fremtidig løsning (fase 2)

Etter denne første utviklingsfasen er intensjonen at verktøyet skal utvikles videre til en interaktiv løsning for forvaltning og konsulenter. For interaktiv bruk av modellverktøyene, vil det bli nødvendig med tettere kobling mellom modellverktøyene og front-end av systemet. Et Content Management-system vil også bli nødvendig slik at brukeren kan legge inn tekst knyttet til utførte beregninger.

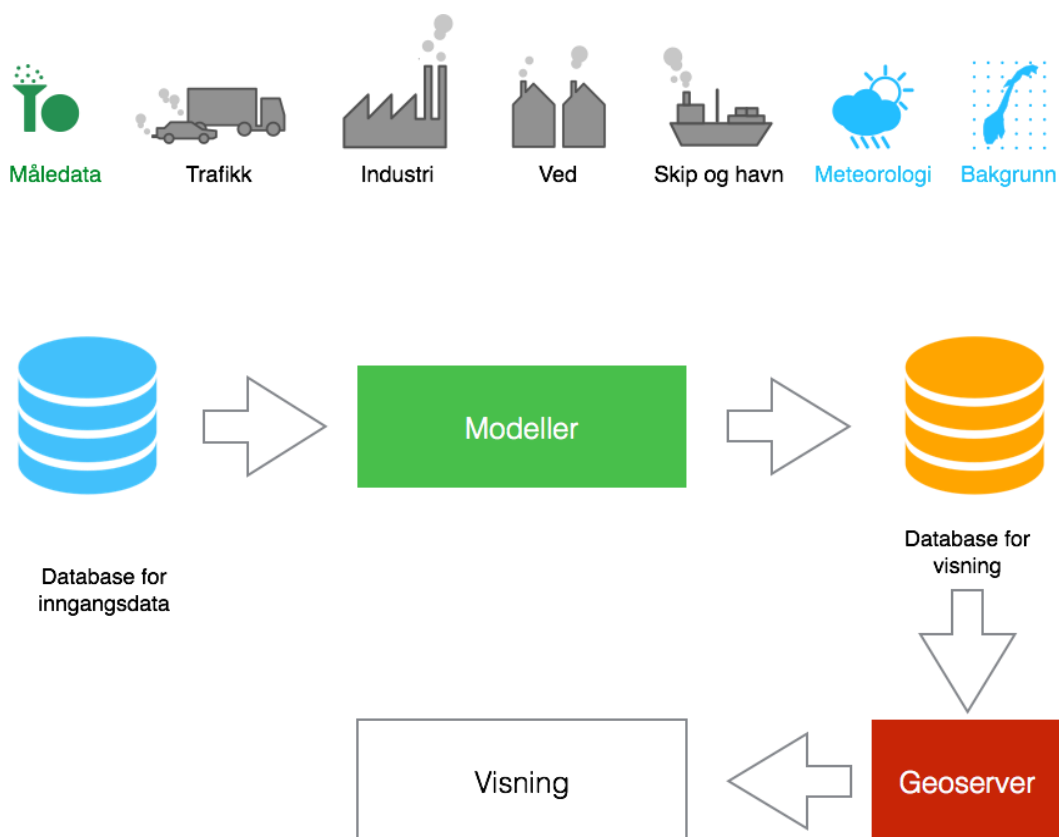
Når systemet går over i en operativ fase vil det stille høyere krav til sikkerhet og oppetid. En spesifisering av infrastruktur for operasjonell løsning anbefales gjennomført når konkrete krav til en interaktiv løsning foreligger og man har bedre kjennskap til forventet bruk og data mengder systemet skal skaleres for.

3 Systemoversikt

I NBV vil vi implementere en fleksibel dataflyt som sørger for at vi har et godt og brukervennlig verktøy for forskere og konsulenter som bruker verktøyet til daglig, samt en robust løsning som inngår i et produksjonsmiljø som skal levere resultater for visning på web. En skjematisk oversikt av systemet er gitt i Figur 1.

Det vil være sentralt å sikre lagring av alle inngangsdata. Resultatene fra modellberegningene vil sikre at dataene blir reproduserbare og dermed gi brukeren mulighet til å kjøre en tilsvarende beregning på samme inngangsdata. Imidlertid må det påregnes noen modifikasjoner av parameterne, som for eksempel å endre trafikkmengde i veinettet.

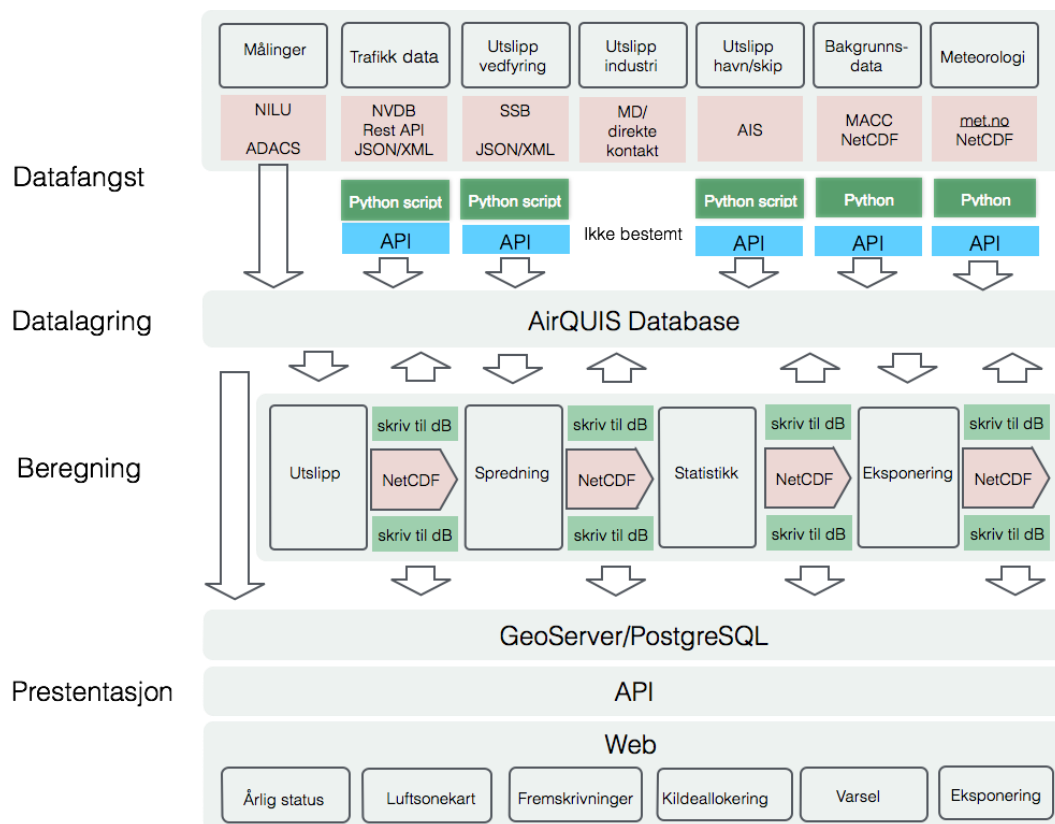
Vi ønsker også en rask og god visning av resultatene på web. Alt som skal vises på web vil derfor bli lagret i et eget visnings-lag med egen tilhørende database. En kartserver (GeoServer) benyttes til backend-løsningen for å generere kartdata til web-løsningen for visning.



Figur 1. Dataflyt. Inngangsdata lagres i en database. Modellene henter nødvendig data fra databasen før de prosesseres. Resultater som skal vises legges i en visningsdatabase og en kartserver (GeoServer) benyttes til å distribuere dataene til web.

3.1 Oversikt

En mer detaljert oversikt over systemet og dataflyten vises i Figur 2.



Figur 2. Dataflyt for det totale systemet delt inn i lagene: Datafangst, datalagring, beregning og presentasjonslaget. Inngangsdata fra Statistisk Sentralbyrå (SSB), Miljødirektoratet (MD), Meteorologisk institutt (met.no) og Nasjonal veg database (NVBD).

4 Datafangst

En viktig oppgave i nasjonalt beregningsverktøy er å automatisere prosessen med å innhente data. Dette er viktig for å effektivisere datainnhenting som kan være en meget tidkrevende prosess. En automatisert prosess reduserer også risikoen for menneskelige feil som skyldes manuell databehandling.



