



cutting through complexity™

Rapport – Forprosjekt for Big Data og byggkvalitet

Direktoratet for byggkvalitet
18. november 2015



Forprosjekt for Big Data og kartlegging av byggkvalitet og byggskadeomfang

KPMG har gleden av å oversende rapport på oppdraget "Forprosjekt for Big Data og kartlegging av byggkvalitet og byggskadeomfang". Som beskrevet i engasjementsavtalen datert 28. april 2015, har KPMG bistått med følgende aktiviteter:

- Kartlegging av mulig relevant informasjon i sentrale aktørers registre/datakilder
- Avklare om registre/datakilder er åpne/tilgjengelige, og på hvilke vilkår
- Identifisere indikatorer som kan benyttes til å trekke ut relevant informasjon fra ulike datakilder
- Lage en enkel testløsning basert på et begrenset utvalg datasett

I tillegg til nevnte aktiviteter, har vi identifisert aktuelle anvendelsesområder for en "Big Data"-løsning, samt hindringer og anbefalte tiltak for å lykkes med en slik løsning. Til slutt har vi skissert forslag til gjennomføring av et eventuelt hovedprosjekt som bygger på dette forprosjektet.

Gjennom vårt arbeid har vi gjort oss noen overordnede observasjoner og vurderinger:

- Vi ser en digitaliseringstrend, der registreringer av skader, feil og mangler i stadig større grad gjennomføres i digitale verktøy og lagres i strukturerte databaser. Dette medfører stadig større mengder og stadig mer tilgjengelig datagrunnlag for analyseformål.
- Vi ser også en høy grad av modernisering og utskifting av systemer, både FDV-systemer, kunde- og klagehåndteringssystemer, skadearkiver mv. Dette medfører at data i fremtiden vil bli stadig mer tilgjengelig, samtidig som det medfører at en løsning med systematisk uthenting av data fra slike systemer må være fleksibel for å kunne tilpasses stadig endrede systemer.
- Vi ser svært lite bruk av statistikk og analyse for å søke innsikt i omfang, trender, årsakssammenhenger mv. Her har med andre ord bransjen et stort uforløst potensiale, som Direktoratet for Byggkvalitet kan være med på å "sparke i gang".
- En siste trend som er verd å merke seg, som ikke er særskilt for denne bransjen, er økt bruk av sensorer og økt mengde data produsert av sensorer. Dette representerer på sikt en meget interessant datakilde for vurdering av visse aspekter ved byggkvalitet, men er ikke vurdert i detalj i denne rapporten.

Gitt trendene og våre observasjoner, synes det som om tidspunktet er riktig for etablering av en løsning for systematisk innhenting, bearbeiding og analyse av data om byggkvalitet, skader, feil og mangler. En slik løsning kan bli til nytte ikke bare for DiBK, men også for dataeiere som utbyggere og forvaltere, forskere og andre aktører i bransjen.

Vi takker for det gode samarbeidet så langt og håper rapporten står til DiBK sine forventninger.

Vennlig hilsen

KPMG AS



Jan-Erik Martinsen

Partner

Innhold

1	Sammendrag	3
2	Innledning	6
3	Teknologiske muligheter: verktøy og metodikk knyttet til Big Data	11
4	Funn	13
5	Vurderinger	31
6	Forslag til hovedprosjekt	40

Vedlegg

	Vedlegg A Gjennomførte workshops og intervjuer	43
	Vedlegg B Detaljerte vurderinger av datakilder	45
	Vedlegg C Detaljer vedrørende test/"Proof of concept"	49
	Vedlegg D Hindre og tiltak for å overkomme identifiserte hindre	54
	Vedlegg E Løsningsskisse for målbildet	59
	Vedlegg F Prosjektplan for foreslått hovedprosjekt	62

1 Sammendrag

Bakgrunn og metode

Per i dag mangler Direktoratet for Byggkvalitet (DiBK) metoder og verktøy for kontinuerlig måling og overvåking av utviklingen i byggkvalitet og for å analysere effekter av regelverksendringer. Et forprosjekt er derfor gjennomført for å vurdere mulighetene for bruk av "Big Data"-metodikk og teknologi for disse formålene.

Forprosjektet har kartlagt og vurdert ulike faktorer og områder som påvirker mulighetsrom og begrensninger. Kartleggingen er gjennomført gjennom idéworkshops, én-til-én-møter med eiere av aktuelle datakilder og øvrige interessenter, samt arbeidsmøter med representanter fra DiBK. I tillegg er det utviklet en testløsning med et begrenset utvalg av datakilder.

Funn og vurderinger

Testløsningen har vist at tradisjonelle metoder som statistiske analyser og visualisering av data i eksisterende systemer hos ulike dataeiere kan gi innsikt i både *regelverkseffekter* og *omfang av skader, feil og mangler*. I tillegg viser testene at såkalt "text mining" kan benyttes for å automatisk kategorisere skader i felles skadekategorier som kan aggregeres og analyseres på tvers av ulike datakilder.

Datakildene som synes å gi størst verdi er registre over *forsikrings saker*, forvalteres *FDV-systemer* og utbygges *saks- og avvikssystemer* med informasjon om avvik, skader feil og mangler.

De mest aktuelle bruksområder som er identifisert kan i hovedsak deles i fire områder:

1. Innsikt i *effekter av regelverk*, gjennom analyser av data fra Finans Norges registre for henholdsvis vannskader og brannskader.

Samme datakilder kan dessuten benyttes til å analysere utvikling og omfang av vannskader og brannskader.

2. Dashboard med *indikatorer på byggkvalitet og byggskader, feil og mangler*, basert på registrerte avvik og kundeforhold i entreprenørens og forvalteres systemer, kan brukes til å overvåke utviklingen i byggkvalitet og byggskader. Samme datakilder kan benyttes til å analysere sammenheng mellom skadeomfang og faktorer som geografi, byggtypen, byggets alder mm.

En teknisk løsning som henter data "som de er" fra ulike datakilder og som automatisk legger på kategorier og utleder indikatorer før resultatene vises for brukerne, har den fordelen at ulike indikatorer og definisjoner kan benyttes for ulike formål og for ulike brukergrupper. Det er dermed ikke

nødvendig med felles definisjoner av begreper som "byggkvalitet" og "byggeskade" for at løsningen skal kunne gi verdi.

3. *Benchmarking* av omfang av skader, feil og mangler mellom ulike entreprenører og forvaltere.
4. Innsikt i *tolkninger og grad av forståelse av regelverk* basert på telling av antall innlegg i relevante diskusjonsfora som omhandler visse typer skader, feil og mangler.

Samme datakilde kan for øvrig benyttes til å analysere hvilke skader, feil og mangler som omtales, og således brukes til en indikator på oppfattet byggkvalitet.

For å få nytte av verktøy for analyser og indikatorer knyttet til byggkvalitet og regelverk, er det visse utfordringer som må adresseres:

- En digitaliseringstrend og høy grad av modernisering og utskifting av systemer hos aktørene i bransjen har den positive effekten at data i fremtiden vil bli stadig mer tilgjengelig. Dette stiller samtidig krav til en eventuell løsning med systematisk uthenting av data fra slike systemer, da løsningen må være fleksibel og må kunne håndtere stadig endrede kildesystemer.
- Standarder og felles begrepsapparat benyttes i liten grad i datakildene som er kartlagt, og datakvaliteten er varierende. For å kunne fremskaffe indikatorer på aggregert nivå og gjøre analyser på data fra ulike kilder, må det være et sett med felles kategorier og øvrige informasjonselementer. En fremtidig løsning bør derfor støtte automatisert kategorisering i type skade/feil og mangel. I tillegg bør løsningen gjøre oppslag mot matrikkelen for å få enhetlig angivelse av informasjon som type bygg, geografisk beliggenhet og byggeår.
- For å kunne sammenstille og benytte data som inneholder personopplysninger vil DiBK måtte søke konsesjon fra Datatilsynet for behandling av disse. Det må samtidig iverksettes nødvendige tiltak knyttet til informasjonssikkerhet, både hva gjelder kompetanse og tekniske løsninger.
- Eierne av aktuelle datakilder har ulik grad av motivasjon for å bidra med data til en fremtidig felles løsning, og det vil kreve noe ressurser fra disse aktørene for å få sendt data til sammenstilling. Vi ser at aktørene i liten grad utnytter potensialet i egne datakilder i dag, og en felles løsning kan således gi verdi også for dataeierne om de får tilgang til verktøyet.

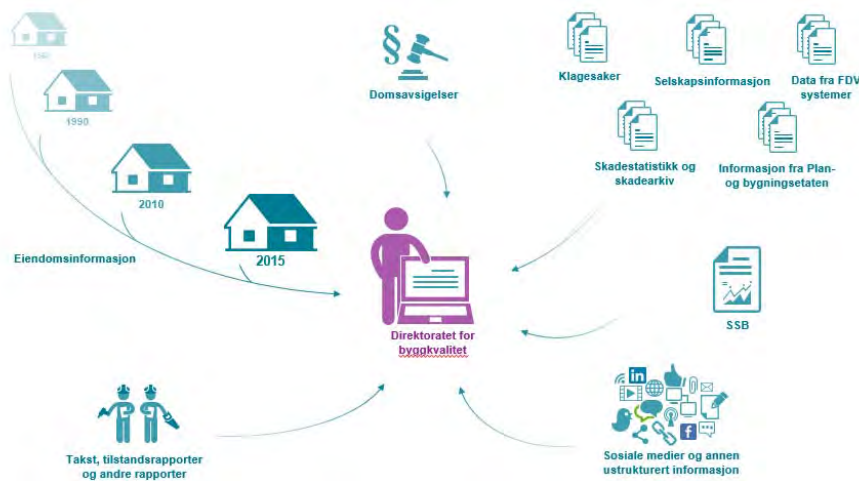
Anbefalinger

- KPMG anbefaler at DiBK *etablerer et hovedprosjekt* for å få opp et system for analyse og kartlegging av byggkvalitet og byggfeil basert på data fra ulike eksterne datakilder. Som del av prosjektet, og før et system kan anskaffes og implementeres, anbefaler vi at DiBK systematisk arbeider med tiltak for å overkomme hindringer som er identifisert i forprosjektet. Vi anbefaler å gjennomføre disse tiltakene gjennom separate arbeidsstrømmer i et hovedprosjekt. Første fase av et eventuelt hovedprosjekt bør være en "scopingfase", hvor mål, mandat og plan for hovedprosjektet utarbeides og besluttes.
- DiBK bør *anskaffe et analyse- og visualiseringsverktøy* for "ad hoc"-analyser av strukturerte data (inklusive tekstinnhold), enten før eller samtidig med hovedprosjektet. Et slikt verktøy er enkelt å ta i bruk og kan benyttes til å gjøre analyser på data fra VASK, BRASK og andre "ad hoc" datakilder. Et slikt verktøy kan også benyttes til å vurdere omfang før regelverksendring trer i kraft, for eksempel ved å bruke uttrekk fra matrikkelen til analyser av antall bygg med relevante egenskaper. Anvendelse av et slikt verktøy vil kunne gi DiBK nyttig innsikt, og vil dessuten kunne gi DiBK nyttig erfaring som kan tas med inn i hovedprosjektet.

2 Innledning

2.1 Bakgrunn og formål

Per i dag finnes det ikke metoder eller verktøy for kontinuerlig måling og overvåking av utviklingen i byggkvalitet eller for å analysere effekter av regelverksendringer. Indikatorer på byggkvalitet og byggefeil blir registrert hos ulike aktører, på strukturert og ustrukturert form. Samtidig ser vi en rivende utvikling innen teknologi og verktøy for søk og analyse av både strukturerte og ustrukturerte data. Basert på dette ønsket DiBK å gjennomføre et forprosjekt for å vurdere mulighetene for bruk av "Big Data"-metodikk og teknologi for systematisk innhenting av indikatordata og analyser på byggkvalitet¹ og byggskader.



