

RAPPORT – ARBEIDSGRUPPE 1 - GEODATASJØ





INNHOLD

SAMMENDRAG AV RESULTATER	3
1. BAKGRUNN	3
2. PROBLEMSTILLING (MANDAT) FOR AP1	4
2.1 Del 1	4
2.1 Del 2	4
3. OPPGAVER – DEFINISJON AV GEODATASJØEN	5
3.1 Datainnhold	5
3.2 Valg av testområder.	6
4. OPPGAVER – ARKITEKTUR	7
4.1 Bakgrunn for foreslått arkitektur	7
4.2 Høynivå konseptskisse	8
4.3 Dataflytløsninger – rådata	9
4.4 Lagringsarkitektur – Datastruktur/modeller/prinsipper (storage)	9
4.4.1 PostGIS - Rådata	9
4.4.2 Blob storage container - Rådata	10
4.4.3 Blob storage container – treningsdata	10
4.4.4 Blob storage container – ML modeller	10
4.4.5 PostGIS - Treningsdata	10
4.4.6 PostGIS – Resultater	10
4.5 Tjenester	11
5. OPPGAVER – DATAFLYT TIL GEODATASJØEN	11
5.1 Høsting av primærdata for testområdene.	11
6. GJENSTÅENDE OPPGAVER	11
6.1 Ikke fullførte oppgaver	11
6.3 Anbefalinger fra arbeidsgruppa	12



SAMMENDRAG AV RESULTATER

Prosjekterfaringer

De interne avhengighetene mellom oppgavene i AP1 og AP2 har vist seg å være omfattende. Dette har stilt til krav til koordinering som er løst gjennom felles workshops og koordineringsmøter. Noen oppgaver er også overført fra AP1 til AP2.

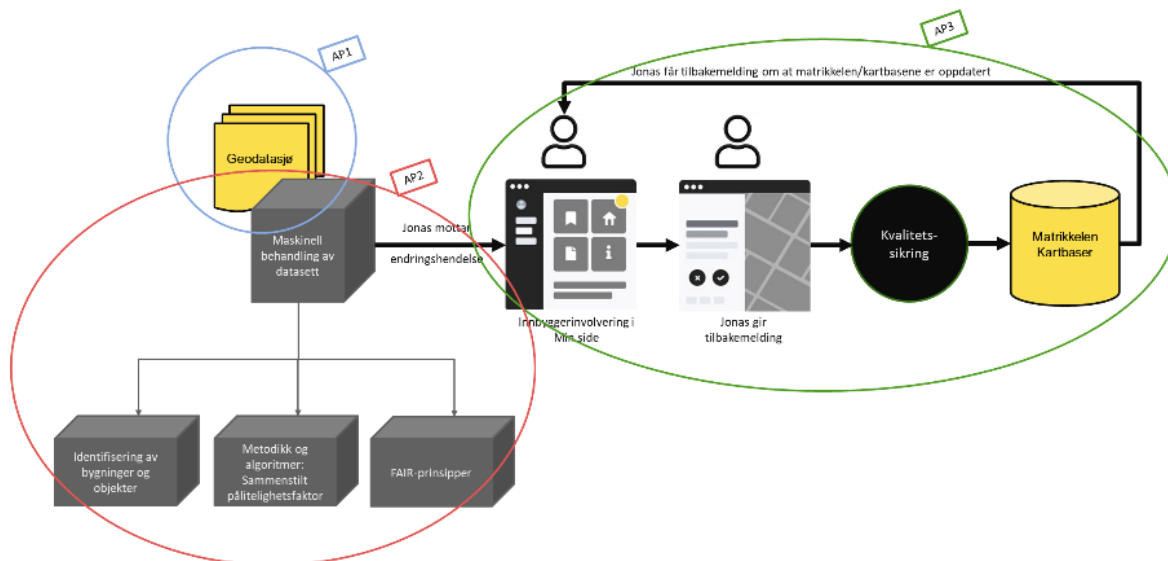
Gjennom prosjektet har en sett behovet for å etablere geodatasjø for å tilgjengeliggjøre dataene for partene i prosjektet. AP2 benytter WMS tjenester for tilgang til ortofoto og det bør vurderes om det er behov for ortofoto i geodatasjøen.