I prosjektet er inngangsdataene inndelt på en hensiktsmessig måte i fire klasser, i form av dataenes opprinnelse samt struktur:

- *Måledata*, data fra landets målestasjoner
- *Utslippsdata*, omfatter samtlige utslipp fra trafikk, industri, vedfyring samt utslipp fra skip- og havneområder
- *Meteorologi*
- *Bakgrunnsdata*

Generelt legges det opp til at vi automatisk innhenter data fra dataleverandørens database. For eksempel kan nødvendig trafikkdata, som benyttes til å beregne utslipp fra trafikk, hentes direkte fra Nasjonal Veidatabase (NVDB). Vi ser at flere og flere dataleverandører tilbyr en web-service hvor man gjennom deres definerte APIer kan hente direkte ut ønskede data.

I NBV vil det etableres et lag mellom dataleverandørens web-service og NBV sin database som kan gjøre enkle operasjoner på innhentet data før det lagres i databasen. Til denne oppgaven er det tenkt å bruke skripting i språket Python. Python vil være «limet» som binder databasene sammen og kontrollerer dataflyten.

For å sikre at data lagres riktig i AirQUIS-databasen, og at det ikke jobbes direkte inn mot databasen, vil det etableres et API rundt databasen. API-laget skal utvikles i .NET rammeverket.

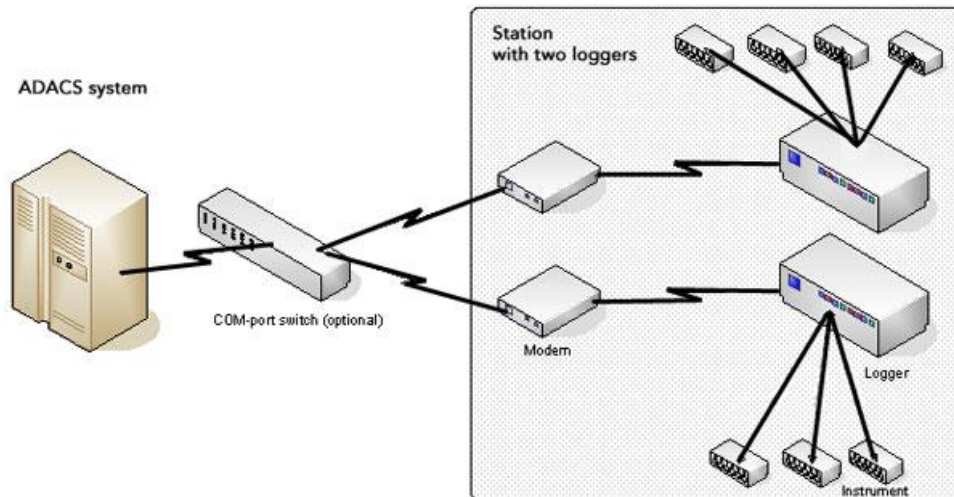
Dataflyten for innhenting av data er gjenspeilet i figur 3.



Figur 3. Generell dataflyt for innhenting av data

4.1 Data fra målestasjoner

Data fra målestasjonene innhentes fra NILU sitt allerede godt etablerte ADACS datainnhentingssystem. ADACS innhenter data fra forskjellige komponenter som måles på målestasjoner rundt om i landet. Dataene lagres deretter i AirQUIS-databasen. NBV kommer ikke til å gjøre endringer i forhold til dagens system, se figur nedenfor.



4.2 Utslippsdata

Utslippsdata omfatter alle typer utslipp som brukes som inngangsdata til modellene. Det er data fra trafikk, industri, vedfyring samt utslippsdata fra skip- og havneområder.

4.2.1 Utslippsdata fra trafikk.

For beregning av utslipp fra trafikk for ulike vegstrekninger, benyttes veinoder. Videre benyttes en klassifisering av kjøretøyer i forhold til kjøretøytype, inndelt i tre klasser henholdsvis; lette, tunge og meget tunge kjøretøy.

Tabell 1: Tabellen viser ønsket informasjon fra NVDB

No.	Name	Definition	Type of data
1	Road classes	Types of road classes	Number of road classes and road types like highways, national roads, city main roads, industrial area roads and some important minor roads
2	Registered Vehicle Classes (RVC)	Type of vehicles	Light, Heavy and very Heavy Vehicles
3	Road nodes	Road nodes define the start and end point for a road link	(x,y) coordinate and height above ground of the nodes
4	Static traffic data	Static properties of the road link	Length, direction, total width (not including sidewalks) for each direction, gradient (%)
5	Dynamic traffic data	Dynamic data on the road links	ADT (annual daily traffic) is total number of vehicles of all vehicle types for the road link direction, speed limit on the road link in km/h
6	Vehicle distribution	Percentage (%) of Registered Vehicle Classes (RVC) and time variation on each road links	Percentage of light, heavy, very heavy vehicles at each road link or road class for each direction
8	Emission Calculated Vehicle Classes (ECVC)	The sub class of Registered Vehicle classes (RVC) is named Emission Calculated Vehicle Classes (ECVC). Classify vehicles into category like motor bikes, diesel/gasoline cars, light trucks, heavy duty trucks, buses,...	Separate between different technologies (Pecat, EURO1, 2,3...) Average Model Year, Average Driving Distance
9	ECVC – RVC Distribution	The coupling between the Emission Calculated Vehicle Classes (ECVC) and Registered Vehicle Classes (RVC)	Percentage of ECVCs for one RVC. The sum of percentage values for one RVC must be 100%.
10	ECVC – Fuel Consumption	The range of fuel consumption for each ECVC depending on the speed	Driving speed (km/h), Fuel consumption (l/km)

No.	Name	Definition	Type of data
11	ECVC Basic and Ageing Factors	The basic factor is the fuel or emission factor for the Average Model Year for a given ECVC. The aging factor adjusts the fuel consumption or emission for the calculation year	Basic factor (g/l if it is fuel based and g/km if it is traffic based). Aging factor for the ECVC.
12	ECVC Speed Dependency Factors Traffic Time variation	Each ECVC/emission component/speed combination has its own speed dependency factor. Only traffic based emission components are using the speed dependency factors Each vehicle class for individual road can be defined with yearly, weekly and daily variation of the traffic flow	Driving speed (km/h), Speed dependency factor Factors from 0-1

Data for å beregne utslipp hentes fra Statens vegvesen sin database, Nasjonal vegdatabank (NVDB). Informasjon fra NVDB opererer under Norsk lisens for offentlige data. Dette betyr at informasjonen kan, med de begrensningene som følger av lisensen, brukes til ethvert formål og i en enhver sammenheng. Man aksesserer databasen gjennom NVDBs REST-API som gir tilgang til alle veg-objekter som er lagret i NVDB, samt tilhørende metadata om veg-objektene.

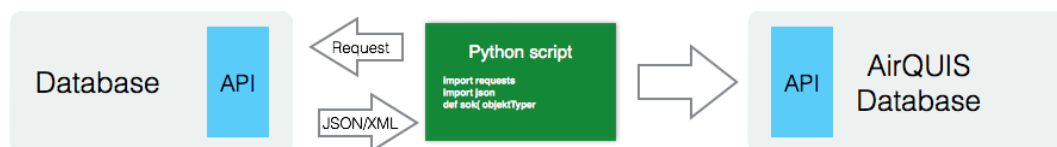
Ved hjelp av et Python skript, hentes dataene fra databasen og struktureres for så å legges over i AirQUIS-databasen, se figur 4.



Figur 4. NBV kobler opp mot NVDB sin database og legger dataen over i databasen.

4.2.2 Utslippsdata fra vedfyring

Vi ønsker å innhente data ved en automatisert spørring mot SSB sin database. Under arbeidspakke 2 (AP2) jobbes det for å identifisere hvilke data som skal hentes fra Statistisk Sentral Byrå (SSB). I fremtiden er det ønskelig å etablere en automatisk innhenting av data, med en automatisert spørring mot SSB sin database, men i denne fasen er utarbeidelse av forslag til dataflyt første prioritering for AP1.



Figur 5. NBV ønsker en automatisk oppkobling mot SSB sin database .

4.2.3 Utslipps data fra industri

Det er foreløpig ikke bestemt hvordan utslipp fra industri skal innhentes, men det er ønskelig å se på mulighetene for å etablere en automatisk innhenting mot Miljødirektoratet sin database. Som en del av arbeidspakke 2 jobbes det for å finne effektive og gode metode for innhenting av nødvendige utslippsdata til NBV. I denne fasen av utviklingen av den tekniske løsningen er utarbeidelse av forslag til dataflyt første prioritering.



Figur 6. NBV ønsker en automatisk oppkobling mot MD sin database .