Figur 1 - Mulige datakilder for byggkvalitet og byggskader

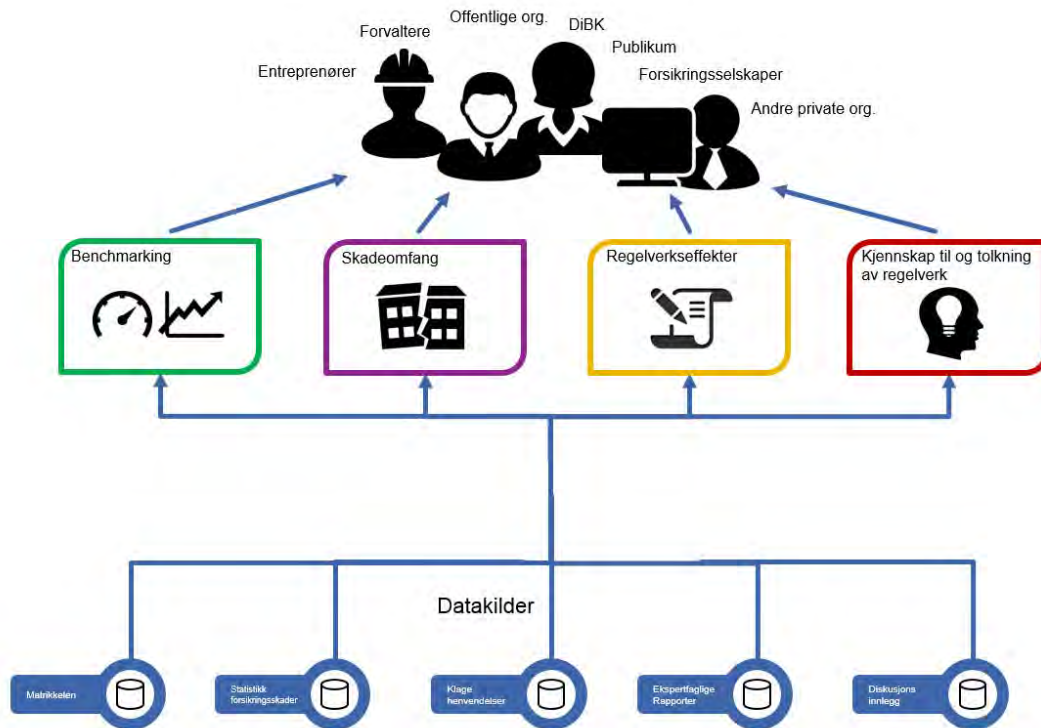
2.2 Visjon og målbilde

Basert på input fra deltakere i idéworkshops i forprosjektet kan man beskrive følgende visjon for en Big Data/analyse-løsning:

"DiBK og aktører i bransjen har et felles verktøy for å skaffe seg innsikt i omfanget av skader, feil og mangler og status på ulike aspekter av byggkvalitet. Aktører som utbyggere og forvaltere kan benchmarke seg mot andre aktører, og DiBK kan vurdere effekter av regelverksendringer og analysere hva som diskuteres vedrørende regelverk."

¹ For dette prosjektet har vi definert byggkvalitet som fravær av skader, feil og mangler.

En fremtidig teknologiløsning kan hente informasjon fra ulike kilder med både strukturerte og ustrukturerte data. Brukerne, som kan representere alt fra utbyggere, forvaltere og forsikringsselskaper til offentlige virksomheter og forskere, kan både hente ut indikatorer på byggkvalitet og gjennomføre detaljerte analyser på skadedata, klagestatistikker mv. Figuren nedenfor illustrerer brukergrupper, bruksområder og et utvalg av aktuelle datakilder.



Figur 2 - Visjon for Big Data-/analyse-løsning

2.3 Metodikk og gjennomføring

Forprosjektet har kartlagt og vurdert ulike faktorer eller områder som påvirker mulighetsrom og begrensninger. Dette legger grunnlag for anbefalinger for videre satsning. Disse områdene er illustrert i figuren på neste side, og funn og anbefalinger innenfor hvert enkelt område er oppsummert i egne avsnitt i det følgende.



Figur 3 - Områder som vurderes i forprosjektet

Kartleggingen er gjennomført gjennom idéworkshops, én-til-én-møter med dataeiere og øvrige interessenter, arbeidsmøter med representanter fra DiBK samt gjennomgang av dokumentasjon fra tidligere, relevante prosjekter.

Grovt sett kan dataeierne deles inn i følgende kategorier:

1. Entreprenører/utbyggere
2. Eiendomsforvaltere
3. Rådgivnings- og takseringsfirma
4. Tilsynsmyndigheter
5. Forsikringsbransjen
6. Tvistenemnder og domstoler
7. Eiere av diskusjonsfora på nett

Innenfor hver av kategoriene har det vært én eller et lite antall aktører som har deltatt i forprosjektet, se liste over informanter i *Vedlegg A Gjennomførte workshops og intervjuer*. Vi mener likevel våre observasjoner på et overordnet nivå er representative for aktuelle aktører i bransjen, da vi ser de samme observasjoner og trender hos de ulike aktørene vi har pratet med.

I forprosjektet er det også utviklet en testløsning ("proof of concept") med data fra et utvalg datakilder. Det er gjennomført ulike tester for å vurdere potensialet for anvendelse av ulike verktøy og metodikker, samt i hvilken grad de utvalgte datakildene egner seg for dette formålet.

Prosjektet er gjennomført i perioden ultimo april 2015 til ultimo september 2015.

2.4 Begreper og definisjoner

For å finne indikatorer på byggkvalitet må forprosjektet ha et bevisst forhold til hva som ligger i "byggkvalitet". Det er tydelig at det finnes ulike oppfatninger av hva som er viktig med hensyn på kvalitet. Dog er det enighet om at kvalitet kan defineres som det motsatte av skader, feil og mangler.

Begrepene "feil" og "mangler" brukes ofte om hverandre, og er i mange tilfeller synonymer. Begrepet "skader, feil og mangler" dekker alt fra "hard facts"-skader via regelverksbrudd og kontraktsbrudd til subjektiv oppfatning av kvalitet.

Sentrale begreper er i dette prosjektet definert som i tabellen på neste side.

Tabell 1 - Begreper og definisjoner

Begrep	Beskrivelse
Byggkvalitet	Det eksisterer ingen omforent definisjon, men de fleste forstår byggkvalitet som fravær av skader, feil og mangler.
Byggskader	Negativt avvik som framkommer gjennom redusert funksjonalitet/yteevne, med nedgradering, nyinvestering eller øking av forutsatte vedlikeholdskostnader. Kan skyldes mangelfull bygging, mangelfullt vedlikehold eller overbelastning/feil bruk. ²
Prosess- forårsakede skader	Skader som skyldes mangelfull bygging eller feil utførte utbedringer.
Byggefeil	Avvik eller svikt som ikke aksepteres av byggeier/tiltakshaver, bygningsmyndighetene eller andre berørte parter. ³
Big Data	Defineres ofte som data som innehar én eller flere av følgende egenskaper: <ul style="list-style-type: none"> 1. Stort volum 2. Høy oppdaterings-/endringshastighet 3. Variasjon i struktur og format (for eksempel tekstlige dokumenter) Det forekommer dog også at begrepet benyttes om visualisering og avansert analyse av interne og/eller eksterne data.
Analytics	Anvendelse av statistiske metoder for å skaffe innsikt og/eller utvikle prediktive modeller i data. Omtales ofte i forbindelse med Big Data.
Text Mining	Anvendelse av avanserte algoritmer og metoder for å gruppere, finne sammenhenger og/eller kategorisere tekstlig informasjon.

² Definisjon fra Byggforskserien (<http://www.sintefbok.no/Download.ashx?bookId=68>)

³ Definisjon fra Byggforskserien (<http://www.sintefbok.no/Download.ashx?bookId=68>)

3 Teknologiske muligheter: verktøy og metodikk knyttet til Big Data

Siden Big Data er et begrep som ofte benyttes også om data og løsninger som ikke nødvendigvis tilfredsstillende den strenge definisjonen av Big Data, dekker denne rapporten også metodikk og teknologi knyttet til berørte områder som "analytics", text mining og minnebaserte analyseløsninger.

Nedenfor forklares ulike typer teknologier/verktøy som kan være relevante for DiBK i en eventuell fremtidig løsning.

Tradisjonelle analyse-, visualiserings- og rapporteringsverktøy

Dette er typisk web-baserte verktøy for både lederrapportering og analyse. Blant verktøyene på markedet kan nevnes IBM Watson Analytics, SAS Visual Analytics, Tableau, Datazen, QlikView, SAP Business Objects og Microsoft BI. De ulike verktøyene har ulike styrker og svakheter både når det gjelder ytelse, funksjonalitet og fleksibilitet. Enkelte av verktøyene har for eksempel innebygget funksjonalitet for å hente inn data fra sosiale medier og websider.

I testløsningen i dette forprosjektet er primært SAS Visual Analytics benyttet, samt noe IBM Watson Analytics.

Verktøy for "Text mining"

Slike verktøy benyttes for å analysere og kategorisere tekstdokumenter ved hjelp av avanserte statistiske algoritmer. SAS Text Miner er det mest utbredte verktøyet av dette slaget, og er dette som er benyttet for tekstanalysene og tekstkategorisering i piloten i dette forprosjektet.

Verktøy for innholdskategorisering

Det finnes også verktøy for å kategorisere innhold basert på predefinerte regler (til forskjell fra text mining hvor systemet selv hjelper med å finne relevante ord og sammenhenger). Eksempel på slike verktøy er IBM Content Classification⁴ og SAS Content Categorization⁵.

Verktøy for dataintegrasjon ("ETL-verktøy")

For å systematisk innhente, bearbeide og sammenstille data fra ulike datakilder, benyttes egne verktøy for dataintegrasjon/ETL ("Extract Transform Load"). De fleste leverandører av database- og rapporteringsteknologi har også egne ETL-verktøy.

⁴ <http://www-03.ibm.com/software/products/no/content-classification>

⁵ http://www.sas.com/en_us/software/analytics/enterprise-content-categorization.html

Social Media Analytics-løsninger

Slike løsninger brukes for å følge med på strømmer på Facebook, Twitter, blogger og andre sosiale medier i sanntid. De brukes typisk av selskaper for å høste informasjon om kundesentimenter og meninger om produkter, samt for å få en indikasjon på selskapets omdømme. De kan også bli brukt for å analysere i hvilken grad et tema eller ord blir omtalt i positiv/negativ/nøytral tone.

Slike verktøy lisensieres ofte gjennom leverandører som drifter løsningen som en "Software as a service" (SaaS)-løsning, slik at organisasjoner ikke behøver å sette opp sin egen infrastruktur. Kostnaden avhenger typisk av antall twitter-meldinger og/eller websider som analyseres.

En begrensning i bruken av slike verktøy er at de gir lite nyttig informasjon om emneord som omtales sjelden. Et eksempel på dette er #tek10.

Web crawlere/scrapere

En *web crawler* søker systematisk gjennom sider på internett, og brukes for å hente inn innhold fra aktuelle nettsider. Roboten starter med et sett av URL, går inn på alle lenker på nettsiden og kopierer informasjonen som ligger under hver enkelt link. Verktøyet lagrer all informasjon som .txt-filer som kan analyseres videre med for eksempel et text mining-verktøy.

En *web scraper* er et lignende program og har som formål å høste data fra nettsider. Her lager man en "agent" for en nettside for å bestemme hvilke deler av nettsiden som skal høstes. På denne måten blir unyttig informasjon silt vekk, til forskjell fra en web crawler som tar med seg all informasjon som finnes på en side.

"*Visual web scraping/crawling*" er en type verktøy der man i stor grad slipper å programmere. Man markerer de delene av en nettside man vil ha ut, og hvordan de skal struktureres. Dette gjør det enklere å bearbeide og analysere innholdet videre enn f.eks .txt-filene som kommer fra en web crawler. Det finnes mange programmer for både web crawling og scraping.

Programvare for lagring av store datamengder

Når man eventuelt skal håndtere sensordata er det behov for andre typer software for lagring og prosessering. Apache Hadoop er den mest kjente software som benyttes for lagring og prosessering av svært store datamengder.

4 Funn

4.1 Funn fra testløsning/"Proof of concept"

Det er satt opp en testløsning som tester muligheten for å kunne trekke ut indikatorer og gjøre analyser basert på et utvalg datakilder. Testløsningen er satt opp på en PC med ulike verktøy fra SAS Institute, MS Excel samt IBM Watson Analytics. Hypoteser og konklusjoner fra disse testene er grundigere beskrevet i *Vedlegg C Detaljer vedrørende test/"Proof of concept"*.

