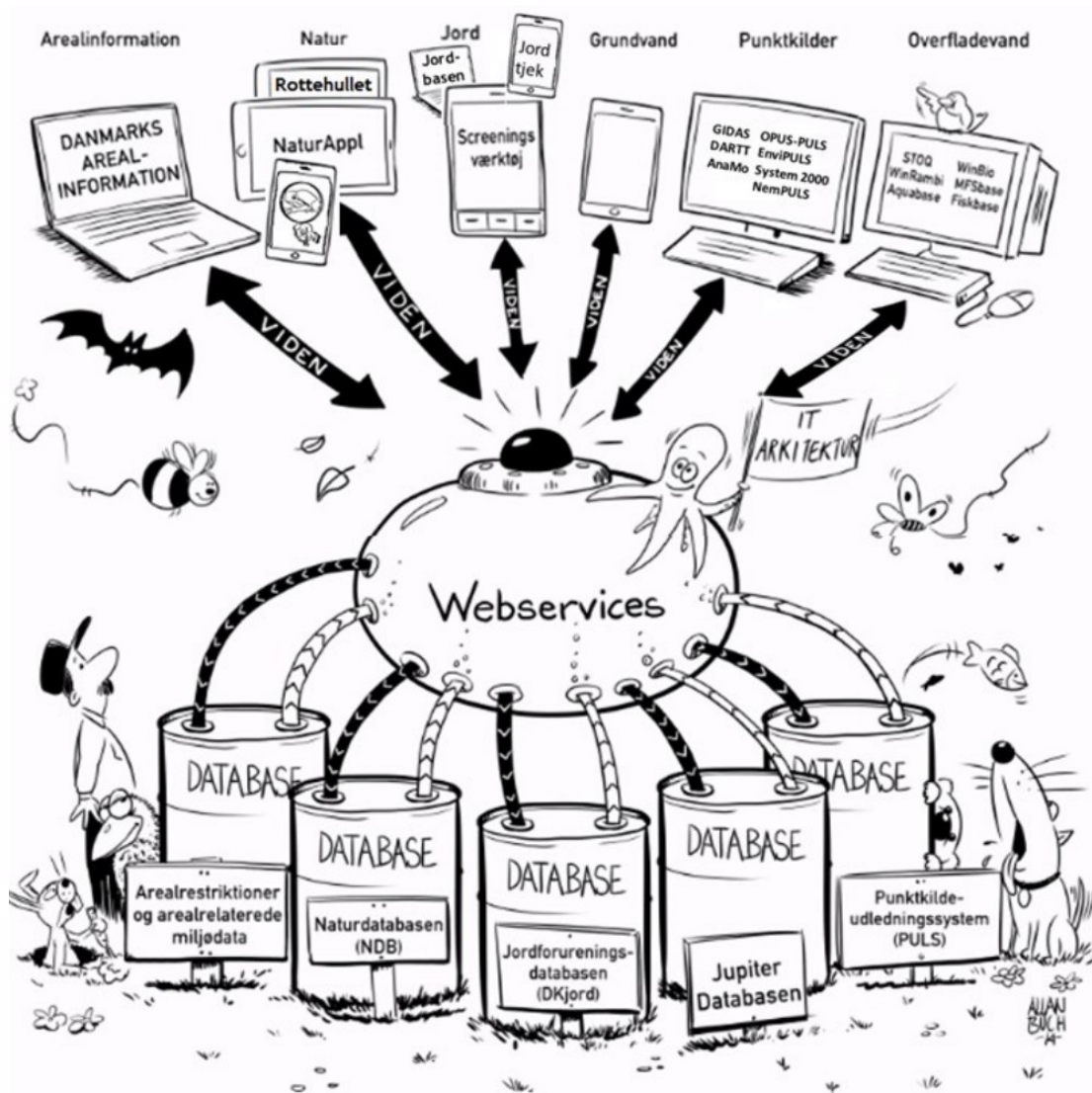


Måarkitektur for Danmarks Miljøportal

13. juni 2018



Indhold

Forord.....	3
Baggrund og behov	3
Nuværende målarkitektur fra februar 2016.....	4
Overvejelser i forhold til hovedprocesser og implementeringsproces	5
Hovedprocesser.....	5
Inddata.....	6
Forædling af data	6
Data ud	7
Opbevaring og databaser.....	7
Brugerstyring	8
Implementeringsproces	8
Implementeringsstrategi baseret på DevOps og stabil drift	8
Bilag 1: Fælles informations- og begrebsmodel på logisk niveau for Miljøområdet i Danmark	10
Bilag 2: Målarkitektur frem til februar 2016	12
Bilag 3: Referencearkitektur pr. 13. juni 2018.....	15

Revisionshistorik			
Dato	Rev. af	Version	Revision
2016-02-09	Jens Jakob Bork	1.0	Første målarkitektur godkendt
2018-05-24	Nils Høgsted	2.0	Tilføjelser til 1. udkast

Forord

Miljøportalens nuværende målarkitektur blev besluttet på et bestyrelsesmøde for DMP den 9. februar 2016. Overordnet set er der ikke fundet behov for at lave de store ændringer til målarkitekturen. Nedenstående opdatering er lavet for at gøre målarkitekturen kortere og mere læsevenlig, desuden er der foretaget nogle få justeringer af indholdet.

Målarkitekturen er baseret på følgende overordnede principper som Danmarks Miljøportal skal arbejde mod:

1. at opsplitte inddata og uddata i databaser.
2. at Danmarks Miljøportal er et samlet systemkompleks med så ens opbygning som muligt.
3. at udnytte standardkomponenter mest muligt. Særligt fokus på modne fællesoffentlige standardkomponenter og Azure og ESRI standardkomponenter.

Målarkitekturen er bygget op i følgende afsnit: Baggrund og behov, Gammel målarkitektur, Nuværende målarkitektur samt overvejelser i forhold til hovedprocesser og implementering

Baggrund og behov

Miljøportalens samlede it systemlandskab består i dag af en række forskellige løsninger, der gør brug af fælles og tværgående komponenter til brugerstyring, logning og lagring/udstilling af geodata. Dette giver en række fordele, men samtidig en afhængighed til de tværgående komponenter.

Disse forhold medfører dels krav til en stram styring af ændringshåndtering (change management) i eksisterende applikationer, dels store krav til koordinering af nye it-applikationer og nye udviklingsprojekter i forbindelse med test og idriftsættelse.