4.2.4 Utslippsdata fra skip og havn

I dag benyttes en manuell prosess hvor data innhentes fra loggbøker gitt av den aktuelle havn. I Nasjonalt beregningsverktøy ønsker vi å se på en automatisert prosess, Automatic Tracking Identification System (AIS). Systemet ble introdusert av FNs International Maritime Organisation (IMO) for å øke sikkerheten til skip, forbedre skipstrafikkovervåkingen, trafikkjenester og miljøhensyn. Systemet tilbyr posisjon, hastighet, kurs og skips-id i sanntid. Dataene kan videre benyttes til å estimere utslipp. Før metoden tas i bruk, må den verifiseres i forhold til datainnhenting og utslippsestimat. Metoden virker som en lovende vei for fremtiden, og vi vil se nærmere på dette i 2015.

4.3 Meteorologidata

Meteorologidata produseres for hele landet i en oppløsning på 2,5 x 2,5 km, og for et gitt antall vertikale lag. I utvalgte byer og tettsteder vil oppløsningen være høyere og økt til 1 x 1 km. Tidsoppløsningen er satt til timenivå slik at det opereres med 24 datasett per døgn.

Datamengden setter et høyt krav til lagringskapasitet om man skulle ta vare på dette datasettet lokalt. Det legges derfor opp til at vi lokalt i NVB kun lagrer nødvendig data for tettsteder og byer som er inkludert i løsningen for å sikre en rask tilgang på data ved prosessering. Meteorologisk Institutt (MET) vil lagre det originale komplette datasettet for hele landet slik at NBV vil ha mulighet til å innhente data også fra områder som er utenfor de allerede definerte områdene. Den prosessen ønskes også automatisert. Datasettet vil hentes for ett år av gangen når nytt eller oppdatert datasett er produsert.

Et komplett datasett vil utgjøre 8,5 TB per år i oppløsning på 2,5 x 2,5 km. Ved på øke oppløsningen på gridet til 1 x 1 km, vil datamengden øke med en faktor på 6,25.

Meteorologidata i form av timesverdier overføres på årsbasis fra MET til NBV for Norge med en oppløsning på 2,5 x 2,5 km, og for de 7 testbyene i en oppløsning på 1x1 km. Filene kommer på NetCDF-format. Det er mulig at dataene av praktiske hensyn grupperes i for eksempel dag, uke eller måned.

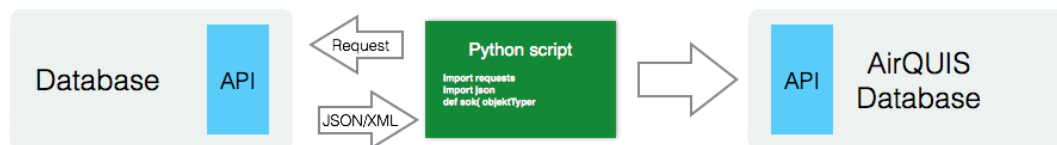
Vi ønsker også å etablere en løsning for å hente ut data fra MET som ikke er lokalt lagret i NBV databasen. MET tilbyr allerede denne tjenesten til andre datasett i form av en web-service. En slik tilgang til hele datasettet vil sikre at vi har en fleksibel løsning for fremtiden.

4.4 Bakgrunnsdata

Bakgrunnsdata vil hentes fra Monitoring Atmospheric Composition & Climate (MACC), tilsvarende det som i dag gjøres for Bedre byluft.

4.5 Helsedata og befolkningsdata

Under AP2 jobbes det for å identifisere hvilke data som skal hentes fra (SSB). I fremtiden er det ønskelig å etablere en automatisk innhenting av data med en automatisert spørring mot SSB sin database, men i denne fasen av er utarbeidelse av forlag til dataflyt første prioritering for AP1.



Figur 7. NBV ønsker en automatisk oppkobling mot SSB sin database .

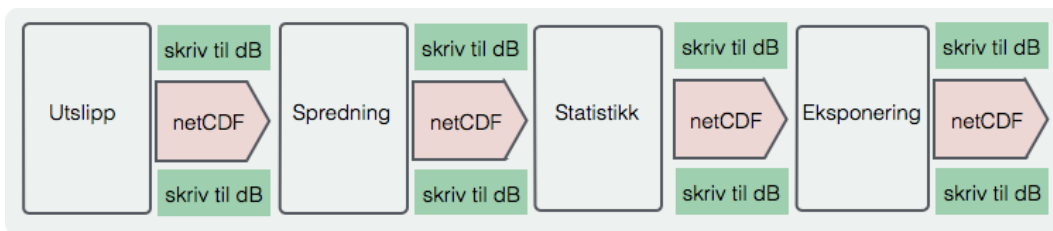
4.6 Preprosessering av data

Preprosessering av data gjøres for å tilpasse inngangsdata slik at de harmonerer med eksisterende databaseformat. For eksempel må grid-data for bakgrunnskonsentrasjoner og meteorologi interpoleres både horisontalt og vertikalt på gridpunktene som spredningsmodellen baserer sine beregninger på.

5 Beregninger

5.1 Oversikt

Modellene som beregner luftforurensning i Nasjonalt beregningsverktøy består av en utslippsmodell, en spredningsmodell, en statistikkmodell og en eksponeringsmodell. Det legges opp til at nødvendige data innhentes fra databasen og lagres i filer som videre brukes som inngangsdata til de ulike modellene. Utslippsmodellen vil være den eneste modulen som er direkte koblet til databasen. De andre modulene vil basere seg på filer som input samtidig som alle resultater skrives til fil. Dette vil gi nødvendig fleksibilitet for forskere, som ikke alltid har tilgang til databasen, til å kunne jobbe isolert med de samme verktøyene som inngår i produksjonsmiljøet, og kun forholde seg til filer. Når dataene skal lagres i databasene, benyttes skript i Python eller modeller i .net som sørger for at resultatene lagres i databasene. Figur 8 viser dataflyten gjennom beregningskjeden.



Figur 4. Dataflyt for beregningsdelen benytter filer for lagring av data. Disse dataene lagres som over videre over i databaser for arkivering og visning.

5.2 Utslippsmodellen

Utslippsmodellen vil bygges opp på nytt fra grunnen av. Den vil basere seg på AirQUIS og tidligere utviklet skriptbaserte modeller. Utslippsmodulen vil hente data direkte fra AirQUIS-databasen som underlag for å beregne timesverdier som skrives til NetCDF format. Modellen vil i hovedsak være skrevet i Python, men Fortran kan også benyttes i deler av modellen hvor det krever raskere prosessering.

5.3 Spredningsmodellen

EPISODE er en modell som beregner spredningen av luftforurensning for urbane og lokale områder. Modellen benyttes til å beregne forurensningskonsentrasjonen fra flere kilder som biltrafikk, fyring og industri samtidig i byer- og urbane områder. Modellen er utviklet i Fortran og kjører per i dag på Windows, men det vil også være mulig å «porte» over til Linux for å kunne tilby en multiplattform løsning.

Modellen i dagens form leser data fra ASCII-filer eller direkte fra databasen i AirQUIS-systemet. Dette må endres slik at EPISODE benytter en struktur hvor man jobber direkte mot filer i NetCDF format gjennom hele prosesseringskjeden.

5.4 Statistiske beregninger

Statistiske modeller vil være matematiske beregninger av statistiske parametere, som for eksempel døgnmiddel, månedsmiddel og årsmiddel. Modellene vil lese inn timesverdier fra filer på NetCDF format og levere resultater på NetCDF format. Modellen vil i hovedsak være skrevet i Python-skript, men Fortran kan også benyttes i deler av modellene hvor det krever raskere prosessering.

5.5 Eksponering

Eksponeringsmodulen er beregningsmodellen for å beregne antall skadelidende personer som eksponeres for ulike verdier av luftforurensning. Modellen vil lese inn timesverdier fra filer på NetCDF format og levere resultater på NetCDF format. Modellen vil i hovedsak være skrevet i Python-skript, men Fortran kan også benyttes i deler av modellen hvor det krever raskere prosessering.

5.6 Datautvekslingsformat NetCDF

Dette dataformatet har mange fordeler og er blant annet strukturert med en navnekonvensjon som gjør at software automatisk kan lese datastrukturens enheter ved å finne variable navn og tilhørende metadata. Det er bred støtte for formatet i alle store programmeringsspråk, som for eksempel C#, C/C++, JAVA og Python. NetCDF støttes også av relevante visualisering- og dataanalyse-verktøy som MATLAB, R og IDL.