Testen er delt inn i tre hovedområder som dekker ulike problemstillinger og benytter ulike metoder og datagrunnlag. Funn fra hvert testområde er beskrevet i påfølgende avsnitt. De tre testområdene er:

- 1) *Kan mer automatisert datainnsamling og analyse av eksisterende informasjon, både strukturert og ustrukturert, gi **innsikt i utvikling og omfang av skader, feil og mangler?***
- 2) *Kan analyse av data i eksisterende databaser gi **innsikt i effekt av regelverksendringer?***
- 3) *Kan analyse av informasjon på internett gi innsikt i bransjens og befolkningens **kjennskap til og tolkning av relevante regelverk?***

4.1.1 Resultater fra testområde 1: utvikling og omfang av skader, feil og mangler

*Kan mer automatisert datainnsamling og analyse av eksisterende informasjon, både strukturert og ustrukturert, gi **innsikt i utvikling og omfang av skader, feil og mangler?***

For dette testområdet benyttet vi et uttrekk av henvendelser som er registrert i en entreprenørs kunde- og klagehåndteringssystem. Uttrekket gjaldt et lite antall byggeprosjekter, og vi fikk informasjon om henvendelsestype, byggtipe, feiltype, produkt mv. Datauttrekket inneholdt også ustrukturert informasjon, i form av et fritekstfelt som beskrev hva henvendelsen gjaldt.

Sak/Kategori/Navn	Byggtype	Sted inne	Sted Ute	Produktgruppe	Produkt Inne	Produkt Ute	Feiltype/ Fullname	Fritekst
2. Reklamasjon	Trehus	Ånnet		VVS	Bereder			
2. Reklamasjon	Trehus	Entrè		Dør	Dørblad		Funksjonsfeil	Dørhåndtak løsnet på utsiden av y
1. Kundeservice	Trehus	Teknisk rom		Ånnet	Ånnet		Utførelse	Lokket i teknisk rom er ikke satt på
1. Kundeservice	Trehus	Kjøkken			Kjøkkenskuff		Funksjonsfeil	Skuffen under platetoppen går ikk

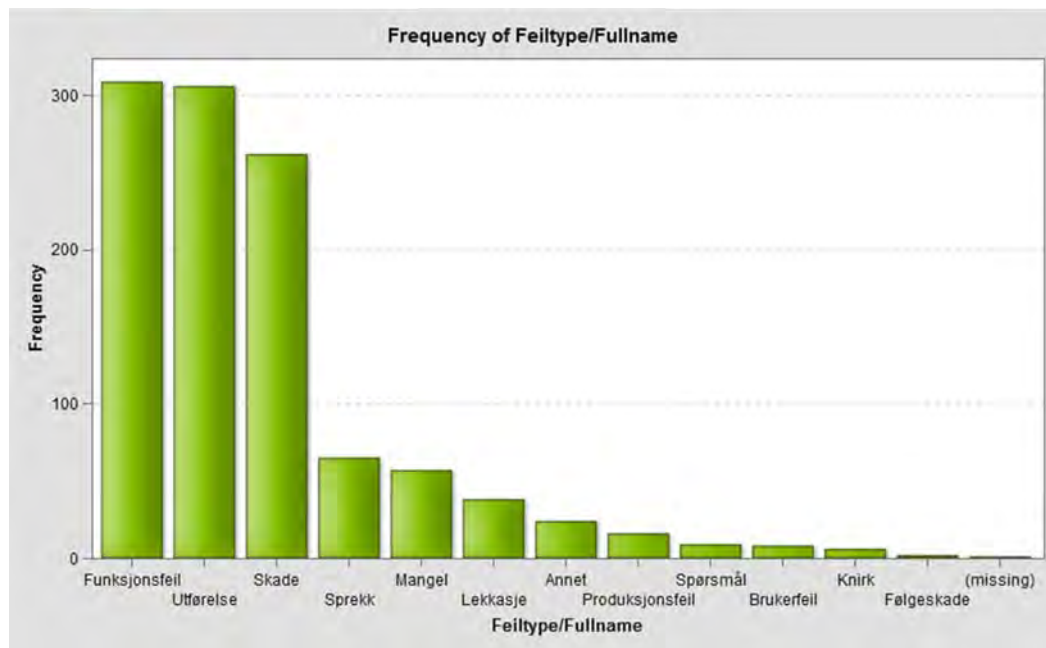
Figur 4 - Utdrag fra datasett om kundehenvendelser

Innenfor dette området ble det gjennomført tester på henholdsvis ren visualisering og på mer avansert text mining, og funnene er oppsummert nedenfor.

1. Visualisering av data om skader, feil og mangler

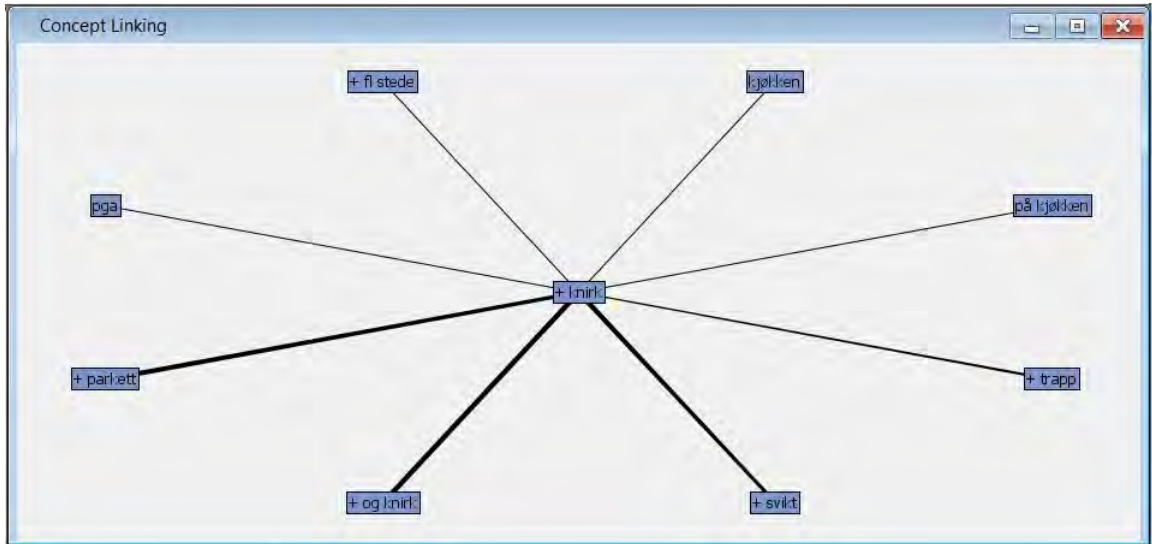
I denne testen ble dataene utforsket i et analyse- og visualiseringsverktøy⁶. Testen viste at man raskt kan få innsikt i omfang av innmeldte skader, feil og mangler ved å visualisere dataene i så vel tradisjonelle stolpediagrammer og grafer som i mer moderne ordskyer. Nedenfor listes noen eksempler på visualiseringer som lages på bakgrunn av slike analyser.

Stolpediagrammet nedenfor gir et raskt overblikk over hvordan henvendelsene/sakene er fordelt på ulike typer skader/feil/mangler.



Figur 5 - Visualisering av frekvens for ulike typer skader/feil/mangler

⁶ SAS Visual Analytics fra SAS Institute



Figur 8 – Visualisering av kobling mellom fraser

2. Automatisert kategorisering av registrerte skader, feil og mangler

I denne testen ville vi undersøke i hvilken grad det er mulig å bygge en modell, eller et regelsett, som kan brukes til å automatisk angi den mest sannsynlige typen skade/feil/mangel en henvendelse gjelder.

Tabellen nedenfor viser utdrag fra utfallet av den automatiserte kategoriseringen (f.eks. at "lys over speil virker ikke" og "heis virker ikke" mest sannsynlig er *funksjonsfeil* mens "Sprekker i overganger" sannsynligvis hører til feiltypen *sprekk*).

Text	Into: Feiltype_Fullname	Why	p_F_Fe.	Feiltype/Fullname
Store sprekker i hjørner og overganger vegg/tak - sjekk overtagelseprotokoll	SPREKK		19	0.994551 Sprekk
Lys over speil virker ikke.	FUNKSJONSFEIL		24	0.981986 Funksjonsfeil
Vannbåren varme i gulv på bad i 3. etasje virker ikke.	FUNKSJONSFEIL		24	0.981358 Funksjonsfeil
Sokkelstøvsuger virker ikke når man forsøker å starte den fra kjøkkenet	FUNKSJONSFEIL		24	0.978753 Funksjonsfeil
Ventilasjonsanlegg går ikke å regulere. Forserngsfunksjonen fungerer heller ikke.	FUNKSJONSFEIL		25	0.974145 Funksjonsfeil
Spot virker ikke helt (kommer og går).	FUNKSJONSFEIL		24	0.970626 Funksjonsfeil
Ventilasjonen vår virker ikke og har slått seg av.	FUNKSJONSFEIL		24	0.960736 Funksjonsfeil
Sprekker i overganger.	SPREKK		19	0.947704 Sprekk
Sprekker i overganger.	SPREKK		19	0.947704 Sprekk
Sprekker i overganger.	SPREKK		19	0.947704 Sprekk
Sprekker i overganger.	SPREKK		19	0.947704 Sprekk
Sprekk hjørne v vindu i stue, v terrassedør, v søyle i taket og v siden av søyle	SPREKK		20	0.940533 Sprekk
Sprekker vegg/tak	SPREKK		20	0.940533 Sprekk
Vanskelig å åpne dører når det er litt is på dekker. Avvist.	FUNKSJONSFEIL		33	0.930984 Funksjonsfeil
Sprekk i fliser vegg 3stk.	SPREKK		18	0.923107 Skade
1 etg. sprekk i 2 flis ved dusjdør	SPREKK		18	0.923107 Sprekk
Sprekk i flis v dusjvegg	SPREKK		18	0.923107 Sprekk
Oppstått nye sprekker i flis på bad i 1 etasje.	SPREKK		18	0.923107 Skade
Sprekker i fliser (2 etg)	SPREKK		18	0.923107 Sprekk
Sprekk i flis ved dusjdør	SPREKK		18	0.923107 Sprekk
Fire fliser med sprekk.	SPREKK		18	0.923107 Skade
Veggene på soverommet ved stuen rister når døren fra felles gang til lekeplassen lukker seg.	FUNKSJONSFEIL		33	0.921278 Skade
Verandadør er blitt skjev, henger ikke bra opp. Vanskelig å åpne og lukke.	FUNKSJONSFEIL		38	0.919301 Funksjonsfeil
Terrassedøren er vanskelig å åpne og lukke.	FUNKSJONSFEIL		38	0.919301 Skade
Heis virker ikke	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil
Virker som det er kun 1 innstilling på varmekabel, sjekkes	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil
Bad 2 etg. Bryter til spotter virker ikke.	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil
Spot over speil på bad i 2 etg virker ikke.	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil
2 etg. Låskasse på badør virker ikke	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil
Gulv har fått en veldig hul lyd i gulv i det ene hjørne, virker som massen under synker.	FUNKSJONSFEIL		24	0.917439 Utførelse
Varmekabler virker ikke som de skal	FUNKSJONSFEIL		24	0.917439 Funksjonsfeil

Figur 9 - Utdrag fra utfall av automatisk kategorisering

Tabellen over har også en henvisning til hvilken regel som ligger bak de enkelte kategoriseringene. Et utdrag fra reglene som verktøyet/algoritmen har kommet frem til er vist i tabellen nedenfor. For eksempel ser vi at ordet "lekk" antyder at saken dreier seg om en lekkasje.

Rule	Rule #	Target Value	Precision
qiø		1PROSJEKTERING	25.00%
fjernvarmeavtale		2INFORMASJON	33.33%
qipsplate		3FØLGESKADE	25.00%
kunde lure		4SPØRSMÅL	100.0%
kunde mene		5SPØRSMÅL	58.33%
av sutakplater		6SPØRSMÅL	56.25%
kunde		7SPØRSMÅL	28.26%
vindusforing		8PRODUKSJONSFEIL	45.45%
knirk		9KNIRK	64.15%
avtale		10ANNET	66.67%
tilsynsrapport		11ANNET	57.14%
klage		12ANNET	42.86%
utbedre		13ANNET	21.28%
lekkasje		14LEKKASJE	87.50%
lekk		15LEKKASJE	82.35%
vannlekkasje		16LEKKASJE	80.82%

Figur 10 - Utdrag fra regler for automatisert kategorisering

For å komme frem til disse reglene, benyttet vi text mining-verktøyet, og det ble anvendt mer avanserte text mining-teknikker. Verktøyet hjelper blant annet med å filtrere ut ord av visse typer ordklasser, foreslår tolkning av ord i ulike bøyingsformer som én og samme frase, samt filtrerer ut ord som forekommer så ofte eller sjelden at de ikke vil gi verdi til regelsettet/modellen.

For å kunne utvikle og benytte en slik automatisert kategorisering i en fullverdig løsning må det brukes tid på å tilpasse og justere synonymer, hvilke ord som skal ekskluderes, manuell overstyring av vektleggingen av enkelte ord, mv.