I de tilfellene hvor datakildene er åpne og tilgjengelige på en form som tilfredsstiller prosjektets krav kan behovet være mindre for å inkludere datasettet i prosjektets geodatasjø. Ved å beholde datasettene i åpne tilgjengelige løsninger sikrer en samtidig oppdaterte data. På bakgrunn av dette anbefales det å benytte åpne tilgjengelige datakilder der det er mulig og ikke inkludere slike datasett i datasjøen.

1. BAKGRUNN

Prosjektet har tre delmål:

- Bygninger skal identifiseres ved hjelp av flere forskjellige datakilder som laserdata, flyfoto og crowd-sourcing (digital dugnad). Dette gjøres ved å danne en «geodatasjø» med datakilder som er optimalisert for kunstig intelligens og stordataanalyse.
- Gi et bedre beslutningsgrunnlag for kommunal saksbehandling. Prosjektet skal identifisere bygninger under 50 kvadratmeter i et utvalgt område i Kristiansand ved å bruke og skape nye kunstig intelligens-algoritmer fra geodatasjøen. Typisk vil dette være bygninger som er unntatt søknadsplikt og ikke er registrert fra før.
- Sikre mer effektiv saksbehandling gjennom pålitelig og automatisert kvalitetsheving av det nasjonale eiendomsregisteret (matrikkelen) og kartbaser, i dialog med innbyggerne.



Arbeidsgruppe 1 jobber med arbeidspakke 1 i prosjektet og har fokus på dataflyt, lagring og uthentingsarkitektur for en Geodatasjø.

Arbeidsgruppe 2 i KartAI, jobber med arbeidspakke 2 i prosjektet (innenfor rød sirkel i figuren). Denne arbeidspakken fokuserer på analyser og objekt-identifikasjon i en Geodatasjø.

I første fase av prosjektet blir det mest interaksjon og samordning mellom AP1 og AP2. Etter hvert vil det også være behov for samordning mot AP3.

2. PROBLEMSTILLING (MANDAT) FOR AP1

Sammendrag av arbeidsgruppens mandat

2.1 Del 1

1) Utarbeide forslag til arkitektur og dataflyt som gjør det effektivt å lagre relevante multikilde-data i en geodatasjø basert på moderne cloud-arkitektur. Forslagene skal ta utgangspunkt i prosjektets skisserte brukerbehov og arkitekturbehov fra Arbeidspakke 2. Det må rettes spesiell oppmerksomhet rundt behovene knyttet til lagring og uthenting av treningsdata og nettverk/modeller. Arbeidet dokumenteres i form av figurer og tilhørende korte tekniske beskrivelser (1-3 sider). Presenteres for prosjektgruppe samt Arbeidspakke 2 med beslutning om videreføring.

2) Implementere MVP-løsning (minste fungerende løsning) av arkitektur med relevante datasett for valgt pilotområde til konseptuell bruk i arbeidspakke 2.

2.1 Del 2

3) Utarbeide veiledninger og referanseimplementasjoner for optimaliseringsbehov for et utvalg relevante datasett for bruk i AI, ML, Analyser. Dokumenteres i rapportform 1-3 sider.



4) Gi innspill til gevinstpotensial i en geodatasjø relativt til målsettinger i nasjonal strategi for kunstig intelligens og andre relevante nasjonale føringer. Dokumenteres i rapportform 1-3 sider.