De senere år er der opstået en række ny behov hos myndigheder og eksterne (bl.a. rådgivere og 3. parts applikationsudviklere), der vedrører:

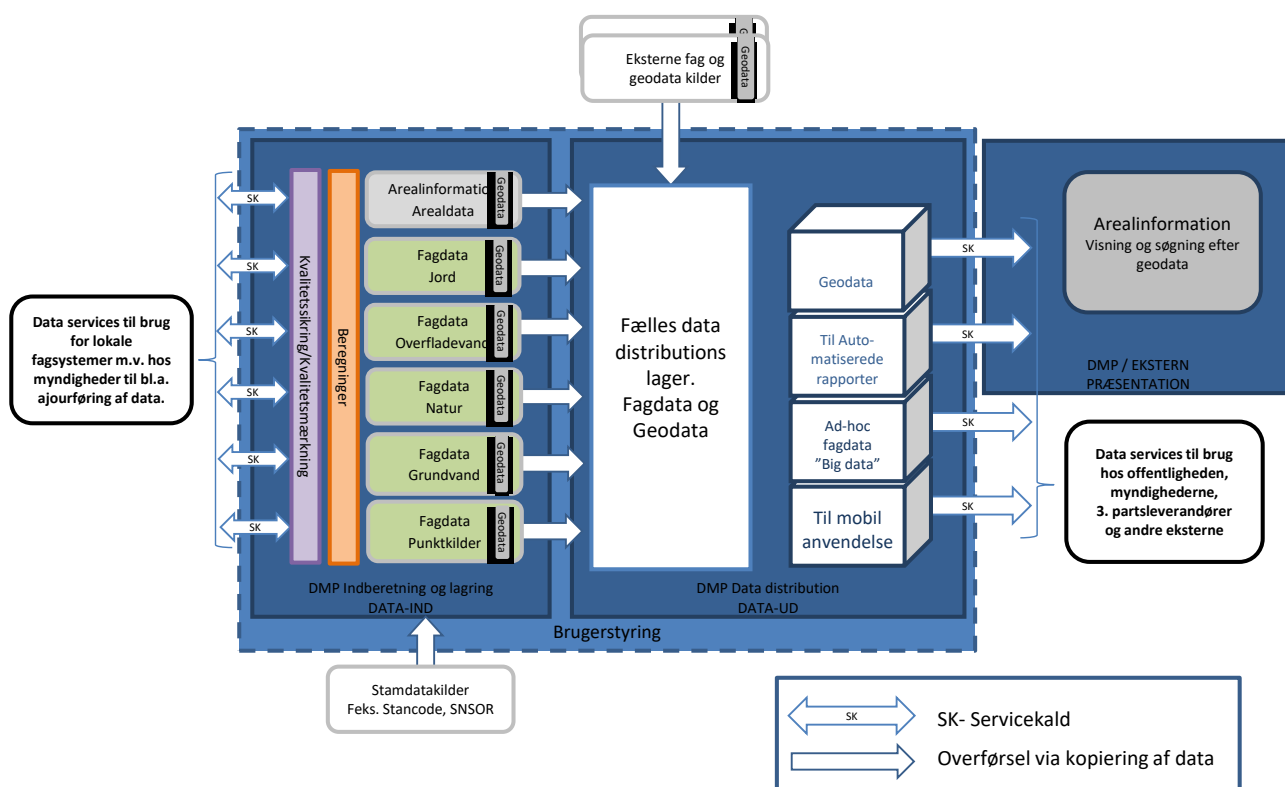
1. Optagelse af eksterne data fra bl.a. VVM redegørelser, universiteter og anlægsmyndigheder.
2. Opbygning af en ny central baseret overfladevandsdatasamling med afsæt i FODS 8.3. med mulighed for massive datamængder i form af online logger data.
3. Eksterne såvel som interne aktører efterspørger at kunne hente større datamængder (både geodata og fagdata) til brug i egne applikationer og analyser. Dette stiller nye krav til både tilgængelighed og systemernes kapacitet.
4. Efterspørgsel af muligheder for at forskningsmæssig adgang til og sammenstille data med andre datakilder.
5. Adgang til miljødata for brug i mobile-applikationer (bl.a. apps), hvilket stiller krav om fleksibel adgang og mulighed for at forespørge data.
6. Effektivisering af rapportering og dataanalyse hos myndighederne, fx hos Miljøstyrelsen og – efterhånden som der er kommet data til flere områder og opgaverne med rapportering er blevet flere.
7. Herunder muligheden for BIG DATA computer power.
8. Efterhånden som brugerkredsen bliver større og større og der kommer flere applikationer til er der opstået behov for flere og flere opdateringer af applikationer og platforme.
9. Krav om tilgang til og fokus på miljødata i offentligheden ift. miljøets tilstand er vokset de senere år. Hvilket giver behov for udstillingsmuligheder til offentligheden, med ”fortyggede” datasæt, med forklaring af data.

Nuværende målarkitektur fra februar 2016

På basis af ovennævnte situation og de nye behov er der opstillet følgende overordnede krav til den nuværende målarkitektur.

1. Den skal give en løsere kobling (færre afhængigheder) mellem applikationerne i forbindelse med adgang til data (data UD).
2. Den skal give nem og integreret adgang til alle DMP's data.
3. Den skal give adgang til at hente store datamængder til analyse m.m. og anvendelse på mobile enheder.
4. Den skal give adgang til at hente data efter kvalitetsmærke.
5. Den skal give mulighed for at såvel interne som eksterne skal kunne tilkoble deres applikationer/integrationer mod betaling direkte i DMP's Driftsmiljø.
6. Den skal give mulighed for at udføre Bigdata analyser og Machinelearning direkte på data.
7. Den skal give et simplere og mere robust driftsmiljø.
8. Den skal give mulighed for adgang til data med basale værktøjer som f.eks. indeholdt i en kontorplatform.
9. Den skal give standardiserede data på tværs af miljøer (data UD).
10. Den skal være baseret på Cloud by design¹ løsninger.
11. Adgangen til data skal kunne tilkøbes.

Figur 1: Målarkitektur



Den nuværende målarkitektur er illustreret på figuren ovenfor og har følgende karakteristika:

1. Decentraliseret placering af geodata simplificerer videreudvikling, men kræver hård standardisering af datamodeller.

¹ Cloud by design betyder at der anvendes komponenter som er udviklet af cloud leverandøren, som kun behøver configuration eller meget lidt tilpasning/kodning for at virke efter formålet.

2. Databaser og datasources på både ind og ud siden implementeret som mindre og mere fokuserede datasæt der er implementeret ift. den/de behov der skal understøttes.
3. Adskilt ind- og uddata – normaliserede data gør det simplere at sammenstille data.
4. "PUSH" IND-data model som i dag.
5. "PUSH" UD-data model betyder, at alle data til udstilling ikke nødvendigvis er 100 % tidstro ift. produktionsdatabaserne.
6. Dataudtræk belaster ikke produktions databaserne.
7. Som minimum standardisering af UD-data.
8. Løskoblet – simplificere "change" og videreudvikling.
9. Fælles kvalitetssikring/kvalitetsmærkningsmetoder eller komponenter. Herunder mulighed for fuldstændig adskillelse af data ud fra data's kvalitets niveau og fortrolighed.
10. Fælles beregningsmetoder-/komponenter. Herunder mulighed for at opbygge beregningsfunktionaliteter, via standardiserede cloud funktionaliteter.
11. Synkroniseringsmekanisme på de data hvor det er forretningsmæssigt nødvendigt at have realtime data i uddata systemerne.
12. Vil kunne understøtte aflevering til Statens Arkiver.
13. Vil kunne understøtte aflevering til den fællesoffentlige datafordeler og andre data initiativer.
14. Giver mulighed for at hente miljødata fra eksterne systemer og udstiller dem som kopidata til eksterne uden, at data bliver tilpasset til datamodeller i Miljøportalens databaser.
15. Giver mulighed for at udstille eksterne datakilder og gøre dem søgbare.
16. Mulighed for at tilgå data på forskellige måder. F.eks. med programmer fra kontorplatformen.
17. Mulighed for integrere interne såvel som eksterne løsninger direkte i driftsmiljøet, mod betaling.
18. Opsætning af mulighed for at sælge adgang til data.