5.7 Eksport av resultater til databasene

Resultater som skal presenteres på web og/eller arkiveres, må skrives til databasene. Nasjonalt beregningsverktøy benytter to databaser for lagring. Den generelle databasen som også benyttes til lagring av inngangsdata, vil også inneholde resultater. I tillegg vil det benyttes en egen GIS-database (PostGreSQL) for resultatene som skal vises på web.

6 Visualisering

6.1 Definerte visninger

Følgende visninger er definert i prosjektet:

1. Verktøyet skal gi informasjon om **årlige status** av den geografiske utbredelsen av luftforurensning i forskjellige områder.
2. Verktøyet skal vise **luftsonekart** etter krav fra regelverket i norske byer og tettsteder.
3. Verktøyet skal vise **framskrivninger** av den geografiske utbredelsen av luftforurensning i forskjellige områder.
4. Verktøyet skal vise **tiltaksberegninger** for konsekvensutredninger etter krav fra regelverket i norske byer.
5. Verktøyet skal gi informasjon om **kildeallokering** etter krav fra regelverket i forskjellige områder.
6. Verktøyet skal vise **varslet luftkvalitet** for informasjon ved brudd på alarmterskelen, og støtte gjennomføring av kortsiktige tiltak i utvalgte områder.
7. Verktøyet skal vise **eksponering** for luftkvalitet, i første omgang beregnet for antall personer utsatt for skadelige nivåer av luftforurensning i utvalgte områder. Beregningene benyttes i forbindelse med planlegging og konsekvensanalyser.

I Tabell 2 ser man de ulike visningene i forhold til de fire kategoriene, hele Norge, norske byer, norske tettsteder og norsk Industri. De blir viktig at de ulike modulene kan gjenbrukes på tvers av disse kategoriene slik at brukeren kan sammenlikne resultater og kjenne seg igjen.

Tabell 2: Tabellens visninger i Nasjonalt beregningsverktøy og når de implementeres i verktøyet 2015 (rosa), 2016 (grønn), videre arbeid i 2017 og utover (lilla).

	Geografisk utbredelse Årlig status	Framskrivninger	Luftsonekart	Tiltaksberegning	Kildeallokering	Varsling	Helseeksponering
Hele Norge	Bakgrunnsatlas i 10x10 km ² Luftkvalitet.info (målinger)		Identifisering av nye byer utover Bedre Byluft som skal ha luftsonekart			MACC-II Hele Norge i 10x10 km ²	Gjennom bakgrunnsatlas i 10x10 km ²
Norske byer	7 byer (inkl. Bedre Byluft) 1x1 km ² Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland	Alle byene i Bedre Byluft 1x1 km ² Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland Scenarier 2020-25-30	Oslo/Bærum Stavanger området Sarpsborg/ Fredrikstad Trondheim + alle byene i Bedre Byluft i 2016	Alle byene i Bedre Byluft 1x1 km ² Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland On-demand online	7 byer 1x1km ² Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland Sarpsborg/ Fredrikstad	Bedre Byluft Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland	Alle byene i Bedre Byluft 1x1 km ² Oslo/ Bærum Drammen Bergen Stavanger/ Sandnes Trondheim Grenland Sarpsborg/ Fredrikstad
Norske tettsteder	FM trafikk + vedfyring + industri	FMT trafikk + vedfyring + industri + tiltak	Identifisering av nye fare soner	FMT trafikk + vedfyring + industri + tiltak	FM trafikk + vedfyring + industri	Gjennom MACC-II og VLUFT (døgn+ timer)	Gjennom hele Norgesatlas (årlig)
Industrianlegg	Konsulentoppdrag ModLUFT	Konsulentoppdrag-Framskrivninger	n.a.	On-demand online Punktkilder	Gjennom hele Norgesatlas	Gjennom MACC-II	Gjennom hele Norgesatlas

Det foreslåes å organisere arbeidet i fire områdetyper som har forskjellige detaljeringskrav vedrørende informasjon om luftkvalitet.

De fire områdetypene er:

1. Hele Norge eller landsdekkende informasjon.
2. Norske byer eller områder med krav til tiltaksutredninger etter regelverket.
3. Norske tettsteder eller områder uten krav til tiltaksutredninger.
4. Industri- eller anleggsområder for konsesjonsoppgaver.

For områdetype 3 og 4, Norske tettsteder og Industri, vil resultatene genereres i modellverktøyet, mens for hele Norge og Norske byer, vil det bli generert som kartbaserte visninger i tillegg tabeller og visuelle diagrammer som kakediagrammer.

6.1.1 Visninger for hele Norge

1. Geografisk utbredelse. Årlig status. 1 x 1 km. Status. Bakgrunnsatlas i 10 x 10 km²
2. Varsling
3. Helseeksponering.

6.1.2 Visninger for norske byer

1. Geografisk utbredelse, årlig status. 1 x 1 km
2. Luftsonkart
3. Kildeallokering
4. Fremskrivninger
5. Eksponering

6.2 Webløsning

Løsningen vil i første fase i stor grad basere seg på å visualisere resultater fra de ulike modellene og inngangsdata brukt i modellene. Webløsningen vil bygges rundt et kartbasert grensesnitt hvor man kan navigere og analysere resultatene basert på sted.

Løsningen vil være basert på MVC.NET, HTML5, CSS og JavaScript slik at den kan brukes på alle relevante plattformer som støttes av moderne web-browsere. Det er et krav at webløsningen skal være responsiv, det vil si at den kan vises på alle flater, mobil, nettbrett og PC.

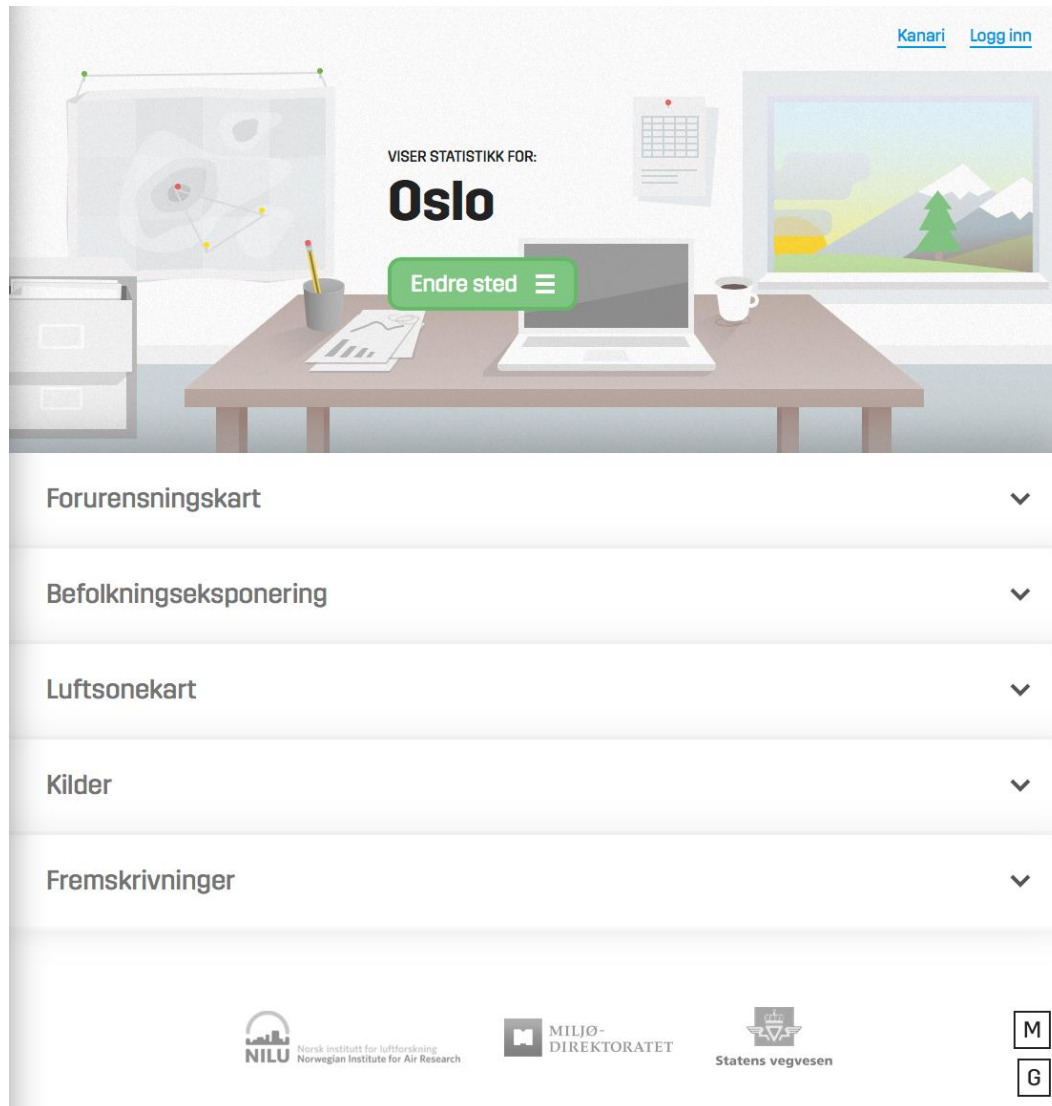
I prosjektfasen vil vi også se nærmere på hvordan NBV-design og visualiseringsløsning vil kunne gjenbrukes i luftkvalitet.info. Det vil blant annet utredes hva som skal til for å få en felles plattform og design for de to webløsningene.