3. OPPGAVER – DEFINISJON AV GEODATASJØEN

3.1 Datainnhold

Innholdet i datasjøen består av Matrikkelen, Felles kartdatabase (FKB), Flybilder, Ortofoto, laser og brukerinvolvering/crowdsourcing. En viktig faktor i prosjektet er å analysere tidsserier for å detektere endringer i bygningsmassen. Ut fra dette kriteriet er det valgt å undersøke datakildene for tidsperioden 2016- 2020 (2021). Følgende datasett er tilgjengelige og inkludert i datasjøen:

Matrikkel:

Prosjektnavn	Referanse dato	Dekning treningsområdet	Dekning prosjektområdet og valideringsområdet
Matrikkel 2020	2021	x	x

FKB (FKB- Bygg, FKB- Tiltak):

Prosjektnavn	Referanse dato	Dekning treningsområdet	Dekning prosjektområdet og valideringsområdet
FKB 2016	2016-12- 31	x	x
FKB 2017	2017-12- 31	x	x
FKB 2018	2018-12- 31	x	x
FKB 2019	2019-12- 31	x	x
FKB 2020	2020-12- 31	x	x



Flyfoto / Ortofoto:

Prosjektnavn	Dekningsnummer	GSD ortofoto(m)	Overlapp % (lengde/bredde)	Referanse-dato	Dekning treningsområdet	Dekning prosjektområdet og valideringsområdet
KRISTIANSAND SENTRUM 2019	CO12167	0.08		2019-04-08	x	x
KRISTIANSAND 2018	TT-14335	0.1	60/30	2018-05-08	x	x
KRISTIANSAND VENNESLA 2017	TT-14268	0.1	60/20	2017-04-08		x
KRISTIANSAND 2016	TT-14215	0.08	60/20	2016-04-09	x	

Flybåren laser:

Prosjektnavn	Dekningsnummer	Punkt tetthet	Overlapp (lengde/bredde)	Referanse-dato	Dekning treningsområdet	Dekning prosjektområdet og valideringsområdet
KRISTIANSAND 2020	41250	5 pkt/m2 (full waveform)	*/20	2020-04-06	x	x

3.2 Valg av testområder.

En av målsetningene til KartAI er å finne uregistrert byggeaktiviteter som kan være relevant for eiendommens utnyttelsesgrad. De ulike scenario som en ønsker å detektere er uendret bygning, tilbygg,

nytt bygg og revet bygg. Ut fra senarioene og lokal kjennskap fant Kristiansand kommune aktuelle områder ut fra følgende kriterier:

- Område bestående i hovedsak av eneboliger og tomannsboliger, men med innslag av annen bebyggelse.
- Tomtestørrelser som gir anledning for byggfortetting
- Et etablert område med noe fortetting i senere år
- Område med generasjonsskifte.

På bakgrunn av dette ble følgende områder valgt:



Figur 1. Inndeling av områder og bruksområder. Treningsområdet for maskinlæring (grønt). Valideringsområdet for justering av hyperparametre (blått) og Prosjektområdet for dokumentasjon av oppnådd resultat (rødt).

De utvalgte testområdene er tilpasset behovene for maskinlæring og kunstig intelligens. Det er tette boligstrøk med noenlunde lik bebyggelse i ulike bydeler i Kristiansand. Det er viktig at man trener på treningsområde, justerer på valideringsområdet og til slutt henter ut statistikk på oppnådd resultat på prosjektområdet.