For at få maksimal udbytte af en målarkitektur skal følgende faktorer iagttages:

1. Hvis data ikke fortsat standardiseres vil der over tid opstå uhensigtsmæssige bindinger mellem fagsystemerne og uddata. Der er derfor fortsat brug for fokus på datastandardisering og datagovernance.
2. Rå data, der gøres tilgængelig skal renses for personhenførbare informationer.
3. Rå data der gøres generelt tilgængeligt bør renses for sagsdata (procesdata, sagsbehandlere, mv) informationer.
4. Der bør skelnes mellem rå-data hvis indhold og struktur principielt kan ændres fra dag til dag og datasæt (fx statistikker, ledelsesinformation og rapporter) hvor Danmarks Miljøportal har aftalt eller er forpligtiget til at levere. For sidstnævnte skal der ved ændringer i fagdatabaserne iagttages de samme forholdsregler og varsler som ved ændringer af systemer og servicegrænseflader.

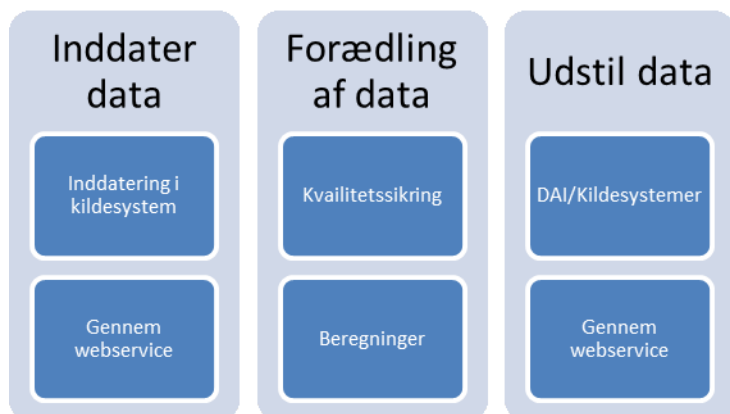
Der arbejdes i et særskilt spor på at udvikle en samlet logisk datamodel for alle Miljøportalens data.

Overvejelser i forhold til hovedprocesser og implementeringsproces

Hovedprocesser

På et meget aggregeret niveau understøtter Danmarks Miljøportal tre overordnede processer: 1) inddatering af data, 2) forædling af data og 3) udstilling af data. I figur 1 er disse tre hovedprocesser skitseret.

Figur 2: DMPs hovedprocesser



I det følgende skitseres først de overordnede behov pr. hovedprocess derefter følger en række beskrivelser af mulige komponenter der kan understøtte disse behov. Det skal bemærkes at de skitserede komponenter er resultatet af en foreløbig brainstorm eksempelvis er komponenter fra andre lande ikke medtaget eller analyseret endnu. Til sidst beskrives et tværgående komponent til brugerstyring.

Inddata.

På inddata området har Danmarks Miljøportals brugere følgende overordnede behov.

1. Det skal være muligt let at oprette stamdata i form af målestationer eller prøve aktiviteter
2. Det skal være muligt, at inddatere et meget bredt spektrum af forskellige måleresultater med så meget validering som muligt.
3. Det skal være muligt at sende prøver med tilknyttede metadata om prøvninger til laboratorier
4. Det skal være muligt at inddatere prøveresultater fra laboratorier et sted, via webservices.
5. Det skal være muligt at lave system integrationer for aflevering af ovenstående data.
6. Data skal være mærket ift. tilstand og brug.
7. Det skal være muligt for borgere og professionelle at indlevere kvalitetetsdeklarerede data

Ovenstående behov bliver i dag understøttet af parterne selv, gennem indkøb procesunderstøttende 3. parts løsninger, som p.t. typisk er meget snævert fagligt specialiserede og gennem løsninger udviklet af Danmarks Miljøportal.

For inddata gælder generelt:

1. Alle services skal kunne håndtere masse upload fra en enkelt partner.
2. Der kan blive tale om system til system integrationer indenfor de områder hvor der er tale om store datasæt som levere data fra mange af DMP's partnere på en gang, herunder f.eks. datafordeler mv.
3. Alle services skal tilgås via Azure Service API. Hvilket udover en nemmere tilgang for udviklingsleverandører også sikre en ensartet opbygning og standardisering.
4. Alle services skal have indbygget Azure Queue og eller Azure Service Bus Queue således data kan overføres til dataud delen så realtime data også kan opnås i forbindelse med alle eller udvalgte dataud data.

Forædling af data

Efter data er inddateret har Danmarks Miljøportals brugere behov for at forædle data både gennem kvalitetssikring af enkelte datasæt og udregning af forskellige indexes. De overordnede behov kan opstilles som følger:

1. Mulighed for at lave automatisk kvalitetssikring eksempelvis ved hjælp af machine learning, så fejl opdages så hurtigt som muligt
2. Mulighed for at rette enkelte datasæt
3. Mulighed for at lave analyser på tværs af store datasæt og på basis af disse analyser identificere behov for rettelser af data.
4. Mulighed for at rette store mængder af data på en gang.

5. Mulighed for at sætte en anden kvalitetsmarkør på data.
6. Mulighed for at beregne diverse indekser på basis af sammenstilling af forskellige data.

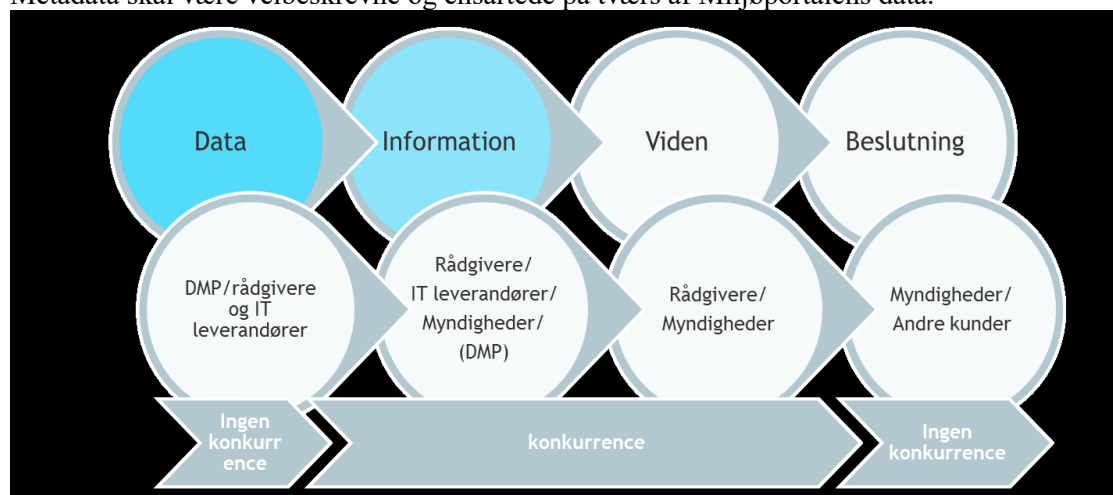
Der er p.t. stor forskel på, hvordan kvalitetssikringen af data laves på tværs af de forskellige fagområder. Noget af denne forskel er faglig begrundet, mens andre dele af forskellen er betinget af historik og mulighed for systemunderstøttelse. Miljøportalen vil arbejde for, at den model der er udarbejdet for VanDa overføres til de andre områder hvis muligt.