6.3 Web design og prototyp

I samarbeid med arbeidsgruppen bestående av representanter for Miljødirektoratet, Statens vegvesen og NILU, har Netlife Research utarbeidet designet til en prototype som tilfredsstillende krav til løsning. Designet er responsivt og fungerer på alle flater. I prototypen inngår også nødvendige designelementer som knapper samt fargepaletter som vil være byggeklosser for å utforme systemet.

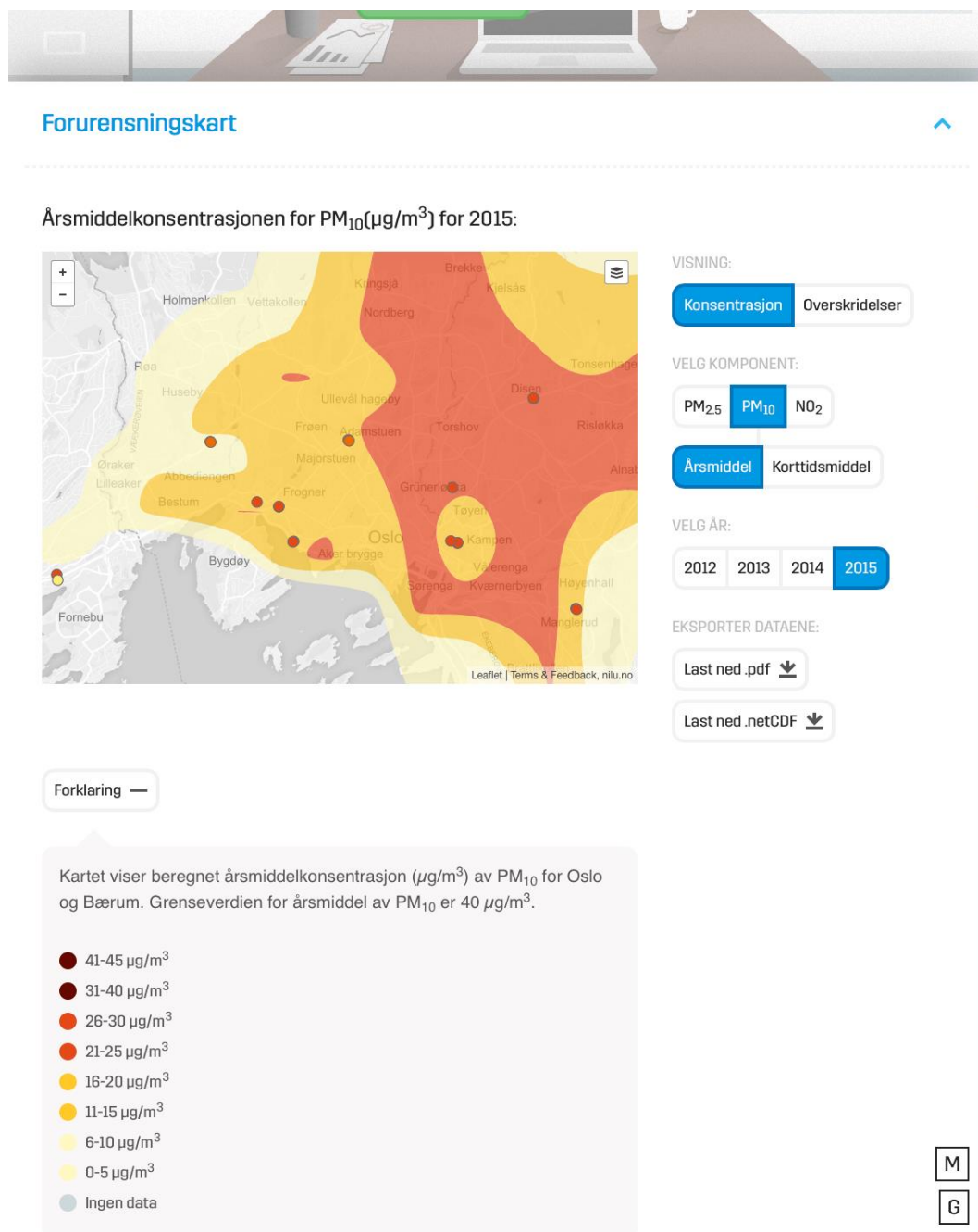
Løsningen er modulbasert og gir stor fleksibilitet i forhold til å legge til nye moduler. Dette innebærer at hver visning blir laget som en separat modul som er uavhengig av de andre modulene og fritt kan endres. Det er i løsningen lagt opp til et system hvor brukeren scroller seg gjennom menyen. Ved å trykke på ønsket valg vil modulen ekspandere.

6.3.1 Navigasjon og menysystem



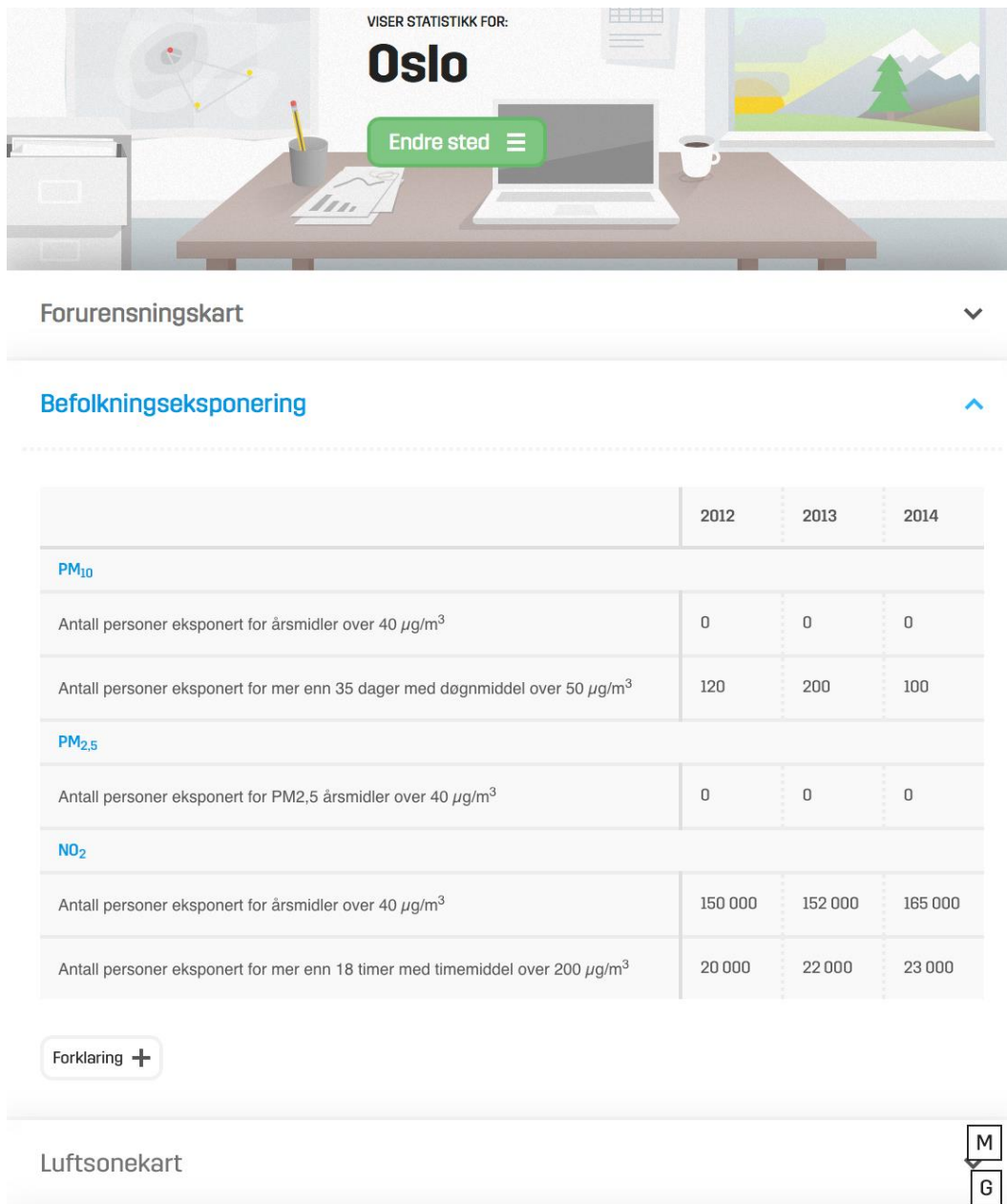
Figur 5. Menyene er bygd opp vertikalt slik at man enkelt kan «scrolle» igjennom den også på mindre enheter som mobiltelefoner. Ved å klikke seg inn, vil menyene ekspandere til den valgte modulen.