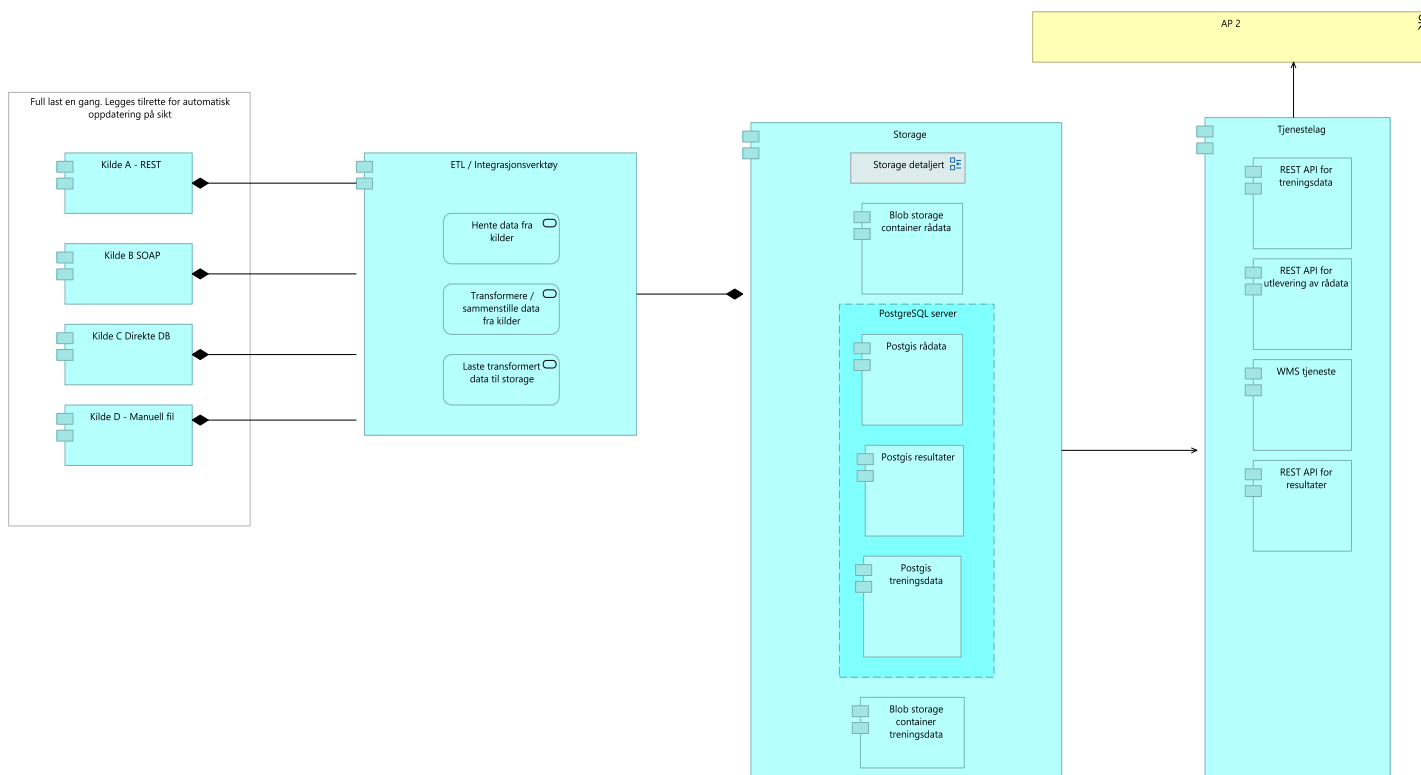
4. OPPGAVER –ARKITEKTUR

4.1 Bakgrunn for foreslått arkitektur

Bakgrunnen for den foreslåtte arkitekturen er å skissere en løsning som gjør det effektivt å innhente, lagre og tilgjengeliggjøre både data fra primærkilder, men også maskinlæringsdata som modeller, treningsdata og resultater generert av AP2.

Arkitekturen bygges i skyen med en modularisert og lagvis løsning som består av flere løst koblede komponenter. Videre skal løsningen bygges med et tjenesteorientert fokus, som vil si at data utleveres og lagres gjennom dedikerte applikasjonstjenester. I skyen velges fortrinnsvis SaaS og PaaS tjenester for å redusere vedlikeholdsarbeid og effektivisere konfigurasjon.

4.2 Høynivå konseptskisse



Skissen viser på et overordnet nivå hvilke komponenter som inngår i løsningen, og hvordan de henger sammen, for å realisere behovene for innhenting, oppbevaring og tilgjengeliggjøring av ulike typer data.

Helt ute til venstre i konseptskissen finnes det kilder med fire forskjellige grensesnitt, REST, SOAP, direkte database tilkobling, og manuelle filer. Skulle en ønske å automatisere innhenting av rådata må en eventuell integrasjonsplattform forholde seg til de ovennevnte grensesnittene. En slik integrasjonsplattform bør og ha mulighet til å transformere data til ønsket format, og lagre dataen videre i KartAIs egne lagringsmedier.

I skissen er det lagt opp til to typer lagringstjenester, postgis og blob storage container. Data som kommer strukturert fra kilden beholdes strukturert, og lagres dermed i en postgis database. Matrikel og FKB er strukturerte data som skal lagres i en egen postgis database for rådata. Ustrukturerte data som flyfoto, laserdata og ortofoto lagres i en egen blob storage container for rådata.