Data ud

De overordnede behov for udstilling af data og beregninger kan sammenstilles som følger:

1. Det skal være nemt og få adgang til alle DMP's data.
2. Det skal være let, sikker og hurtigt at bruge Miljøportalens webservices.
3. Det skal være let at udvikle på basis af Miljøportalens data.
4. Det skal være muligt at lave dataudtræk på tværs af fagområder.
5. Det skal være muligt for at såvel interne som eksterne skal kunne tilkoble deres applikationer/integrationer direkte i DMP's driftsmiljø – for eksterne mod betaling.
6. Det skal være muligt at udføre Bigdata analyser og Machinelearning direkte på data evt. mod betaling for driftsomkostninger.
7. Der skal være adgang til alle relevante data i forbindelse med myndighedsforvaltningen.

Metadata skal være velbeskrevne og ensartede på tværs af Miljøportalens data.



Ovenstående figur 1 viser kæden: data, information, viden og beslutning og skitserer, hvem der er de primære spillere indenfor de forskellige områder og om hvorvidt der er en konkurrencesituation. Som det fremgår er den primære konkurrence når data skal omformes til information og videre til beslutning. Det er her 3. parts leverandørerne primært er interessante og det er her rådgiverne omsætter data til viden, som kan bruges af myndigheder i en beslutningssituation. Hvor data der ligger i Miljøportalen er rimeligt statiske og både rådgivere og 3. parts leverandører er villige til at levere data, da der ikke hersker en konkurrencesituation er det anderledes i forhold til omdannelsen af data til information og viden. Her er den primære konkurrence og der er så mange forskellige behov, at det vil være umuligt for Danmarks Miljøportal at dække dem. I stedet bør Danmarks Miljøportal indgå i et netværk med parternes IT afdelinger, 3. parts leverandører og rådgivende ingeniører for at tilfredsstille behovene bedst muligt.

Opbevaring og databaser

Opbevaringsplatformen vil forsat umiddelbart bestå af relationelle databaser, men der vil også kunne anvendes andre former for data storage, herunder NOSQL databaser og storagemetoder som Azure tablestorage m.v. Valget afhænger af datatypen og størrelsen på data.

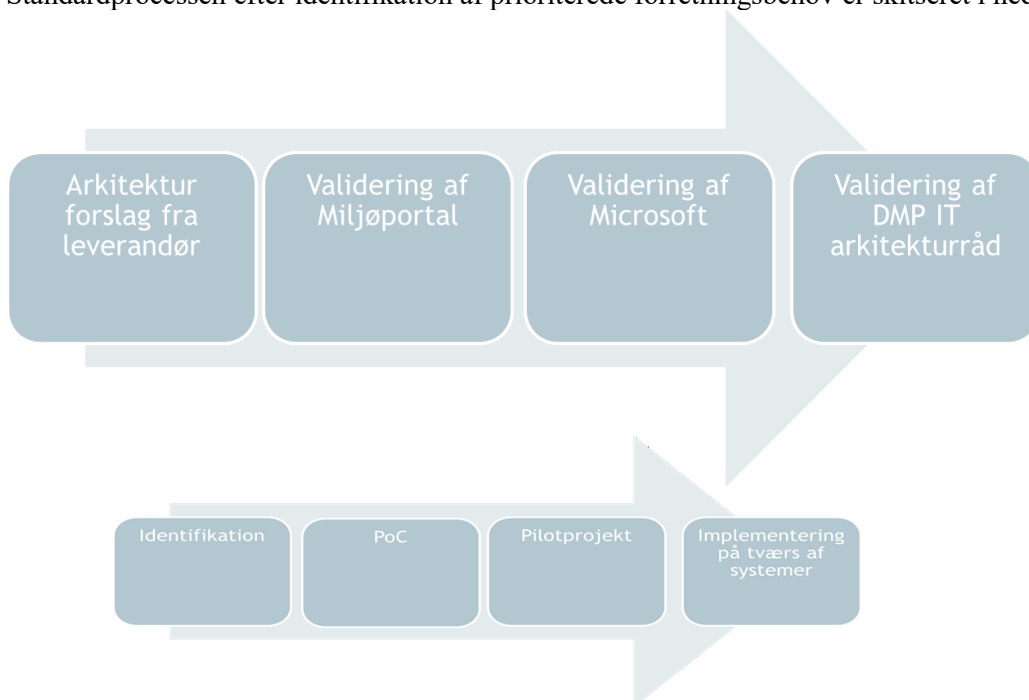
Der skal arbejdes henimod at alle databaser laves mere fokuserede og dermed også mindre i størrelse og udstrækning. Således skal de opbygges til formålet, fremfor at kunne indeholde store datasæt. Skal der f.eks. laves masse inddateringsmetoder, må dette ikke kunne påvirke andre former for inddatering på data fra samme område eller datasæt.

Brugerstyring

Danmarks Miljøportals brugerstyringsløsning understøtter Danmarks Miljøportals fokus på sikring af dataintegritet samt effektivitet ift. det sammenhængende login. Løsningen har primært beskyttet Danmarks Miljøportals egne løsninger samt løsninger fra myndigheder som har relation til parterne. Brugerstyringen skal udbredes til at alle brugere, private applikationsleverandører kan tilkoble sig imod på betaling.

Implementeringsproces

DMP er allerede nu ved at undersøge fordele og ulemper ved en gradvis omlægning af Danmarks Miljøportals systemer til den nye målarkitektur. Denne omlægning vil foregå gradvist i takt med, at der opnås finansiering til at understøttes nye forretningsbehov eller modernisere gamle systemer. Standardprocessen efter identifikation af prioriterede forretningsbehov er skitseret i nedenstående figur::



Implementeringsstrategi baseret på DevOps og stabil drift

Danmarks Miljøportals implementeringsstrategi er baseret på DevOps kombineret med et ønske om at sikre stabil drift gennem forbedret test og dokumentation samt proaktiv overvågning af såvel performance som brugen af systemerne.