4.1.2 Resultater fra testområde 2: effekt av regelverksendringer

*Kan DiBK få økt **innsikt i effekt av regelverksendringer** ved analyse av data i eksisterende datakilder?*

For dette testområdet benyttet vi uttrekk av data fra Finans Norges registre for henholdsvis vannskader ("VASK") og brannskader ("BRASK"). Uttrekkene inneholder oppgjorte skadesaker med informasjon om utbetalt sum samt informasjon om blant annet skadetype og skadeårsak.

Innenfor dette området ble det testet om det har vært signifikant endring i skadeomfang etter 2010, da nye regler for henholdsvis vannskadesikring og brannskadesikring trådte i kraft. Funnene er oppsummert nedenfor.

1. Analyse av endringer etter innføring av krav til vannskadesikring

Hensikten med denne testen var å finne ut om man kan si noe om effekter av innføring av nye krav til vannskadesikring⁸ ved å analysere og gjøre statistiske tester på data fra VASK. Skadedataene ble visualisert i et analyse- og visualiseringsverktøy⁹, mens de statistiske testene ble gjennomført i Excel. Testen viste at VASK inneholder tilstrekkelig mengde med data for å kunne analysere om det har vært signifikante endringer i omfanget av skader knyttet til vanninstallasjoner uten overløp i bygg som er registret som nye (0 år), men at det knyttes en viss usikkerhet til resultatene på grunn av en betydelig andel manglende informasjon om byggets alder i registeret.

Det bør også bemerkes at grunnlaget ikke kan brukes til å si noe sikkert om *årsakssammenhenger* og den direkte koblingen til *regelverket*, da det kan være andre forhold knyttet til installasjoner i nye bygg som er med å påvirke forekomsten av skader.

Nedenfor beskrives datagrunnlaget, analysene som er gjennomført og vurdering av potensialet for anvendelse av disse dataene.

Om datagrunnlaget

VASK inneholder skadedata for tidsperioden 2008 – 2015 (tall per juni). Vi har imidlertid tatt skader fra 2015 ut av datagrunnlaget i analysene for å få et så likt grunnlag som mulig for de enkelte årene.

For hver enkelt skadesak finnes informasjon om utbetalt sum til den forsikrede, byggets alder, skadedato og geografi. Skadene er kategorisert i type installasjon (vannrør, avløp, vaskemaskin osv.), kilde (metallrør, vanntilkoblet maskin osv.), årsak (produktfeil, slitasje, frost, drenering osv.) og bransje (privat eller bedrift).

⁸ I praksis automatisk vannskadesikring

⁹ SAS Visual Analytics fra SAS Institute

utbetalt	fastsatt	skddato	grov_navn	detaljert_navn	kommune_fk	kommune_navn
63767	63767	07Jan2013	Utstyr	Varmeanlegg	1601	Trondheim
33920	33920	04Nov2013	Utstyr	Varmeanlegg	1924	Målselv
36811	36811	27Jan2014	Utstyr	Varmeanlegg	602	Drammen
0	35000	26Jan2015	Annet	Annen	1432	Førde
4584	4584	04Nov2013	Annet	Annen	219	Bærum
0	44159	09Feb2015	Annet	Annen	1925	Sørreisa
15370	15370	28Dec2008	Bygg	Inntrenging utenfra	1942	Nordreisa

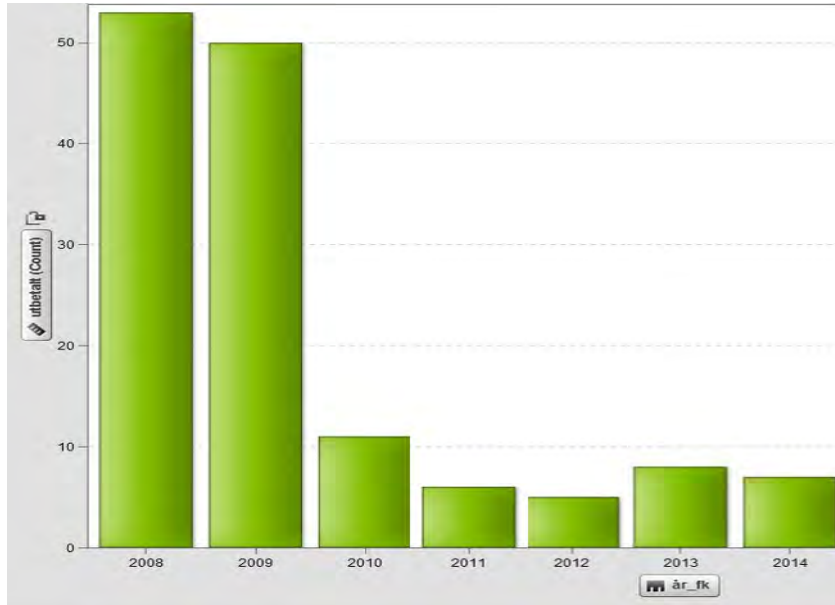
fylke_navn	år_fk	bygg_alder'	bransje_n	installasjon_beskrivelse
Sør-Trøndelag	2013	NULL	Bedrift	Varmeanlegg, gulvvarme, radiator
Troms	2013	38	Privat	Varmeanlegg, gulvvarme, radiator
Buskerud	2014	NULL	Bedrift	Varmeanlegg, gulvvarme, radiator
Sogn og Fjord.	2015	NULL	Bedrift	Varmeanlegg, gulvvarme, radiator
Akershus	2013	10	Privat	Varmeanlegg, gulvvarme, radiator
Troms	2015	11	Privat	Vannintr. utenfra over grunn
Troms	2008	NULL	Privat	Vannintr. utenfra over grunn

Figur 11 - Utdrag fra VASK

Statistiske tester

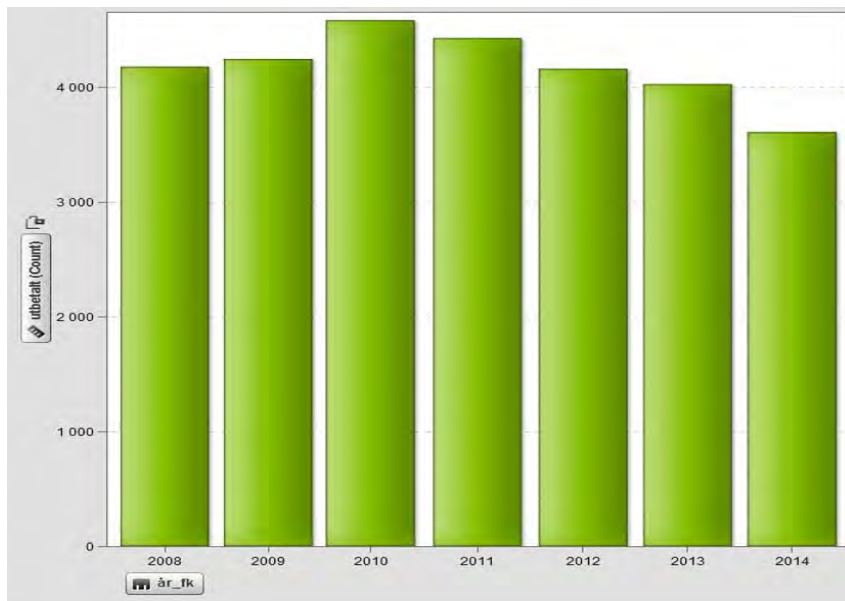
For å teste hypotesen om nedgang i skadesaker knyttet til vanntilkoblede maskiner etter 2010, har vi sammenlignet gjennomsnitt for henholdsvis *antallet skader* og *sum utbetalt* for skader knyttet til installasjonstype "Vaskemaskin, oppvaskmaskin og beholder" i nye bygg for periodene før og etter 2010¹⁰. Testen viser en signifikant nedgang i *antallet* slike vannskader, mens nedgangen i utbetalinger innenfor denne kategorien ikke har vært signifikant. Figuren på neste side viser utviklingen år for år i antall skader av nevnte type for nye bygg.

¹⁰ Gjennomsnitt for årene henholdsvis før og etter 2010 er sammenlignet ved hjelp av en to-utvalgs t-test.



Figur 12 - Antall vannskader for aktuelle installasjoner i nye bygg

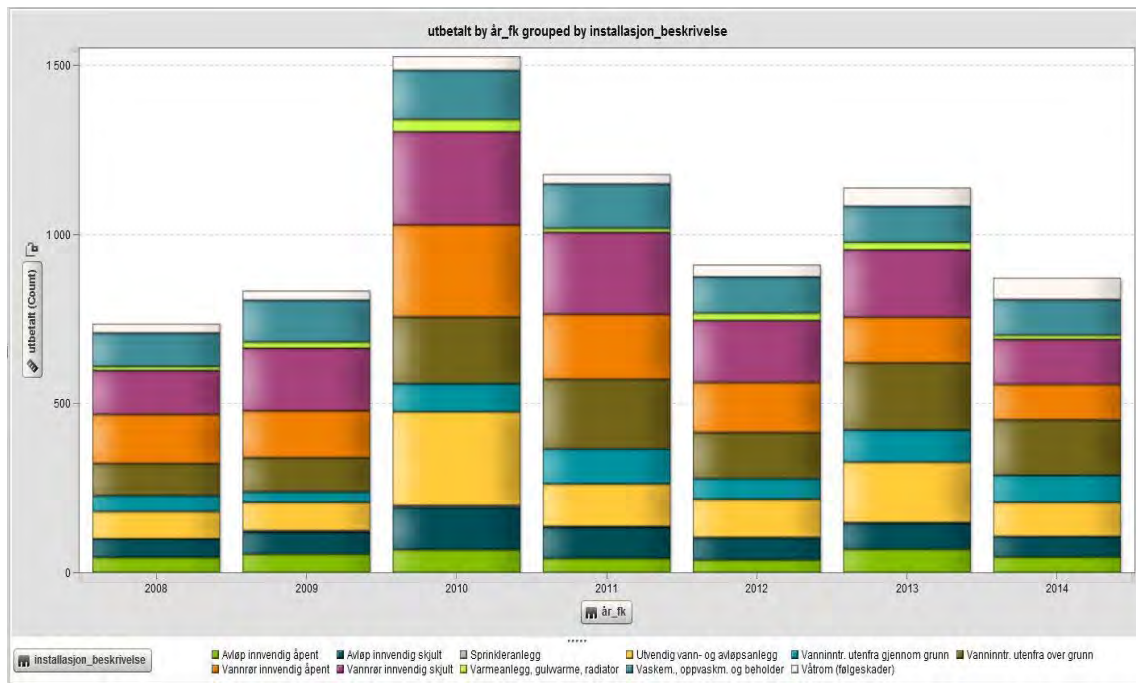
Antall skader når en ser bort fra byggets alder, har også gått ned de senere år, men statistiske tester viser *ikke* signifikant nedgang fra 2010. Dette støtter opp under hypotesen om at nedgang i skadeomfang kan ha sammenheng med endringer krav knyttet til nye bygg.



Figur 13 - Antall vannskader for aktuelle installasjoner i bygg av alle aldre

Visualisering

Dataene fra VASK kan analyseres og presenteres i mange ulike dimensjoner for å få innsikt i utviklingen i skadeomfang. Figuren nedenfor viser ett eksempel, der antall skader år for år er fordelt på de ulike installasjoner som hver skade er knyttet til.



Figur 14 - Visualisering av skadeutvikling og fordeling på aktuell installasjon

Kriterier for nytte

For at VASK skal kunne gi sikrere analyser, bør informasjon om byggets alder være mer komplett og pålitelig. Dette kunne vært løst ved å koble på informasjon om byggeår fra Matrikkelen, og/eller at forsikringsselskapene ble flinkere til å fylle ut denne typen informasjon.

Detaljer rundt metodikk og resultater kan finnes i *Vedlegg C Detaljer vedrørende test/“Proof of concept”*.

2. Analyse av effekter av innføring av krav til automatisk slukkeanlegg

Hensikten med denne testen var å finne ut om man kan si noe om effekter og/eller utilsiktede konsekvenser av innføring av nye krav til brannskadesikring¹¹ ved å

¹¹ I praksis automatisk vannskadesikring

analysere og utføre statistiske tester på data fra Finans Norges brannskaderegister (BRASK) og eventuelt vannskaderegister (VASK). Skadedataene ble visualisert i et analyse- og visualiseringsverktøy¹², mens de statistiske testene ble gjennomført i Excel. Testen viste at BRASK ikke inneholder tilstrekkelig informasjon om bygningstype til å kunne identifisere og isolere de skadesaker som gjelder bygg som er omfattet av regelverksendringen. Testen viste også at forekomsten av vannskader tilknyttet sprinkleranlegg i VASK er for liten til å kunne si hvorvidt forekomsten av dette har økt etter innføringen av de nye kravene til brannskadesikring.