For å få tilgang på disse dataene skal man enten gjennom et dedikert REST API eller en WMS for utlevering av rådata. Ved å definere slike tjenester kan man gi ut data etter den formen/modellen man ønsker samtidig som man har kontroll på dataen.

Maskinglæringsdata fra AP2 lagres i egne blob storage containere og postgis databaser. Også på toppen av disse dataene foreslås det å bruke dedikerte tjenester for inn- og utlasting av data.



4.3 Dataflytløsninger – rådata

AP1 har ikke designet en detaljert løsning for automatisert flyt av rådata til geodatasjøen. Teamet foreslår en løsning basert på standard ETL verktøy. Muligheten for automatisering vil avhenge av om forvaltningssystemene (SFKB, Matrikkel, NorgelBilder, hoydedata.no) eller andre dataverter kan tilby tjenester som dekker prosjektets behov. Dette må undersøkes før man går videre med detaljert design/realisering.

Dataflyten som er laget for å hente gjeldende FKB og Matrikkeldata kan med relativt enkle grep automatiseres (settes opp i schedule) eller trigges gjennom api-kall.

4.4 Lagringsarkitektur – Datastruktur/modeller/prinsipper (storage)

4.4.1 PostGIS - Rådata

FKB og matrikkeldata:

Datasettene er lagret i PostGIS, separate schema for hvert datasett

20210528_matrikkel - inneholder alle data som kommer fra matrikkelen

20210528_fkb - inneholder alle data som kommer fra fkb

Ved opprettelse av ny versjon av dataene lages nye schema prefixet med dato for når uttrekket ble gjort, f.eks vil kjøring 25 juni gi 20210625_matrikkel og 20210528_fkb

Hvert schema inneholder en tabell 'datastore' med alle dataene i følgende felter:

Feltnavn	Type	Beskrivelse
id	uuid	Unik id
attribs	jsonb	Alle attributter lagret i jsonb
change_date	date	Dato for når objektet ble oppdatert (mottatt hos Norkart)
omraade	text	Holder tre verdier: 'trening', 'prosjekt', 'validering'
komm	text	kommunennummer
areal	int	objektets areal
geom	geography	objektets geometri (epsg:4326)

I tillegg til rådatatabellen er det laget 3 materialized views i fkb-schemaet, og 1 i matrikkel- schemaet altså ett view for hver av de 4 objekttypene vi har tatt med inn.. Disse lager en enkelt tabellstruktur med noen



få nøkkelegenskaper hentet fra datastore, og er indeksert for raskere og enklere tilgang. Disse viewene kan utvides /endres ut fra AP2 sine behov

FKB-views:

- annenbygning (Bygninger uten bygningsnummer, ikke matrikkelført)
- bygning (Matrikkelførte bygg)
- pbltiltak (Tiltak som ikke er påbegynt eller er under oppføring)

Matrikkel-views:

- bygningspunkt (Bygningspunkter i matrikkelen)

4.4.2 Blob storage container - Rådata

Ortofoto:

Lagres som bildefiler i geodatasjøens BLOB storage. AP2 har ansvar for innhenting av ortofoto fra WMS tjenester og lagring av bildefiler.

Laserdata:

Lagres som filer på LAZ format i geodatasjøens BLOB storage

«Rå» flybilder:

Lagres som filer på tiff format + egne filer for georeferering i geodatasjøens BLOB storage

4.4.3 Blob storage container – treningsdata

Treningsdata produsert av AP2 får sin egen blob storage container. Ustrukturert treningsdata lagres her.

Treningsdata som er tilrettelagt av Kristiansand kommune (digitaliserte takflater / bildefiler)

4.4.4 Blob storage container – ML modeller

Maskinlæringsmodeller produsert av AP2 får sin egen blob storage container.

4.4.5 PostGIS - Treningsdata

Treningsdata produsert av AP2 lagres på egen PostGIS database.