I forhold til DevOps så har Danmarks Miljøportal valgt Visual Studio Team Services som DevOps tool chain. Danmarks Miljøportal vil i forhold til implementering af DevOps arbejde mod at:

- Deploys foregår gennem automatisk deploy ved at trykke på et link
- Miljøer oprettes ved hjælp af scripts

- Systemerne dokumenteres ved hjælp af en overordnet komponentbeskrivelse samt nedbrydelse af komponenterne i Visual Studio Team Services
- Al kildekode ligger i Visual Studio Team Services og eksekveres herfra
- Test cases er beskrevet i Visual Studio Team Services og de test cases der giver mening er automatiseret
- Der opbygges standard performance tests i Visual Studio Team Services
- En stigende grad af automatisering af integrationstest
- Alle ændringsønsker, fejl og issues dokumenteres i Visual Studio Team Services.
- Etablering af automatiseret unit test og test af kodekvalitet i Visual Studio Team Services.

I forhold til at sikre stabil drift, så vil den primære driver være det øgede fokus på test. Derudover vil sekretariatet arbejde for en bedre og mere proaktiv overvågning af systemer ved at:

- Opsætte performancemålinger for specifikke kritiske forretningsprocesser
- Opsætte bedre generel overvågning af performance og integrationer på tværs af Miljøportalens systemkompleks.

Bilag 1: Fælles informations- og begrebsmodel på logisk niveau for Miljøområdet i Danmark

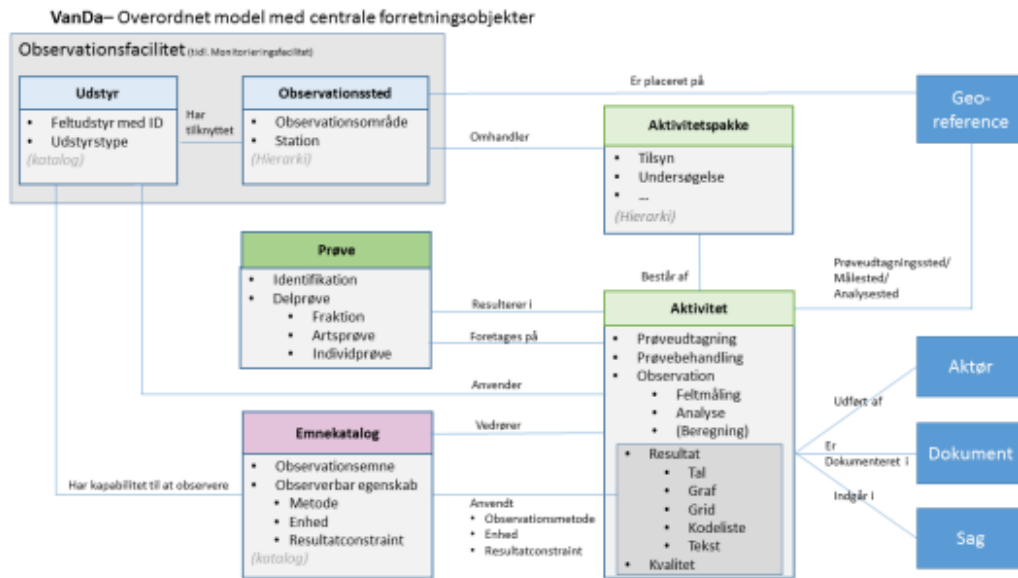
Informationsarkitekturen udgør det forretningsmæssige grundlag for udviklingen af logiske og fysiske datamodeller for it-systemer og i sidste ende for opbygningen af systemernes database og snitflader. Informationsarkitekturen har derfor stor indflydelse på, hvor let det er for forskellige systemer at snakke sammen og hvor let det er at kombinere data fra forskellige systemer.

I forbindelse med VanDa projektet blev der udarbejdet en begrebs- og informationsmodel for observationssteder og målbare miljødata (<http://info.rammearkitektur.dk/byggeblokke/Vanda/vanda.htm>). Modellen kan ses som en fortsættelse af arbejdet med de første fælles grundbegreber defineret i sektorstandardiseringsudvalget for miljødata fra 2006, men med en række elementer fra nyere initiativer indarbejdet. Særligt arbejdet på europæisk plan med udvikling af dataspecifikationer i regi af INSPIRE-direktivet, men også med arbejde i regi af øvrige nationale og sektorspecifikke initiativer og projekter.

Denne datamodel er siden blevet udvidet med PULS data, hvor begrebs- og informationsmodellen viste sin robusthed i forhold til at kunne rumme andre fagområder. Det er planlagt i 2018, at udvide denne begrebs- og informationsmodel til også at omfatte Jupiter data og opdatere den så den overholder den nye hvidbog for datamodellering vedtaget i regi af den fællesoffentlige digitaliseringsstrategi. På sigt bør begrebs- og informationsmodellen udvides til at gælde alle Miljødata.

En fælles begrebs- og informationsmodel på logisk niveau betyder ikke at alle fysiske datamodeller i systemerne skal være ens. Begrebs- og informationsmodellen bør tjene som et udgangspunkt for den fysiske modellering og de fysiske modeller bør kunne mappes op mod den logiske model. Så selvom formålet er at standardisere, så skal der ske en afvejning mellem standardiseringsfordele og forretningsmæssige hensyn. Et konkret eksempel kunne være, at der er enighed om at der både kan være en 1 – 1 relation og en 1 – mange relation mellem en prøve og et analyseresultat, Det vil sige en prøve kan give et eller flere analyseresultater, men at det omvendte ikke er tilfældet dvs. et analyseresultat kan ikke stamme for mange prøver.

Arbejdet med en fælles begrebs- og informationsmodel forankres i en datamodelleringsgruppe bestående af repræsentanter fra parterne. Modellen sendes i høring i takt med at den bliver udviklet. Det skal bemærkes, at udbredelsen af datamodellen er centralt for at høste gevinsterne af arbejdet. Der er derfor behov for ledelsesmæssig opmærksomhed for at sikre, at datamodellen ikke kun bliver brugt i regi af Danmarks Miljøportal, men også i regi af de enkelte parters udviklingsprojekter og i sektoren generelt.