Nedenfor beskrives datagrunnlaget, analysene som er gjennomført og vurdering av potensialet for anvendelse av disse dataene.

Om datagrunnlaget

BRASK inneholder skadedata for tidsperioden 1980 – 2015 (tall per mars). Vi har imidlertid tatt alle skader fra 2015 ut av datagrunnlaget i analysene for å få et så likt grunnlag som mulig for de enkelte årene.

For hver enkelt skadesak finnes informasjon om utbetalt erstatningsbeløp til den forsikrede, byggets alder (aldersgruppe), skadedato og geografi. Skadene er videre kategorisert i bransje (Hjem/innbo, villa/hus, bedrift osv.), næring (beboelse, industri, undervisning, osv.), branntype (kald eller varm, kilde (ildsted, varme arbeider, elektroniske apparater osv.)) og årsak (antatt påsatt, selvantennelse, lynnedslag osv.).

erstatning	skadedato	kommune_navn	fylke_navn	år_fk	aldersgruppe	branntype	næring_navn	årsak_beskrivelse
22563	2014-10-20	Halden	Østfold	2014	Ukjent	Kald	Jordbruk, skogbruk og fiske	Elektrisk fenomenskade
5000	1988-08-22	Halden	Østfold	1988	Ukjent	Kald	Industri	Elektrisk fenomenskade
400	1985-01-28	Halden	Østfold	1985	Ukjent	Varm	Varehandel, reparasjon av motor	Annet eller ukjent
1832	1988-02-01	Halden	Østfold	1988	Ukjent	Kald	Varehandel, reparasjon av motor	Annet eller ukjent
434789	1987-10-19	Halden	Østfold	1987	Ukjent	Varm	Overnattings- og serveringsvirksomhet	Annet eller ukjent
12296	2002-05-27	Halden	Østfold	2002	Ukjent	Kald	Industri	Annet eller ukjent
11850	1987-11-16	Halden	Østfold	1987	Ukjent	Kald	Industri	Teknisk svikt
0	2011-10-17	Halden	Østfold	2011	Ukjent	Kald	Varehandel, reparasjon av motor	Annet eller ukjent

Figur 15 - Utdrag fra BRASK

Statistiske tester

For å teste hypotesen om nedgang i brannskadesaker for bygg hvor automatisk slukking er påbudt, ønsket vi å sammenligne gjennomsnitt for henholdsvis *antallet skader* og *sum utbetalt* for skader knyttet til *blokker* for periodene før og etter 2010¹³. Mangel på kategorier som kan brukes til å identifisere de bygg som har krav til automatisk slukking, gjorde at vi ikke fikk testet dette.

¹² SAS Visual Analytics fra SAS Institute

¹³ Gjennomsnitt for årene henholdsvis før og etter 2010 er sammenlignet ved hjelp av en to-utvalgs t-test.

Visualisering

Tilsvarende som for VASK, egner BRASK seg godt for utforskende analyser og visualiseringer da datasettet har mange dimensjoner som det er aktuelt å analysere på.

Kriterier for nytte

For at BRASK skal kunne gi sikrere analyser, bør informasjon om byggets alder og øvrig informasjon om bygget være mer komplett og pålitelig. Dette kunne vært løst ved å koble på informasjon fra Matrikkelen, og/eller at forsikringsselskapene ble flinkere til å fylle ut denne typen informasjon.

Detaljer rundt metodikk og resultater kan finnes i *Vedlegg C*
Detaljer vedrørende test/"Proof of concept".

4.1.3 Resultater fra testområde 3: kjennskap til og tolkning av regelverk

*Kan DiBK få **innsikt i** bransjens og befolkningens **kjennskap til og tolkning av relevante regelverk** gjennom analyse av informasjon på internett?*

For dette området testet vi på henholdsvis uttrekk av Twitter-meldinger om relevante emneknagger og uttrekk fra innlegg på byggebolig.no. Funn fra hver av testene er oppsummert nedenfor.

1. Analyse av innlegg på Twitter

Hensikten med denne testen var å finne ut om man kan få innsikt i hva som diskuteres om regelverk ved å analysere Twitter-meldinger med emneknagger som #TEK10 og #byggutenasoke. Twitter-meldinger ble testet, samlet inn og analysert via to ulike analyse- og visualiseringsverktøy (SAS Visual Analytics og IBM Watson Analytics). Testen viste at forekomsten av meldinger med relevante emneknagger er for lav til å kunne gjennomføre analyser. Dersom DiBK klarer å skape økt aktivitet rundt utvalgte emneknagger, kan imidlertid de testede verktøyene benyttes til å innhente og analysere relevante meldinger.

2. Analyse av innlegg på byggebolig.no

Hensikten med denne testen var å finne ut om man kan få bedre oversikt over det som diskuteres om regelverk, og således få kunnskap om både kjennskap til og

tolkning av regelverk, ved å innhente og analysere tekst fra innlegg i diskusjonsforumet på byggebolig.no. For denne testen brukte vi "web crawler"-funksjonaliteten i analyse- og visualiseringsverktøyet fra SAS Institute, som henter inn all tekst som finnes på oppgitt webside samt sider som det linkes til.

Testen viste at web crawler-funksjonalitet ikke er tilstrekkelig for å få hentet inn tekst fra diskusjonene på en hensiktsmessig måte. Enten må det benyttes mer spesialiserte verktøy som henter tekst fra pre-definerte områder på en nettside (se informasjon om "web scrapers" i avsnitt 3), eller man må hente informasjonen direkte fra databasen til de som drifter diskusjonsforumet.

Manuell utforsking av innleggene på byggebolig.no viser at det finnes en rekke diskusjoner som gjelder både kjennskap til, tolkning av og meninger om regelverk.



Figur 16 - Eksempel på diskusjonsinnlegg på byggebolig.no

4.2 Potensialet i ulike datakilder

På overordnet nivå ser vi størst potensiale i bruk av strukturerte data i FDV- og sakssystemer som brukes til registrering av avvik, skader, feil og mangler, samt forsikringsdata, og mindre potensiale i bruk av dokumenter i "rapportform". Nedenfor følger oppsummering av observasjoner og innspill fra den/de aktørene vi har pratet med innenfor hver type av "dataeiere", samt enkelte, generelle observasjoner.

4.2.1 Potensialet i data fra dataeiere som er vurdert

Entreprenører og utbyggere sitter på informasjon om skader, feil og mangler som blir oppdaget under bygging, ved overtakelse og den første tiden etter overtakelse.

Tidligere ble denne informasjonen registrert med "penn og papir". Det er fremdeles en del som registreres på papir og i ustrukturert digital form, men hos våre informanter registreres henvendelser fra kundene etter overtakelse i et CRM-/saksbehandlingssystem. Uttrekk fra systemet til en av deltakerne i dette forprosjektet tyder på god datakvalitet, men noe inkonsistens i kategorisering av skader.

Eiendomsforvaltere sitter på informasjon om avvik (skader, feil og mangler) som oppdages eller oppstår i bruks-/driftsfasen av et bygg. Informasjon fra deltakerne i forprosjektet, inklusive demonstrasjon av avviksmodulen i FDV-systemet, indikerer at disse dataene er lagret på en strukturert måte og at det bør være relativt enkelt å benytte seg av disse dataene. Forprosjektet har ikke mottatt uttrekk/testdata fra eiendomsforvaltere, og kan dermed ikke konkludere på datakvaliteten.

Rådgivnings- og takseringsfirma utarbeider rapporter og vurderinger i dokumentform. I tillegg blir noe registrert i egne systemer. Uttrekk av disse rapportene i seg selv bør være relativt enkelt, men nytten anses å være begrenset da eksempel rapporter vi har sett indikerer at sakene omhandler flere ulike tema og at det vil være komplisert å trekke ut essensen eller å kategorisere basert på text mining/automatisert kategorisering. Takseringsforbundet opplyser også at stadig større andel av registreringer gjøres digitalt og at data lagres i strukturerte databaser.

Tilsynsmyndigheter, som plan- og bygningsetatene, sitter på tilsynsarkiv og byggesaksarkiv. Blant sakene som er registrert i tilsynsarkivet er både saker som representerer *faktiske* avvik og saker der det er meldt om *mulig* avvik, og disse er ikke nødvendigvis enkle å skille. Forprosjektet har ikke analysert uttrekk fra disse datakildene.

Forsikringsbransjen sitter på informasjon om skadesaker. *Finans Norge* drifter og vedlikeholder databasene "VASK" (vannskadestatistikk) og "BRASK" (brannskadestatistikk) som inneholder informasjon om registrerte skader. Databasene inneholder kun de saker som har endt med utbetaling. Databasen inneholder informasjon om blant annet årsak, type installasjon, geografi (kommunenivå) og byggets alder, men datakvaliteten er varierende. Særlig for skader fra lengre tid tilbake er det en del manglende data.

VASK og BRASK er allerede i dag åpne og tilgjengelige for statistiske uttrekk¹⁴, dog uten filtermulighet på geografi. Geografi er holdt utenfor da skadene kan bli identifiserbare ved statistiske uttrekk på små kommuner.

Tvistnemnder og domstoler besitter informasjon om saker der boligeier ønsker å klage på ulike forhold knyttet til kvalitet og ansvar for skader, feil og mangler.

¹⁴ <https://www.fno.no/statistikk/skadeforsikring/vask/> og <https://www.fno.no/statistikk/skadeforsikring/Brannstatistikk/>

Boligtvistnemnda publiserer på sine nettsider oppsummering av alle saker de har behandlet, i form av tekstlige dokumenter. Forprosjektet har analysert disse dokumentene, men kategorisering viste seg utfordrende da hver tvistesak omhandlet en rekke ulike punkter og resulterte i klagers medhold på visse punkter mens andre klagepunkter ble avvist. I registrene for *domsavsigelser* i tingretter og lagmannsretter finnes informasjon om blant annet hva boligkjøpere klager på. På lovdata.no er det åpen tilgang til avsagte dommer, men begrenset søkefunksjonalitet gjør det vanskelig å hente ut relevante dommer.

Diskusjonsfora på nett inneholder innlegg som omhandler både spørsmål om tolkning av regelverk, praktiske spørsmål knyttet til bygging og rehabilitering, erfaringer med utbyggere og spørsmål knyttet til håndtering av ulike typer skader, feil og mangler. Eieren av diskusjonsforumet vi har pratet med har forklart at de lagrer alle innlegg og tilhørende metadata (bruker, tidspunkt, emner etc.) i en database. Forprosjektet har ikke hatt tilgang på uttrekk av denne databasen og har således ikke kunnet vurdere datakvalitet.

4.2.2 Andre datakilder

Blant andre dataeiere som kan være av interesse, er firmaer som analyserer **radonmålinger**. Det foretas årlig mange tusen radonmålinger i norske bygg. Registreringene fra disse målingene er imidlertid ikke samlet i ett felles system, men er lagret hos de ulike aktørene som tilbyr radonmålinger og analyse av målingene. Statens Strålevern vurderer å kjøre et prosjekt for å samle inn radonmålinger fra 2010 og frem til i dag for å kunne vurdere utviklingen etter innføringen av TEK10, men dette arbeidet har ikke startet ennå.

En annen type data som ikke er vurdert i detalj i dette prosjektet, men som vi observerer at får større og større utbredelse, er **sensordata**. Etter hvert som sensorer i større og større grad blir installert i nybygg, blir det skapt enorme mengder data om blant annet temperatur, luftfuktighet og energiforbruk, som kan benyttes som indikatorer på luftkvalitet og energieffektivitet. Statsbygg er med sitt SmartGrid-prosjekt¹⁵ blant aktørene som har startet å teste ut nytten av slik teknologi. Det skjer en rivende utvikling på dette området, både når det gjelder utvikling av selve sensorene og når det gjelder infrastruktur og verktøy for overføring og analyse av disse dataene.

Se for øvrig *Vedlegg B*

Detaljerte vurderinger av datakilder for detaljer vedrørende hver enkelt datakilde som er vurdert.

¹⁵ <http://www.statsbygg.no/Samfunnsansvar/fou/Effektiv-drift/SmartGrid/>

4.2.3 Generelle observasjoner knyttet til aktuelle datakilder

Vi har gjort følgende *generelle observasjoner* når det gjelder struktur, kvalitet og andre faktorer som påvirker kompleksiteten ved å hente ut data/dokumenter og muligheten til å få nytte av datakildene.