4.4.6 PostGIS – Resultater

Strukturerte resultater av maskinlæringen lagres på egen PostGIS database.



4.5 Tjenester

Krav til tjenestelag for geodatasjøen må ta utgangspunkt i AP2 sine behov og detaljert design og realisering er overført til AP2.

5. OPPGAVER – DATAFLYT TIL GEODATASJØEN

5.1 Høsting av primærdata for testområdene.

Høsting av data til geodatasjøen for testområdene er i hovedsak gjort manuelt ved innhenting av filer som er kopiert til geodatasjøens BLOB storage. FKB og matrikkeldata er i tillegg lest inn geodatasjøens PostGIS base.

Følgende primærdatasett er etablert:

- Ortofoto
AP2 baserer seg på bruk av WMS tjeneste for tilgang til ortofoto. Bildene lagres i geodatasjøens BLOB storage som filer inndelt i kartblad.
- FKB
5 års versjoner av FKB bygning og tiltak på SOSI format er hentet fra Kartverket og lagt inn i geodatasjøens PostGIS database.
- Matrikkel
Nåsituasjonen for bygninger er hentet fra matrikkelen og lagt inn i geodatasjøens PostGIS database.
- Laserdata
LAZ filer er hentet fra hoydedata.no og lagret i geodatasjøens BLOB storage.
- «Rå» flybilder
Er hentet fra sentralarkivet for flyfoto og lagret i geodatasjøens BLOB storage.

6. GJENSTÅENDE OPPGAVER

6.1 Ikke fullførte oppgaver

Teamet har ikke arbeidet med del 2 av mandatet og dette må gjøres i en videreføring av prosjektet.

Tjenestelag / AI'er mot storage er ikke detaljspesifisert. Vi foreslår at dette overføres til AP1

Dataflyten for primærdata til geodatasjøen er ikke automatisert.

Tilrettelegge databasestruktur for lagring av info om bygg som er identifisert fra ustrukturerte datakilder er en oppgave som gjenstår. Informasjonen benyttes som grunnlag for endringsanalysen.

Different data sources establish whether the building exists

Cadaster	National Map Database	Arial Images	Ortophoto	Lidar	Public respond
Yes/no	2016: yes/no	2016: yes/no	2016: yes/no	2020: yes/no	yes/no
date	2017: yes/no	2017: yes/no	2017: yes/no		
areal	2018: yes/no	2018: yes/no	2018: yes/no		
	2019: yes/no	2019: yes/no	2019: yes/no		
	2020: yes/no				

Figuren gir en skjematisk fremstilling av de ulike datakildene hvor enkelte av kildene har flere tidsepoker.

De ulike datakilder og tidsepoker tilegnes unik uuid. Deretter vil endringsanalysen i AP2 tilegne en overordnet unik bygningsID og sannsynliggjøre hvilke unike uuid som tilhører hvert enkelt bygningsobjekt.

6.3 Anbefalinger fra arbeidsgruppa

Endret oppgavefordeling mellom AP1 og AP2

Med utgangspunkt i erfaringene med at de to teamene sine oppgaver henger tett sammen foreslår vi at arbeidet med ytterligere detaljering av storage design og etablering av tjenester på storage overføres til AP1 sammen med aktuelle ressurser.

Revurdere innhold i geodatasjøen

I de tilfellene hvor datakildene er åpne og tilgjengelige på en form som tilfredstiller prosjektets krav er behovet mindre for å inkludere datasettet i prosjektets datasjø. Ved å beholde datasettene i åpne tilgjengelige løsninger sikrer en samtidig oppdaterte data. På bakgrunn av dette anbefales det å benytte åpne tilgjengelige datakilder der det er mulig og ikke inkludere slike datasett i datasjøen.

Vurdere automatisering av dataflyt til geodatasjøen

Høsting av primærdata til geodatasjøen er så langt i prosjektet gjort «manuelt». Vi foreslår en kartlegging av hvilke primærdatasett som er tilgjengelige via tjenester som gir mulighet for automatisert høsting. Dette brukes som grunnlag for å lage en automatisert løsning. Inkludert i dette ligger også valg av ETL verktøy.