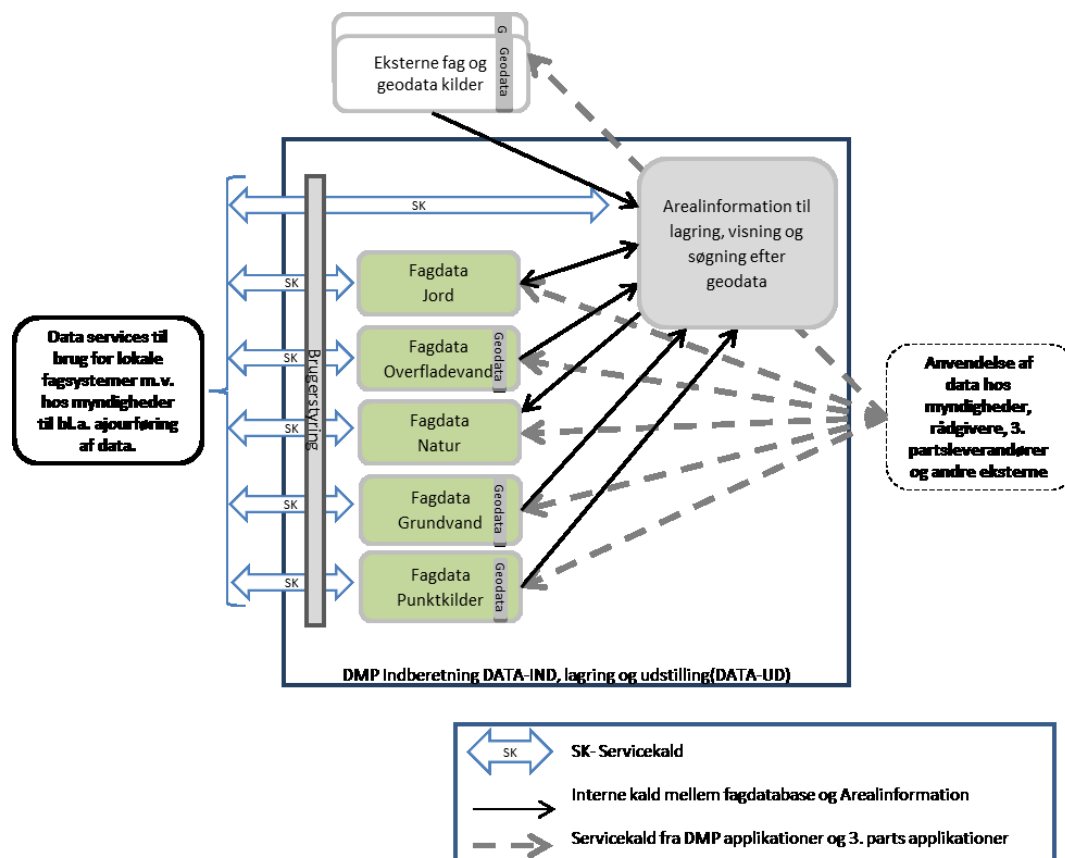


FIGUR 8: OVERORDNET MODEL MED CENTRALE FORRETNINGSOBJEKTER

Bilag 2: Målarkitektur frem til februar 2016

Den gamle målarkitektur frem til februar 2016 var baseret på en centralisering af geodata (i arealinformation), dette er kun delvist gennemført. Dels var der en række gamle applikationer, hvor det ikke har været økonomisk eller praktisk fornuftigt at flytte geodata til Arealinformation, dels var der behov for at hente eksterne geodata fra tredjepart, bl.a. PlansystemDK og NaturErhverv.

Figur 3: Eksisterende arkitektur



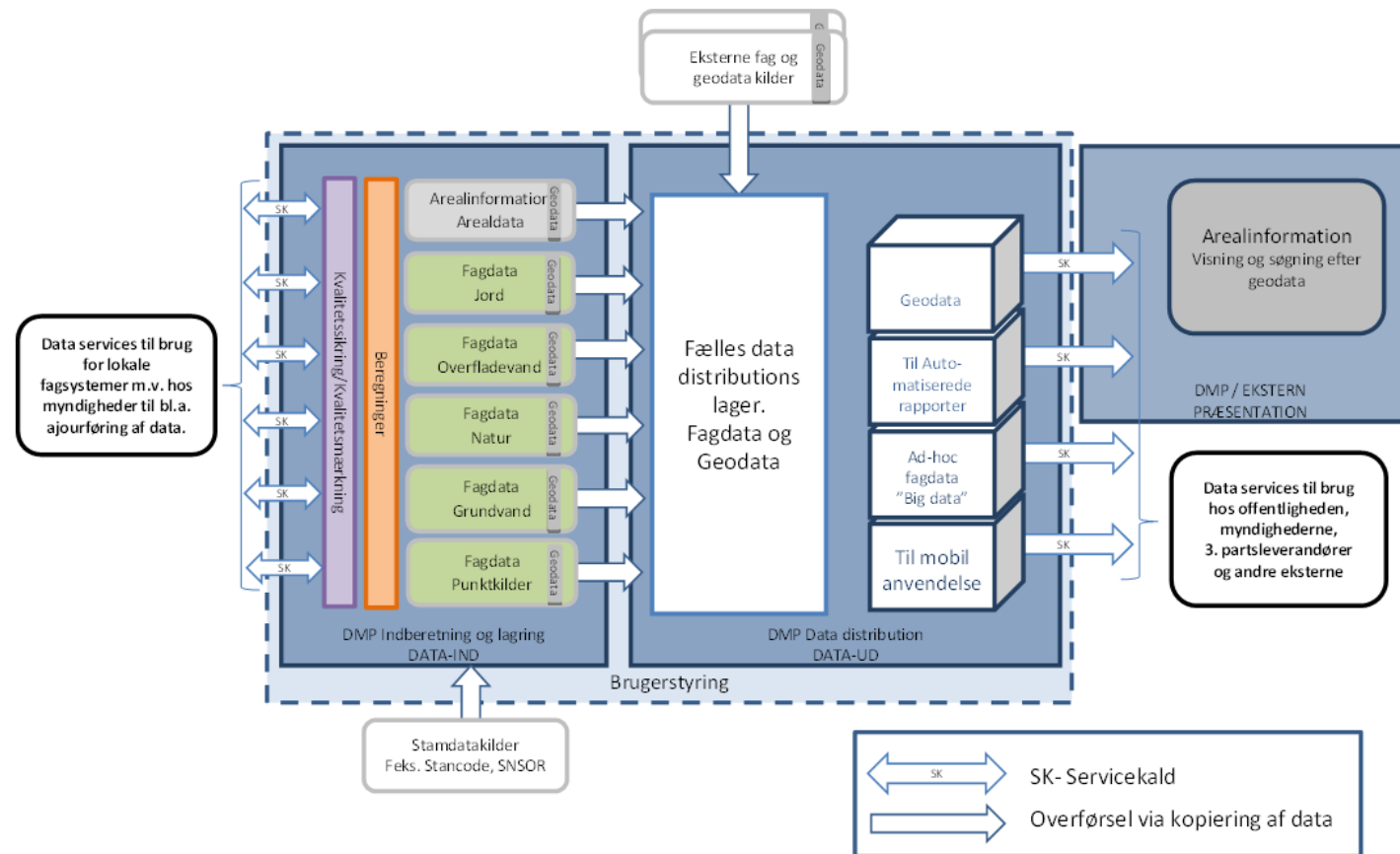
Den nuværende målarkitektur er illustreret på figuren ovenfor og har følgende karakteristika:

1. Delvist centraliseret geodata, hvilket medfører kompleks videreudvikling og kræver hård standardisering af datamodeller.
2. Direkte adgang til fagsystemer med forskellige datamodeller, der komplicerer sammenstilling af fagdata.
3. Geodata er standardiseret gennem lagring i Arealinformation
4. "PUSH" IND-data model som i dag.
5. "PULL" UD-data model – Data altid ajourført, der samtidig medfører, at dataudtræk belaster fagdatabase med kildedata. Desuden er udtræks mulighederne begrænset til de funktionaliteter som leveres i basisløsninger eller de klienter der er bygget til sagsbehandlerne.
6. Der findes ingen decideret udstilling af alle fagdata eller endsige download af stor datasæt til videre forarbejdning.
7. Der findes ingen mulighed for at interne såvel som eksterne kan lave store dataudtræk på tværs af DMP's løsninger.