Høy grad av digitalisering og modernisering

- Den teknologiske utviklingen har medført høy grad av modernisering av verktøy og teknologi i bransjen, og det er også planer for videre digitalisering og modernisering de nærmeste årene. Dette gjelder også utrulling av ny teknologi som kan representere nye datakilder, slik som sensorer som registrerer luftfuktighet, temperatur etc.
- Aktørene vi har pratet med registrerer i stor grad skader, feil og mangler i digitale systemer, og lagrer dette i databaser. Det er lite som nå kun registreres i form av skriftlige rapporter.
- Data fra lengre tid tilbake er i mindre grad digitalisert og strukturert, og består av alt fra arkiverte papirrapporter til skannede maskinskrivne dokumenter. Struktur og form varierer altså over tid, og det vil derfor være en omfattende oppgave å sette opp innhenting av skriftlige rapporter.
- Mange aktører planlegger migrering til nye løsninger i nær fremtid, hvilket stiller krav til fleksibilitet i løsningen som skal trekke ut/motta dataene.
- Overgang til digitale arbeidsverktøy med registrering av stadig mer data gir et stadig mer omfattende analysegrunnlag. Samtidig gjør dette at datagrunnlaget og forutsetningene for dette stadig er i endring.

Liten grad av standardisering

- Aktørene benytter i liten grad standarder slik som bygningsdelstabellen, standardiserte skade- og feiltyper og bygningstype. Når det gjelder identifikasjon av bygget anvendes heller ingen standard; her benyttes vekselvis gateadresse, gårds- og bruksnummer og bygningsnummer. Ingen av dataeierne vi har pratet med, har i dag en kobling til Matrikkelen.
- Det observeres heller ingen standardisering når det gjelder systemer/teknologi og IT-infrastruktur.

Datakvalitet

- Datakvaliteten i de analyserte kildene er jevnt over god.
- Kvaliteten varierer *over tid* da systemer har blitt skiftet ut over de siste årene.

4.2.4 Standardiseringshindre

Forprosjektet har identifisert et behov for å styrke standardiseringsarbeidet, og DiBK bør anbefale aktørene i bransjen å benytte gitte standarder, som for eksempel bygningsdelstabellen fra Byggforsk-serien fra SINTEF Byggforsk. Det er også aktuelt å samarbeide tettere med verktøyleverandører som for eksempel Vianova for å sikre at standarder blir anvendt i deres registreringssystemer.

For å bøte på utfordringene anbefales det også å videreutvikle løsningen for "automatisert kategorisering" av skader, feil og mangler som er testet ut i forprosjektet.

4.3 Øvrige observasjoner

4.3.1 Varierende grad av motivasjon hos dataeiere

Dataeierene har i utgangspunktet varierende evne og motivasjon for å dele data med DiBK. Enkelte dataeiere ser på sine data som en viktig eiendel som de selv ønsker å tjene penger på og dermed ikke ønsker å dele med andre. Andre dataeiere har ingen motforestillinger mot å dele data, men vil ha kostnader knyttet til å tilgjengeliggjøre sine data som må forsvares gjennom forretningsmessig nytte, pålegg eller andre motivasjonsfaktorer.

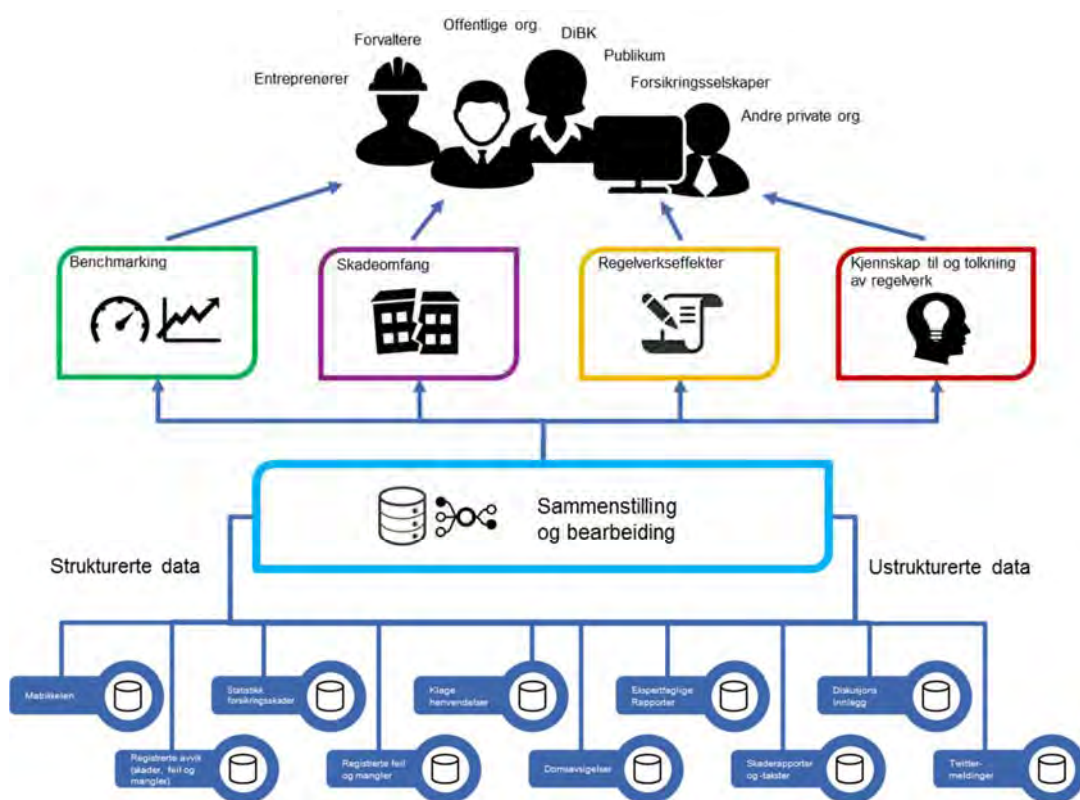
4.3.2 Liten grad av anvendelse av statistikk og analyse

Bransjen er relativt umoden på bruk av analyse- og rapporteringsverktøy. Få av aktørene vi har pratet med jobber systematisk med analyser av avviks- og skadedata. Kjennskapen til og kompetansen innenfor analyse- og rapporteringsverktøy samt dataintegrasjon er relativt lav.

5 Vurderinger

5.1 Nytteverdi og mulige bruksområder

Figuren nedenfor illustrerer konseptuelt ulike brukergrupper og bruksområder samt aktuelle underliggende datakilder og databehandling. De mest aktuelle bruksområder som vi har identifisert etter testing og kartlegging av et begrenset utvalg datakilder, kan i hovedsak deles i fire områder: innsikt i effekter av regelverk, indikatorer og analyse på skadeomfang, benchmarking og innsikt i kjennskapet til og tolkning av regelverk.



Figur 17 - Konseptuell løsningskisse

1. Innsikt i effekter av regelverk gjennom analyser av data fra ulike dataeiere.

Myndighetene ønsker bedre innsikt i effekter av regelverk, og vil kunne ha nytte av å analysere utviklingen i visse typer skader for å få bedre forståelse for effekter av ulike regelverk. Virkningen av regelverk kan vurderes ved hjelp av et analyseverktøy hvor data kan leses inn fra ulike eksterne kilder. Ved hjelp av statistikk- og analyseverktøy som de som er testet i dette forprosjektet, kan det gjennomføres grundige analyser, inkludert statistiske tester, av utviklingen av skader, feil og mangler over tid samt analysere årsakssammenhenger og mulige drivere.

Finans Norges registre for henholdsvis vannskader og brannskader er blant datakildene som kan benyttes til dette formålet.

2. Dashboard med indikatorer på byggkvalitet og byggskader, feil og mangler.

Basert på funnene i forprosjektet mener vi at offentlige myndigheter og øvrige interessenter kan følge med på utviklingen av indikatorer for omfang av byggskader, feil og mangler gjennom et dashboard med analysemuligheter. Den skisserte løsningen støtter funksjonalitet for visning av dashboard med viktigste indikatorer samt "drilling" ned i spesifikke indikatorer og skadetyper. Løsningen gir også gode muligheter for å analysere utvikling over tid, samt analysere sammenhenger med underliggende faktorer.

Datakildene som ut fra nytte- og kompleksitetsvurderinger synes mest hensiktsmessige å prioritere for etablering av dataauthenting, er *avviksregistre* fra forvalteres FDV-systemer, *saks- og avviksregistreringssystemer* hos utbyggere, databasen for diskusjoner på byggebolig.no og *skadedata* for forsikringsdata. For å komplettere informasjonen som er tilgjengelig i Finans Norges registre samt få tilgang på flere detaljer om skadene, kan og bør forsikringsselskapene selv involveres i et eventuelt hovedprosjekt.

Det vurderes *ikke hensiktsmessig* å sette opp automatisert uttrekk av tekstdokumenter, da nytten av disse synes begrenset på grunn av varierende innhold og tolkning over tid. De er dessuten tilgjengelige fra en begrenset tidsperiode, og for de senere år og fremtiden synes den mest verdifulle informasjonen tilgjengelig på strukturert form.

Blant **indikatorene** som kan være aktuelt å vise i et slikt dashboard, er for eksempel:

- **Antall skader, feil og mangler** som er registrert hos et utvalg utbyggere, entreprenører, og forvaltere, **kategorisert innen gitte skadekategorier**. Som tidligere nevnt, inneholder de ulike systemene ulike skadekategorier, men ved hjelp av algoritmer for automatisert kategorisering som angir sannsynlig skadekategori, kan data fra ulike kilder aggregeres og til sammen utgjøre indikatorer på utviklingen. Merk at kategoriseringsmetoden bør testes på data fra flere aktører før den endelig rulles ut.
- **Skadedata fra forsikringsselskapene** om innmeldte skader og oppgjorte skader innenfor ulike typer (for eksempel vannskader og brannskader)
- Antall **klager** som er meldt inn til boligvistnemnda
- Hvilke **skader, feil og mangler** som er mest omtalt på utvalgte **diskusjonsfora** på nett.

Samme datakilder kan benyttes til analyser av utvikling av skadeomfang og dypere analyse av skader fordelt på ulike dimensjoner som geografi, byggtyper mm. Eksempelvis kan utvikling av omfang av vannskader og brannskader analyseres

basert på data fra VASK og BRASK. Indikatorer på utvikling i opplevd/oppfattet byggkvalitet kan skaffes fra analyse av innlegg i diskusjonsfora.

3. Benchmarking av omfang av skader, feil og mangler mellom ulike entreprenører og forvaltere

Den enkelte entreprenør/utbygger og forvalter kan benchmarke seg mot andre aktører. Er fordelingen av feiltyper lik som for andre aktører? Den enkelte entreprenør/utbygger og forvalter kan også følge utviklingen over tid i omfang av byggskader, feil og mangler knyttet til egne prosjekter/bygg og gjøre "ad hoc"-analyser av sammenhenger i datagrunnlaget.

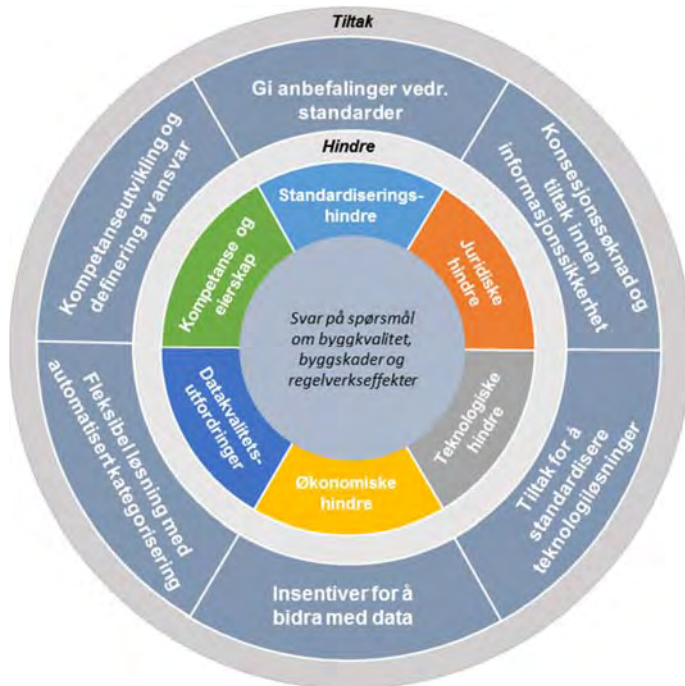
4. Innsikt i tolkninger og grad av forståelse av regelverk, basert på telling av antall innlegg i relevante diskusjonsfora som omhandler visse typer skader, feil og mangler.