6.3.2 Forurensningskart



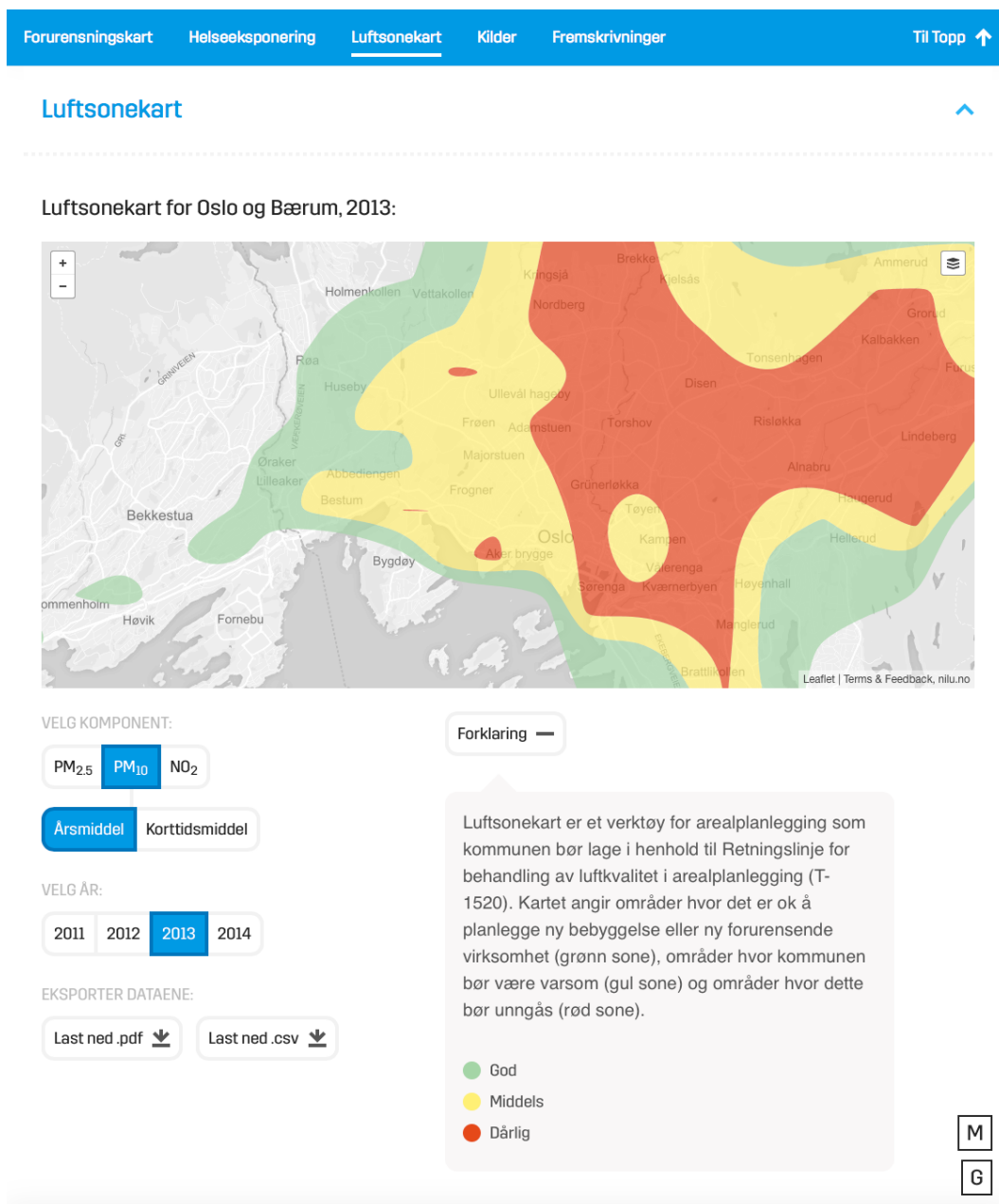
Figur 6. Forurensningskartet viser årsmiddelkonsentrasjonen for Norge og utvalgte domener.

6.3.3 Befolkningseksposering



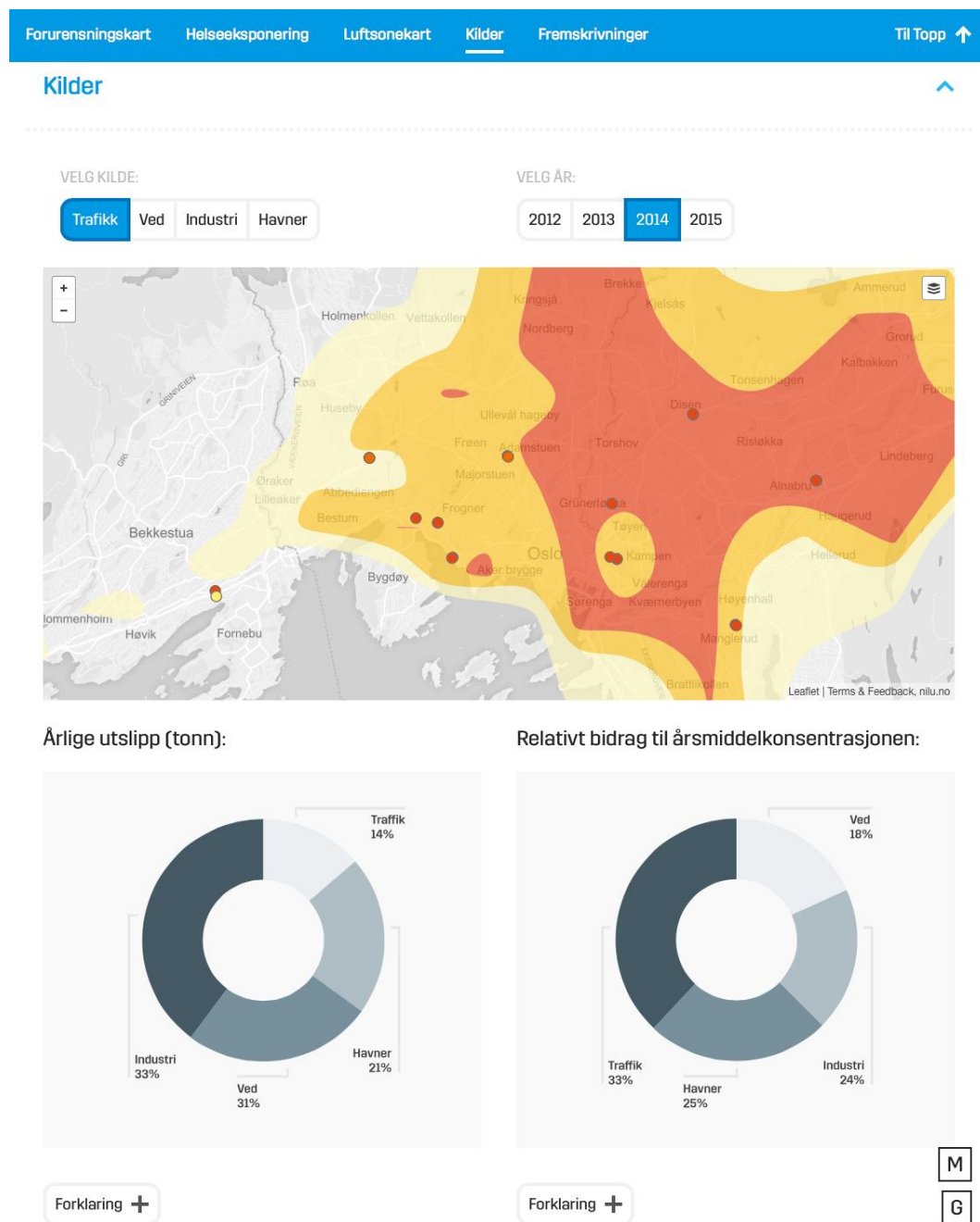
Figur 7. Befolkningseksposering viser i tabellform antall eksponerte mennesker.