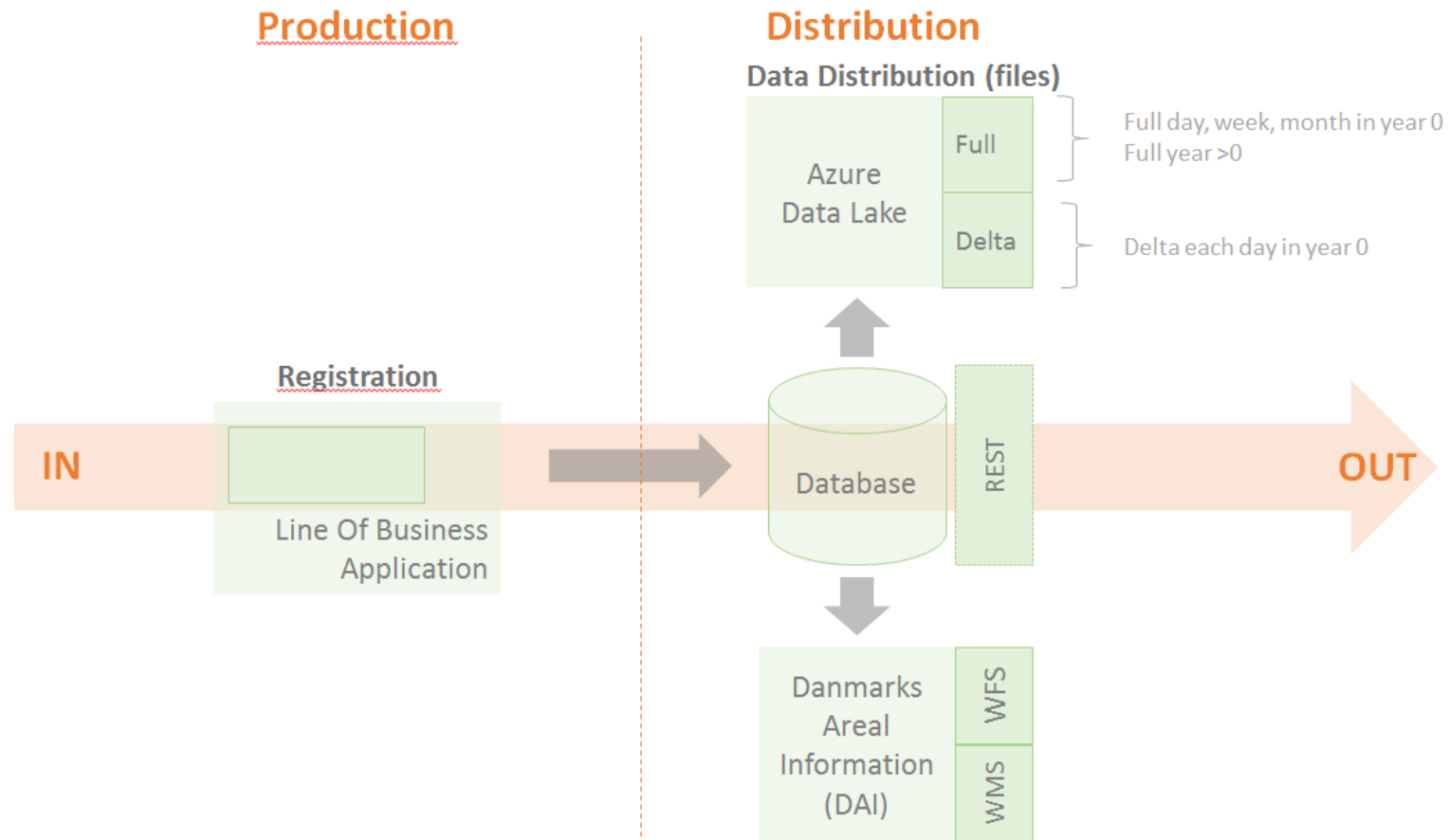
8. Der findes ingen mulighed for at interne såvel som eksterne kan integrere deres løsninger i DMP's driftsmiljø.
9. Der findes ingen mulighed for at tilkoble applikationer fra kontor platformen direkte til data.
10. Tæt afhængighed mellem fagdatabase og arealinformation, hvilket komplicerer "change" og videreudvikling.
11. Manglende standardisering af begrebsapparatet på tværs af fagområder.
12. Manglende data-governance på alle områder
13. Der er historik, og sporbarhed i data i forbindelse med løsnings af myndighedsbehandling, fx afgørelser. Sker typisk i Myndighedernes fagsystemer, men kan principielt også ske på grundlag af uddata fra Miljøportalens databaser.