Da Twitter viste seg å ikke egne seg for analyse av omtale av relevant regelverk, anbefales det å sette analyser av sosiale medier på vent inntil det eventuelt blir flere merkinger av aktuelle emneknagger.

Direktoratets og kommunenes ansatte kan følge med på hvor godt regelverk er forstått ved å få rapporter fra utvalgte diskusjonsfora på antall innlegg som omtaler utvalgte tema. De kan også grave seg dypere ned i diskusjoner om tema som de ser går igjen. En slik løsning forutsetter at DiBK får tilgang til data direkte fra eieren/driftsansvarlig for forumet.

5.2 Hindre og mulige tiltak

Basert på funnene som er beskrevet i avsnitt 4 *Funn* vurderer vi at det er seks ulike hindre som kan motvirke oppnåelse av visjon og målbilde for en Big Data-løsning; *Juridiske hindre, teknologiske hindre, økonomiske hindre, datakvalitetsutfordringer, hindre knyttet til kompetanse og eierskap samt hindre knyttet til manglende standardisering.* Vi vil her beskrive hvordan vi vurderer disse hindringene, og hvilke tiltak vi vurderer at bør iverksettes for å løse utfordringene.



Figur 18 - Hindre og mulige tiltak

5.2.1 Juridiske hindre

5.2.1.1 Lovgivning knyttet til personopplysninger

Ved bruk av datakilder bestående av personopplysninger er det nødvendig å ta hensyn til relevant lovgivning. For DiBK vil dette innebære både etablering av avtaler med aktørene det innhentes data fra og å iverksette rutiner for informasjonssikkerhet.

Personopplysninger er opplysninger eller vurderinger som kan knyttes til enkeltpersoner (eks. navn, adresse, telefonnummer, e-postadresse og bilnummer) eller informasjon om atferdsmønstre (eks. hva du handler, hvor du beveger deg og netthistorikk)¹⁶. Sensitive personopplysninger er opplysninger knyttet til eks. etnisk bakgrunn, politisk, filosofisk eller religiøs oppfatning, kriminelle forhold, helseforhold og liknende.

Datatilsynet utarbeidet i 2013 en rapport som omfatter Big Data og personvern¹⁷.

Det rettslige utgangspunktet er først og fremst den norske **personopplysningsloven**¹⁸ og **personopplysningsforskriften**. Datagrunnlag

¹⁶ Hva er en personopplysning (<https://www.datatilsynet.no/personvern/Personopplysninger>)

¹⁷ Datatilsynet: Big Data – personvernprinsipper under press (http://www.datatilsynet.no/Global/04_planer_rapporter/Big%20Data_web.pdf)

¹⁸ Personopplysningsloven, lov av 14. april 2000 nr. 31 om behandling av personopplysninger, bygger på det europeiske personverndirektivet (Europaparlamentets og Rådets direktiv 95/46/EF av 24. oktober 1995 om beskyttelse av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger).

benyttet som grunnlag for Big Data-analyser utløser plikter og rettigheter etter denne lovgivningen dersom datagrunnlaget inneholder personopplysninger.

Samtykke

Personopplysningsloven presiserer at opplysninger ikke kan brukes til andre formål enn det opprinnelige formålet for innsamlingen, med mindre den som har registrert dataen samtykker til dette. For å innhente samtykke skal virksomheten som innhenter personinformasjonen utarbeide en personvernerklæring som blant annet skal presisere hvem som er behandlingsansvarlig og hva som er formålet med innsamlingen¹⁹. Dette er del av informasjonsplikten i personopplysningsloven.

Krav til informasjonssikkerhet

Krav til informasjonssikkerhet ved behandling av personopplysninger finnes i personopplysningslovens § 13 og personopplysningsforskriftens kapittel 2. Kravene omfatter blant annet at virksomheten skal ha:

- Definerte sikkerhetsmål²⁰
- Sikkerhetsstrategi²¹
- Gjennomført risikovurdering²²
- Organisasjonskart som beskriver ansvar og myndighet i forhold til informasjonssikkerhet og drift²³
- Beskrivelse av informasjonssystemet²⁴

Databehandleravtaler

Hvis en virksomhet setter ut hele eller deler av behandlingen av personopplysninger til andre virksomheter benyttes det såkalte databehandlere. Forholdet mellom en behandlingsansvarlig virksomhet og databehandleren skal være regulert i en avtale; databehandleravtale. Dette reguleres av personopplysningsloven § 13, jf. § 15. DiBK vil opptre som en databehandler dersom de i en fremtidig løsning innhenter data inneholdende personopplysninger fra tredjeparter, selv om datagrunnlaget anonymiseres før det gjøres tilgjengelig for brukerne. Dersom DiBK setter ut drift og forvaltning av løsningen til en ekstern part, vil denne igjen opptre som en databehandler for DiBK.

En databehandleravtale kan være en frittstående avtale mellom partene, eller en integrert del av annet avtaleverk. En *databehandler* kan ikke behandle personopplysninger på en annen måte enn det som er skriftlig avtalt med den behandlingsansvarlige. Den behandlingsansvarlige skal forsikre seg om at databehandleren har tilstrekkelig sikkerhetsnivå, jf. personopplysningsloven § 15. En

¹⁹ Personvernerklæring: Hva skal den inneholde? (<http://www.datatilsynet.no/Sikkerhet-internkontroll/Personvernerklæringer>)

²⁰ Personopplysningsforskriften § 2-3

²¹ Personopplysningsforskriften § 2-3

²² Personopplysningsforskriften § 2-4

²³ Personopplysningsforskriften § 2-7

²⁴ Personopplysningsforskriften § 2-7

veileder til databehandlingsavtaler og mal for kontrakt kan lastes ned fra Datatilsynet sine hjemmesider²⁵.

Lovgivning under endring

EU-kommisjonen har jobbet med modernisering av EUs personvernregler siden 2012. Kommisjonen foreslår å erstatte det gjeldende personverndirektivet med en *forordning*. Den nye personvernlovgivningen vil bli gjeldende to år etter at de har blitt vedtatt, trolig fra 2017 eller 2018²⁶. KPMG kjenner per d.d. ikke til om den nye lovgivningen direkte vil påvirke lovgivningen rundt bruk av Big Data-teknologi.

5.2.1.2 Tiltak for å imøtekomme juridiske hindre

KPMG har identifisert tre tilnærminger DiBK kan benytte seg av for å kunne benytte data med personopplysninger til analyseformål:

1. Innhente samtykke

Personopplysningsloven presiserer at opplysninger ikke kan brukes til andre formål enn det opprinnelige formålet for innsamlingen, med mindre den som har registrert dataen samtykker til dette. For å innhente samtykke skal virksomheten som innhenter personinformasjonen utarbeide en personvernerklæring som blant annet skal presisere hvem som er behandlingsansvarlig og hva som er formålet med innsamlingen²⁷. Dette er en del av informasjonsplikten i personopplysningsloven. Hvis DiBK ønsker å gjennomføre et hovedprosjekt hvor data som inneholder personopplysninger regelmessig skal innhentes anbefaler KPMG at DiBK avtaler med aktører som samler inn data å oppdatere sine standard personvernerklæringer. Ved at kommuner, aktører i byggeprosessen og byggevaremarkedet oppdaterer sine personvernerklæringer med nødvendig informasjon kan DiBK løpende innhente ønsket informasjon til analyseformål.

2. Anonymisere data

For de data der kobling mot eiendomsregister/matrikkelen ikke er nødvendig, er den sikreste måten å sikre seg mot brudd på personopplysningsloven å anonymisere datagrunnlaget som skal analyseres. I så fall vil den videre behandlingen ikke være regulert av personopplysningsloven, uavhengig av om det er den opprinnelige innsamleren eller en ny aktør som står ansvarlig for behandlingen. Anonyme opplysninger defineres som opplysninger som det ikke er mulig å knytte til et identifiserbart individ når man tar i betraktning alle de hjelpemidlene som med rimelighet kan tenkes brukt for å identifisere vedkommende, enten av den behandlingsansvarlige eller av en annen aktør.

Anonymisering av opplysninger blir imidlertid stadig mer utfordrende. I forbindelse med innhenting av datasett til Big Data-analyse, kan en virksomhet ofte ikke med

²⁵ Databehandleravtaler etter personopplysningsloven og helseregisterloven (https://www.datatilsynet.no/Global/04_veiledere/Databehandleravtaler_20090526.pdf)

²⁶ Datatilsynet: Ny personvernlovgivning på vei fra EU (<https://www.datatilsynet.no/Nyheter/20151/Ny-personvernlovgivning-pa-vei-fra-EU>)

²⁷ Personvernerklæring: Hva skal den inneholde? (<http://www.datatilsynet.no/Sikkerhet-internkontroll/Personvernerklæringer>)

sikkerhet vite hvorvidt de anonymiserte datasettene kan la seg re-identifisere av en ukjent tredjepart. Gjennom sammenstilling av data fra flere kilder kan det oppstå *risiko* for at enkeltindivider kan identifiseres fra datasett som i utgangspunktet er anonyme.

Det er derfor viktig å teste anonymiserte data i henhold til akseptabelt risikonivå og dokumentere dette. Datatilsynet har utarbeidet en veileder for anonymisering av personopplysninger. Veilederen inneholder anbefalinger for sikker anonymisering og ble publisert på Datatilsynets nettsider sommeren 2015²⁸.

3. Innhente konsesjon fra Datatilsynet

Juridisk må DiBK ha konsesjon for å kunne lagre og behandle personopplysninger, inkludert informasjon om adresse/eiendomsnummer som er tilknyttet registrerte skader, klager mv. Ved at DiBK søker Datatilsynet om konsesjon og de godtar denne, kan DiBK få forhåndsgodkjent sin innhenting og behandling av personopplysninger. For at Datatilsynet skal godkjenne en konsesjon vurderer de blant annet personopplysningenes art, mengden data og formålet med behandlingen. Et elektronisk skjema for konsesjon ligger på Datatilsynet sine nettsider, men skjemaet må underskrives før det sendes til Datatilsynet²⁹. KPMG antar at Datatilsynet vil kunne godkjenne en konsesjon da DiBK sine analyser vil ha et samfunnsnyttig formål.

5.2.2 Teknologiske hindre

Den høye graden av digitalisering og modernisering i bransjen representerer både muligheter og utfordringer knyttet til systematisert analyse og anvendelse av registrerte data. Hyppig utskifting av systemer krever at løsningen som henter data fra de ulike kildene må være svært fleksibel og håndtere endringer i datakilder og formater.

Det er også viktig å være oppmerksom på at når innholdet i datakildene endrer seg over tid, vil dette kunne gjøre utslag på indikatorene. Jo flere datakilder som danner grunnlag for indikatorene, jo mindre utslag vil endringer i én enkelt kilde gjøre.

5.2.3 Økonomiske hindre/forretningsmessige forhold

Dataeierne har i utgangspunktet varierende evne og motivasjon for å dele data med DiBK.

Fagmiljøet hos dataeierne må ha tilstrekkelig tid til involvering da de er "bestillere" og "sponsorere" for prosjektet, og IT/tekniske ressurser hos dataeierne vil måtte bruke betydelige ressurser på å sette opp uttrekk og overføring av data til DiBK. Motivasjon blir derfor meget viktig.

²⁸ ANONYMISERING AV PERSONOPPLYSNINGER, Veileder, 2005, Datatilsynet

²⁹ Søke konsesjon (<http://www.datatilsynet.no/personvern/Melding-og-konsesjon/soke-konsesjon>)

For å imøtekomme de økonomiske hindrene bør det iverksettes tiltak som forenkler arbeidet hos aktørene. Videre bør kartlegging av motivasjon og investeringsvilje (både hos brukere og dataeiere) gjennomføres i første fase av hovedprosjektet for å sikre forankring og at løsningen tilfredsstillende dataeierne krav. Benchmarking-funksjonalitet og muligheter for dataeierne til å gjøre "ad hoc"-analyser er foreslått for å gi dataeierne ekstra insentiver til å delta i prosjektet.

En full løsning som integrerer og tilrettelegger data fra alle de kartlagte kildene, vil medføre kostnader knyttet til både maskinvare, programvare, utvikling, drift og forvaltning. For å få dekket kostnader til et eventuelt hovedprosjekt og implementering, drift og vedlikehold av en systemløsning og verktøy, bør DiBK i eller før et hovedprosjekt gå i dialog med Difi og andre offentlige aktører for å høre om det er initiativer eller infrastruktur man kan koble seg på for å holde kostnadene nede. Det bør også vurderes løsninger der dataeierne kan bli med på å dekke kostnader mot at de får anvende løsningen til benchmarking mot andre aktører og analyse av egne data.