6.3.4 Luftsonekart



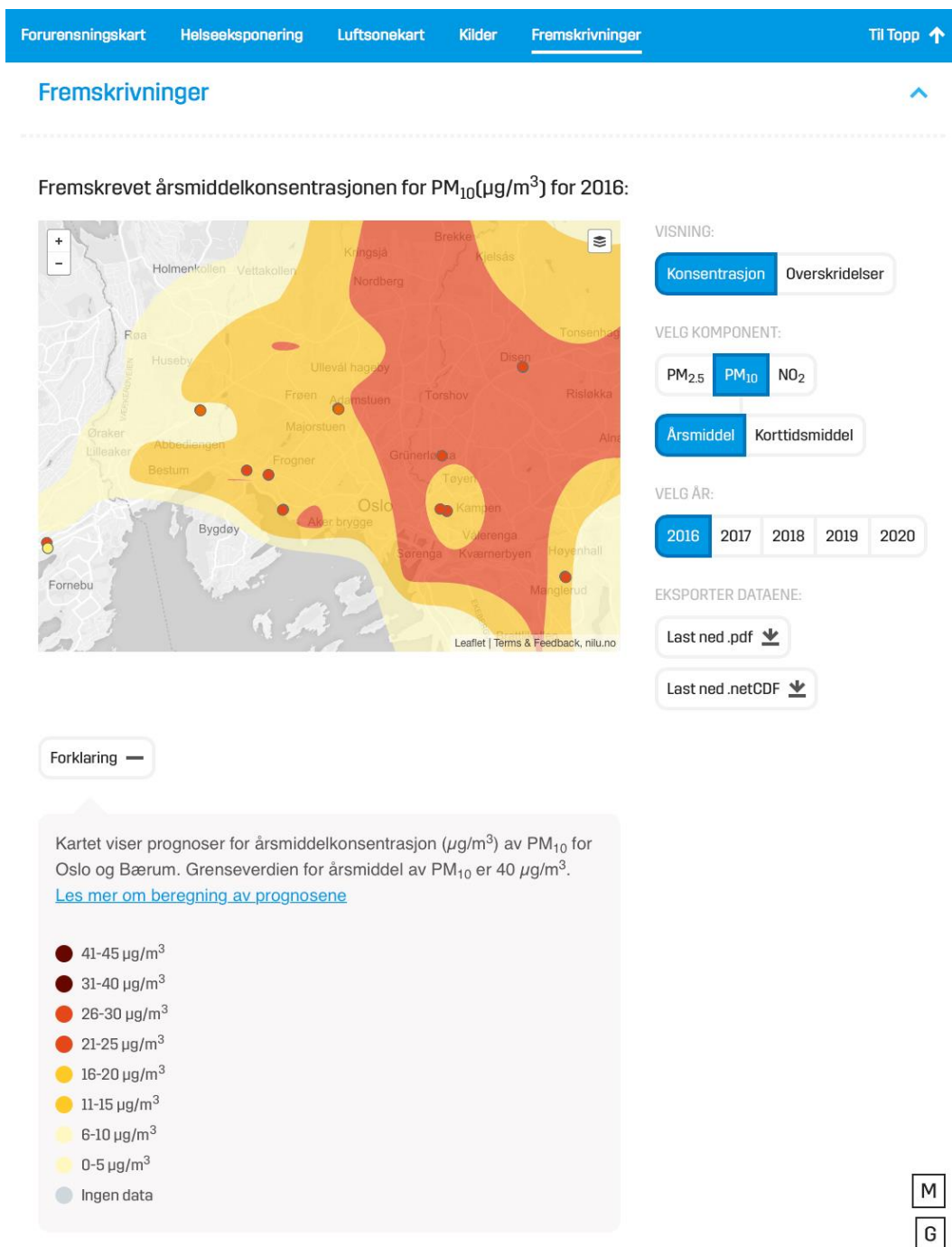
Figur 8. Luftsonekart viser geografisk utbredelse.

6.3.5 Kilder



Figur 9. Kilder, viser relative bidrag til årsmiddelkonsentrasjonene. Brukeren kan i kartet velge lokasjon og få opp spesifisert bidrag beregnet ved valglokasjon.

6.3.6 Fremskrivninger

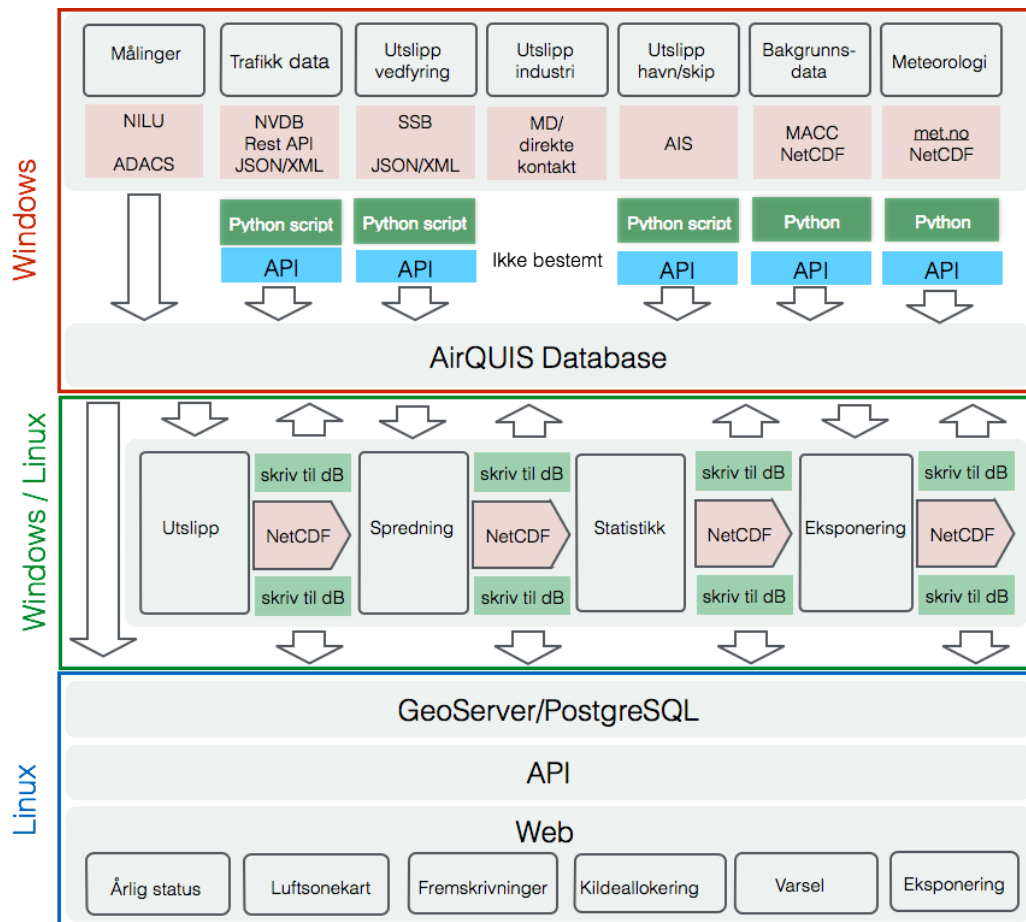


Figur 10 Fremskrivninger viser prognose for årsmiddelkonsentrasjonen i form av geografisk utbredelse.

7 Infrastruktur

7.1 Krav til infrastruktur

Nasjonalt beregningsverktøy vil kjøre på to plattformer. Database og modeller vil kjøre på en Windows-plattform, mens visningslaget vil kjøre på en Linux-løsning. Alle modulene i beregningskjeden (utslipp, spredning, statistikk og eksponering) vil bli utviklet som multiplattformer. Dette gir oss god fleksibilitet, og vil gi hensiktsmessige ytelsesvurderinger for valg av plattformer som henholdsvis Linux eller Windows. Figur 9 viser fordelingen slik den er tenkt i dag. I utviklingsfasen av NBV vil vi benytte NILUS eksisterende infrastruktur.



Figur 11. Oppdeling mellom ulike plattformer brukt i systemet.