Bilag 3: Referencearkitektur pr. 13. juni 2018

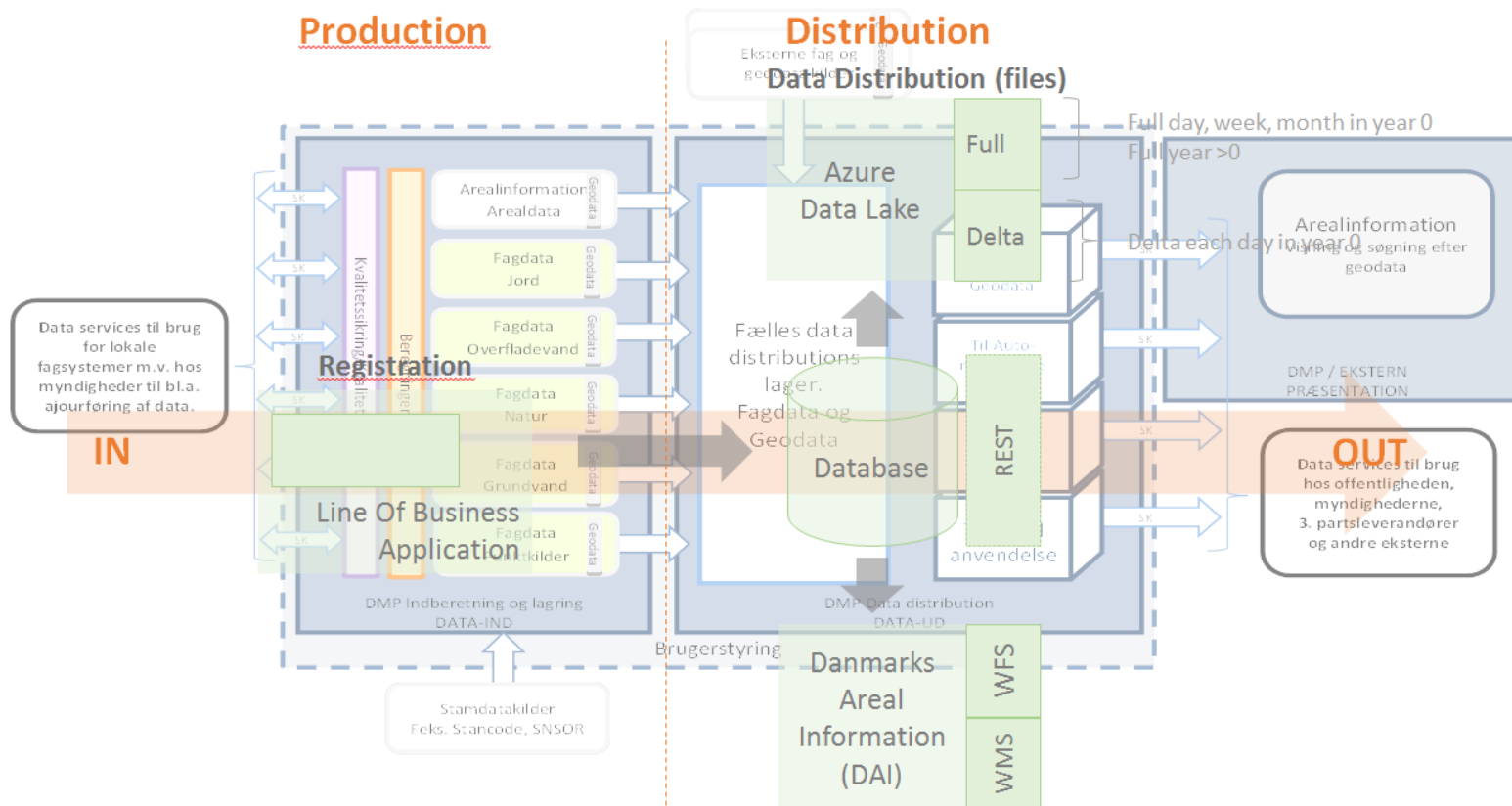
Måarkitekturen



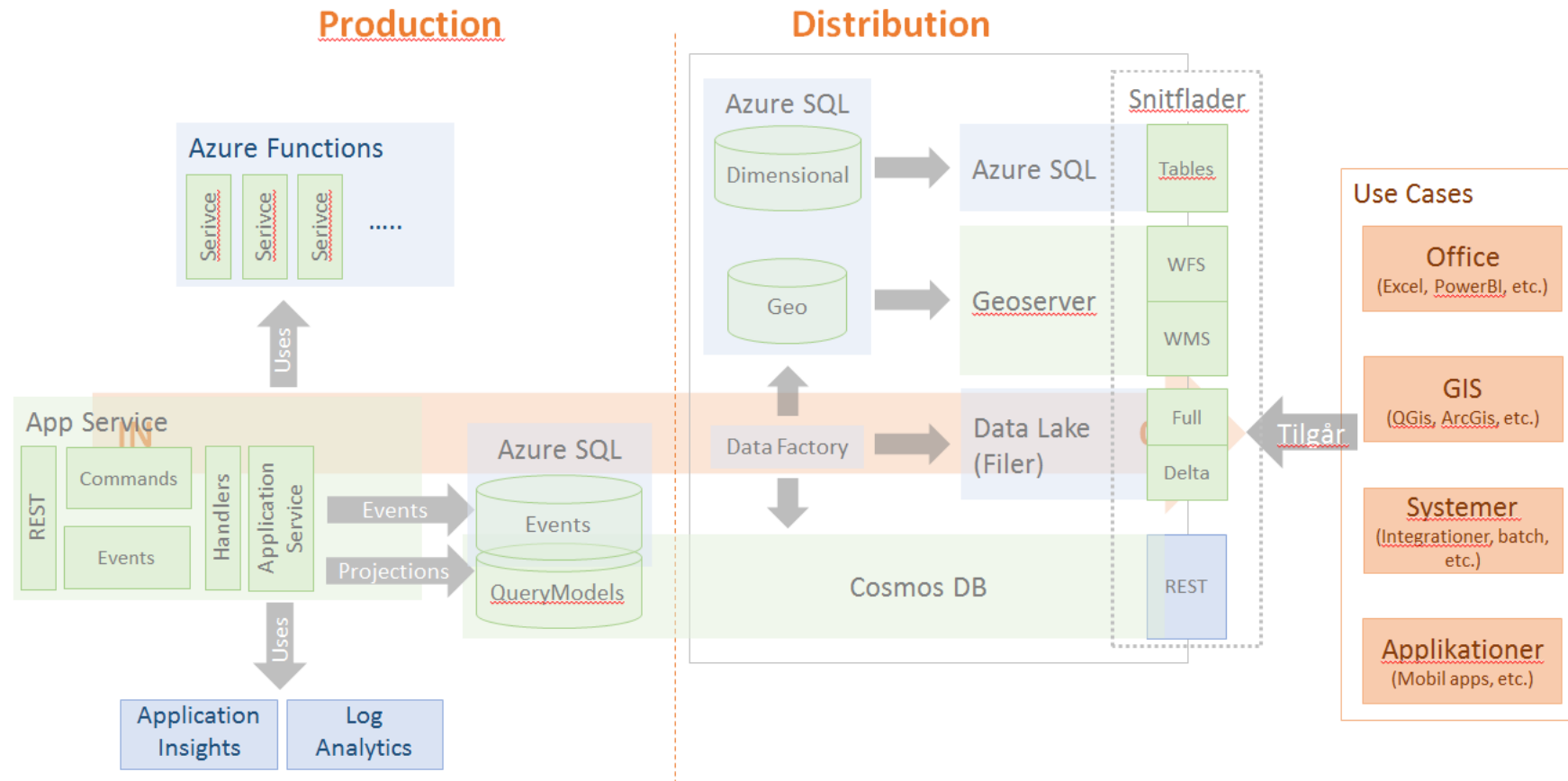
Referencearkitekturen (meget overordnet)



Sammenhængen mellem mål- og referencearkitektur



Referencearkitekturen



Referencearkitekturen, designprincipper

- Baseret på event sourcing [som er baseret på CQRS]
- Produktionssiden benytter aggregates, commands og events
- Distributionssiden baseret på projektions oven på event streams
- Events persisteres i Azure SQL
- Projektions udstilles vha. Azure Cosmos
- Revisions- og transaktionslog vha. event sourcing og vedligeholdes som projection.
- Snitflader implementeret som app services
 - REST API defineret i OpenAPI
 - Sikkerhed med OAuth og OpenID Connect vha. DMP Brugerstyring
- Stateless logic der genbruges på tværs implementers med Azure functions
- Application insights til telemetry og hændelser
- Log Analytics til Infrastruktur logning

Referencearkitekturen – Status på implementering

Anvender eller planlagt

- DKJord Screening
- GrundRISK (under udvikling)
- VanDa (under udvikling)
- PULS2 (under udvikling)

Anvender ikke og ikke planlagt

- DKJord service
- RotteWeb
- DinNatur
- PULS

Status på brug af komponenter