Modulene i beregningskjeden (utslipp, spredning, statistikk og eksponering) vil bli utviklet slik at de kan fungere på ulike plattformer. Hvis det er hensiktsmessig kan dette flyttes over på Linux. Inngangsdata fra Statistisk Sentralbyrå (SSB, Miljødirektoratet (MD), Meteorologisk institutt (met.no) og Nasjonal veg database (NVBD).

7.2 Prosesseringsbehov

Kravene til prosessering er forholdsvis beskjedne før systemet settes ut i drift. Systemet vil i testfasen kjøre på NILU sin eksisterende infrastruktur.

7.3 System og infrastruktur for visualiseringer

Presentasjonslaget vil hovedsakelig bestå av fire komponenter; en PostgreSQL-database en med PostGIS-utvidelse for lagring av visningsdata, en GeoServer for kartvisualiseringer, en backend-løsning for å levere data til klientsiden, og en webløsning basert på HTML5, CSS og JavaScript.

7.3.1 PostGIS

PostGIS er en romlig databaseutvidelse til PostgreSQL, og er en objektsrelasjons-database. PostGIS tilbyr støtte for geografiske objekter og lokasjonsbaserte spørringer til databasen.

7.3.2 GeoServer

Geoserver er en åpen kildekode-server som brukes for å dele og editere romlig data. Den støtter diverse romlige dataformater basert på åpne standarder. Serveren vil være bakendeløsningen visualiseringen på web mot frontend som OpenLayer eller Leaflet.

7.4 Lagringsbehov

Det er identifisert et lagringsbehov på rundt 3.0 TB pr. år. For å ha rom for endringer, er det ønsket å skalere systemet til 6 TB per år. Dette vil naturligvis påvirkes av hvor mye systemet blir brukt og antall beregninger som blir lagret. Det vil være viktig at systemet er skalerbart slik at man kan ta vare på beregninger for flere år tilbake samtidig som man fyller på med nye inngangsdata og utfører nye beregninger. I tabell 3 og 4 vises fordelingen av indentifisert lagringsbehov for hele Norge og de syv definerte domener i Norge.

Tabell 3. Lagringsbehov for beregninger og visning av data for hele Norge

Datatype	Behov per år	Totalt behov for utviklingsfasen
Bakgrunnsatlas (hele Norge)	50 MB	100MB
Tidsserier per punkt (hele Norge)	200 MB	400MB

Tabell 4. Lagringsbehov for beregninger og visning for de syv definerte domeneene.

Datatype	Behov per år	Totalt behov for utviklingsfasen
Rå meteorologi	55 GB	110 GB
Preprosessert meteorologi	120 GB	240 GB
Utslippsdata	50 GB	100 GB
Rå bakgrunnsdata	220 GB	440 GB
Preprosessert bakgrunnsdata	112 GB	224 GB
Sprednings	480 GB	960 GB
Backend Oracle DB	360 GB	720 GB
Frontend PostgreSQL DB	2 GB	4 GB

7.5 Skyløsning

Det er nærliggende å tenke seg en skyløsning for både prosesseringskjeden og visningene. Sky-løsninger som blant annet Microsoft AZURE og Amazon elastic cloud har blitt vurdert. I tillegg vil det i relativ nær fremtid, komme flere solide norske aktører på markedet. Da det ligger noen år frem i tid til systemet settes i drift, og fordi det er en rask utvikling på dette feltet, vil det være fornuftig å vurdere dette på et senere tidspunkt. Vurderinger som bør legges til grunn ved valg av skyløsning vil blant annet være:

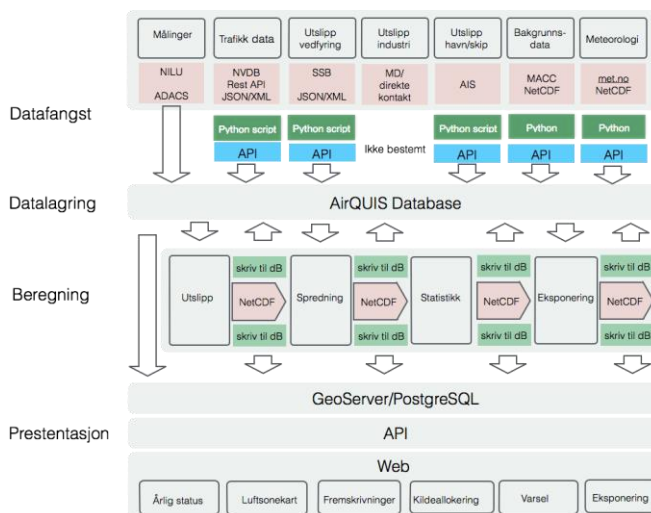
- Skalerbar ytelse, for prosessering og dataoverføring
- Skalerbar lagring (estimert behov 10-15 TB pr. år)
- Tilgjengelighet. Hvilke krav vi har til oppetid?
- Sikkerhet mot hacking.
- Backup-løsning for data
- Hvor dataene skal lagres

8 Åpen kildekode

Det er et krav at løsningen er åpen. Dette vil innebære at all kildekode utviklet for innhenting av data, kildekode for modeller brukt i løsningen og kildekode til visualisering, blir gjort fritt tilgjengelig. Dette er viktig med henblikk på at verktøyet blir tilgjengelig og fritt kan brukes og videreutvikles av konsulenter og forvaltningen.

All kildekode som i NBV blir frigitt som åpen kildekode, vil bli underlagt lisensen «Reciprocal Public License» Versjon 1.5 (RPL 1.5).


- Alle skript i produksjonskjeden vil være tilgjengelig.
- Alle APIer vil være tilgjengelige, men beskyttet.
- Beregningsmodulene blir frigitt som åpen kildekode. Lisensert under RPL 1.5.
- PostgreSQL og Geo Server er åpen kildekode.
- All data vil være tilgjengelig gjennom åpene APIer.
- All kode i presentasjonslaget vil være åpen. Mulig lisens er Creative Commons



8.1 Utviklingspråk og åpen kildekode

I Nasjonalt beregningsverktøy vil det gjennomgående bli benytte programmeringsspråk som er åpen kildekode.

- **Python** vil benyttes for skripting og vil være limet som binder det hele sammen.
- **Fortran** vil benyttes i beregningsmodulene hvor krav til ytelse er meget viktig.
- **.net**, som i sin helhet nylig er frigjort fra Microsoft som åpen kildekode, vil benyttes når det er behov for kommunikasjon med databasene.
- **Javascript /html5** vil benyttes i visualiseringslaget.

RAPPORTTYPE OPPDRAGSRAPPORT	RAPPORT NR. OR 13/2015	ISBN: 978-82-425-2770-7 (trykt) 978-82-425-2771-4 (elektronisk) ISSN: 0807-7207	
DATO 30/06/2015	ANSV. SIGN 	ANT. SIDER 25	PRIS NOK 150.-
TITTEL Teknisk løsning for nasjonalt beregningsverktøy Dataflyt, systemdesign og krav til infrastruktur		PROSJEKTLEDER Leonor Tarrasón	
		NILU PROSJEKT NR. O-114092	
FORFATTER(E) Morgan Kjølervbakken og Islen Vallejo		TILGJENGELIGHET * A	
		OPPDRAGSGIVERS REF. Isabella Kasin	
KVALITETSSIKRER: Leonor Tarrasón			
OPPDRAGSGIVER Miljødirektoratet, Statens vegvesen			
REFERAT Rapporten beskriver dataflyt, systemdesign og krav til infrastruktur for et system som modellerer og beregner luftkvalitet. Rapporten inneholder også beskrivelse av brukergrensesnittet i en webløsning.			
STIKKORD Dataflyt	Systemdesign	Beregningsverktøy	
TITLE System design NBV			
ABSTRACT System design for «nasjonalt beregningsverktøy» describing data flow and infrastructure requirements			

* Kategorier A Åpen – kan bestilles fra NILU
 B Begrenset distribusjon
 C Kan ikke utleveres

REFERANSE: O-114092
DATO: JUNI 2015
ISBN: 978-82-425-2770-7 (trykt)
978-82-425-2771-4 (elektronisk)

NILU – Norsk institutt for luftforskning er en uavhengig stiftelse etablert i 1969. NILUs forskning har som formål å øke forståelsen for prosesser og effekter knyttet til klimaendringer, atmosfærens sammensetning, luftkvalitet og miljøgifter. På bakgrunn av forskningen leverer NILU integrerte tjenester og produkter innenfor analyse, overvåkning og rådgivning. NILU er opptatt av å opplyse og gi råd til samfunnet om klimaendringer og forurensning og konsekvensene av dette.