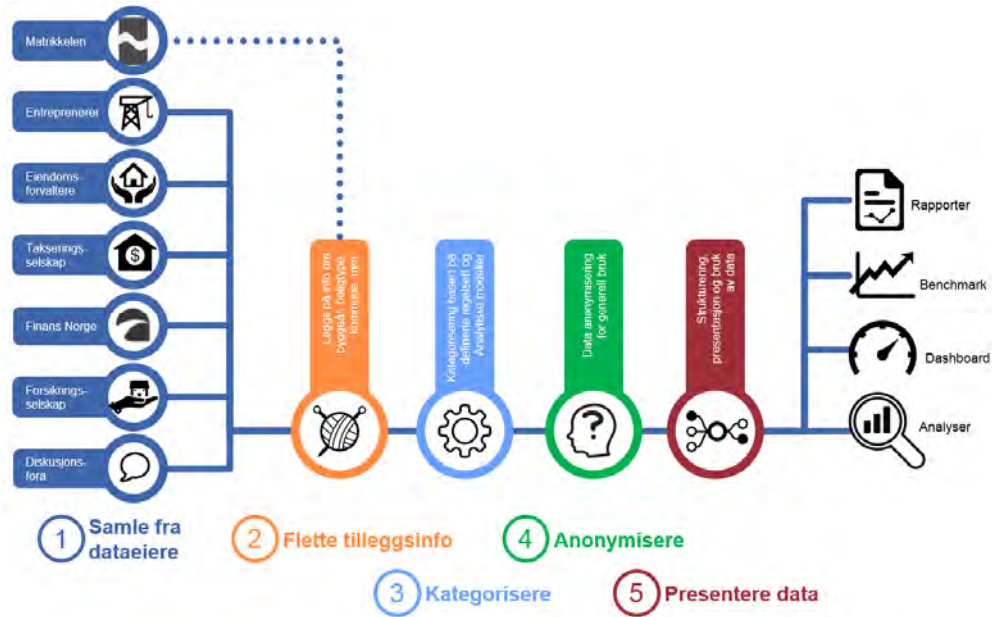
5.2.4 Datakvalitetsutfordringer

For å bøte på varierende datakvalitet og manglende bygningsinformasjon, foreslås det å koble på informasjon fra matrikkelen. Gitt at informasjon om adresse/bygningsnummer er korrekt, vil kobling med informasjon fra Matrikkelen gi bygningsinformasjon av høy kvalitet tilknyttet den enkelte sak/klage/skade/feil/mangel. Kompetanse og eierskap DiBK har begrenset kompetanse på områdene databehandling, datalagring og bruk av analyse- og rapporteringsverktøy, og det bør derfor sterkt vurderes å sette ut utvikling, drift og forvaltning av løsningen. Samarbeid med andre aktører i offentlig sektor som har initiativer innen Big Data og analyse bør vurderes.

Bransjen totalt sett er også relativt umoden på bruk av analyse-, statistikk- og rapporteringsverktøy. Dette fordrer at de tenkte brukerne av løsningen får tilstrekkelig opplæring. Samtidig representerer umodenheten en mulighet for en eventuell Big Data-løsning for byggkvalitet, da det er mange aktører som ikke har utviklet noe tilsvarende selv og som kan være interessert i å delta i investeringer innenfor området.

5.3 Krav til funksjonalitet i eventuell fremtidig løsning

For å imøtekomme utfordringer knyttet til *ulike typer systemer, liten grad av standardisering og varierende datakvalitet*, foreslår vi en løsning som i tillegg til å motta/hente data fra aktuelle dataeiere også fletter inn eiendomsinfo fra matrikkelen, legger på automatisert kategorisering av skade-/feiltyper og anonymiserer dataene før data tilrettelegges for analyser og dashboards. Disse stegene er kort forklart nedenfor.



Figur 19 - Overordnet løsnings-skisse for anbefalt løsning

1. Samle inn data

Dataeierne tilgjengeliggjør regelmessig oppdaterte data for DiBK via hensiktsmessig filoverføringsteknologi.

2. Flette inn tilleggsinfo

Basert på informasjon om adresse, bruks- og gårdsnummer og/eller bygningsnummer den bygningen som den enkelte skade/feil/mangel/klage gjelder, gjøres oppslag i Matrikkelen. Basert på data i Matrikkelen påføres dermed bygningsinformasjon som bygningstype og byggeår etter samme standard.

3. Kategorisere

Basert på innhold i felter med skriftlig innhold blir den enkelte skade/feil/mangel/klage tilordnet én eller flere kategorier som saken mest sannsynlig tilhører. Logikken i den automatiske kategoriseringen kan være av typen som er beskrevet i avsnittet om funn fra piloten.

4. Anonymisere

Informasjon om hvilken bygning en sak gjelder, enten det er adresse, bruks- og gårdsnummer og/eller bygningsnummer, blir fjernet for å gjøre analysegrunnlaget anonymt.

5. Presentere data

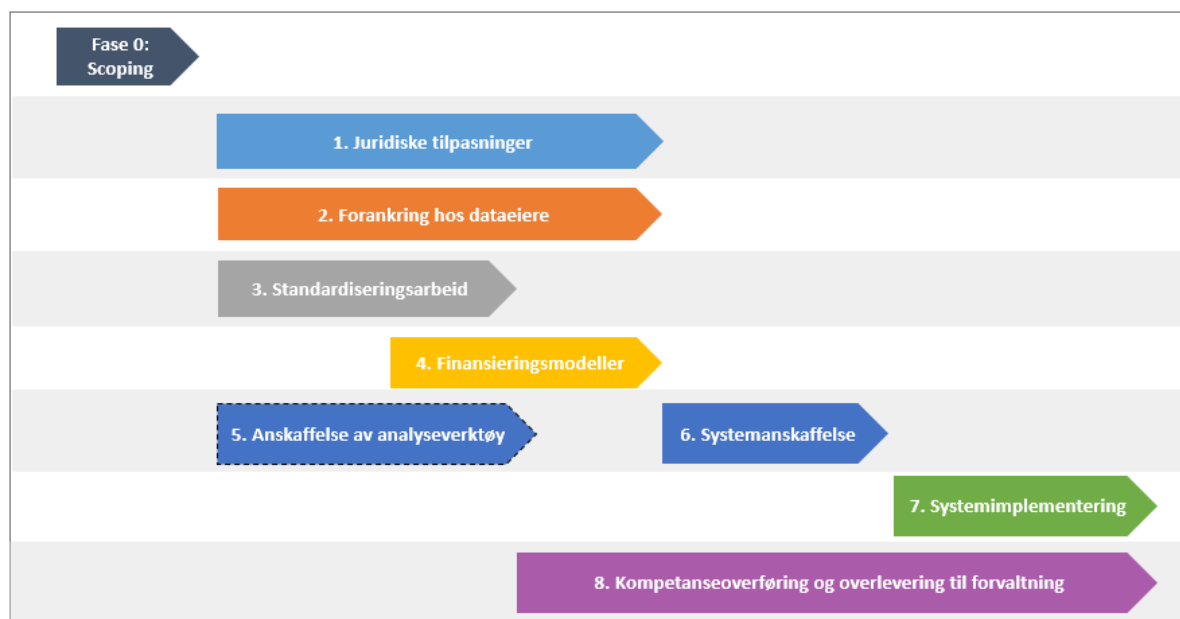
Data tilrettelegges i relevante analyse- og rapporteringsverktøy. I tillegg til anonymiserte data, som kan være tilgjengelig for alle brukergrupper, kan verktøyet inneholde ikke-anonymiserte data for bruk av de som eier dataene. Personvernet vil da ivaretas gjennom tilgangsstyring.

6 Forslag til hovedprosjekt

Vi vil her presentere vårt forslag til hovedprosjekt, som har til hensikt å sørge for at Direktoratet for Byggkvalitet når målbildet om en analyseplattform for byggkvalitet og byggskaadeomfang. Før selve hovedprosjektet settes i gang for fullt, anbefaler vi å kjøre en fase 0 hvor mandat for hovedprosjektet blir definert og det legges detaljerte planer for gjennomføringen av hovedprosjektet.

Selve hovedprosjektet kan bestå av arbeidsstrømmer som sikrer forankring, juridiske tilpasninger, standardisering og overlevering til systemeier, samt evaluering av finansieringsmodeller. I tillegg kan det være egne faser for henholdsvis anskaffelse og implementering av systemløsning/verktøy. Anskaffelsen av analyse- og visualiseringsverktøy kan gjennomføres sammen med anskaffelsen av resten av systemet, eller den kan gjennomføres som et eget delprosjekt som raskt gir verdi for DiBK.

Nedenfor er det laget en tentativ skisse til overordnet prosjektplan for videre arbeid. Et mer detaljert forslag til prosjektplan er vedlagt i *Vedlegg F Prosjektplan for foreslått hovedprosjekt*. Denne viser overordnet oversikt med foreslåtte arbeidsstrømmer, og deretter følger forslag til omfang, milepæler, leveranser og aktiviteter for hver arbeidsstrøm.



Figur 20 - Overordnet prosjektplan med arbeidsstrømmer

Fase 0, Scoping-fasen, har til hensikt å konkretiser og etablere prosjektmandat for hovedprosjektet. Når denne er ferdigstilt kan flere arbeidsstrømmer foregå parallelt. Arbeidsstrøm for *juridiske tilpasninger* som sikrer at nødvendige tiltak blir iverksatt for å møte juridiske krav (særlig knyttet til personopplysningsloven og informasjonssikkerhet samt mulig konsesjonssøknad). Det anbefales å kjøre egen

arbeidsstrøm for *forankring hos dataeierne* for å sikre motivasjon og deltakelse fra de aktuelle dataeierne, samt arbeidsstrøm for *standardiseringsarbeid* som skal sikre en utvikling mot stadig større konsistens i registreringer og datagrunnlag.

Videre bør det arbeides med *finansieringsmodeller*, for å sikre at kostnader for løsningen dekkes på en hensiktsmessig måte. I denne sammenheng er det også naturlig å se på driftsmodeller. Vi anbefaler å *anskaffe og implementere analyse- og visualiseringsverktøy* som kan benyttes til "ad hoc"-analyser og analyser på filer som er semi-manuelt overført. Denne anskaffelsen kan kjøres separat eller som del av neste arbeidsstrøm, som er knyttet til *anskaffelse av systemløsning/verktøy* for den totale dataflyten.

Til sist kommer *implementering av systemet for hovedløsning*, og parallelt med dette vil vi sørge for *kompetanseoverføring og overlevering* til forvaltning.

Vedlegg

Vedlegg A

Gjennomførte workshops og intervjuer

Det er gjennomført både workshops og intervjuer med ansatte i DiBK og representanter fra andre aktører i bransjen for å belyse både muligheter, faktisk situasjon og status på data. Tabellen nedenfor angir de workshops og intervjuer som har blitt gjennomført i forbindelse av dette oppdraget.

Type gjennomføring	Organisasjon	Type selskap/organisasjon
Idé- og oppstartsworkshop med interne aktører	DiBK, prosjektleder og representanter fra avdeling for bygg og byggeprosess som dekker fagområdene tekniske forskrifter (nybygg), eksisterende bygg og universell utforming.	DiBK
Idé- og oppstartsworkshop med interne og eksterne aktører	DiBK: Direktoratets leder (innleder) og representanter fra avdeling for bygg og byggeprosess. Eksterne aktører: - Trondheim kommune (Byggesakskontoret) - Vianova Systems TiPS FM - Norges Takseringsforbund - Finans Norge - Bygg21	Internt + Tilsynsmyndighet Programvareleverandør Rådgivnings-/takseringsfirma Bransjeorganisasjon
1:1-møter med aktører som sitter på data som kan gi indikatorer på byggkvalitet og oppfatning av regelverk	Skanska Bolig	Utbygger/entreprenør
	Sintef Byggforsk Bolig	Rådgivningsfirma
	Statsbygg (eiendomsavdelingen)	Eiendomsforvalter
	Boligtvistnemnda (telefon)	Tvistnemnd

Type gjennomføring	Organisasjon	Type selskap/organisasjon
	Finans Norge	Hovedorganisasjon for finansnæringen i Norge
	Undervisningsbygg	Utbygger og eiendomsforvalter
	Opak	Rådgivningsfirma
	Statens strålevern (telefon)	Tilsynsmyndighet
	Arkitektbedriftene	Bransje- og arbeidsgiverorganisasjon
	Tryg	Forsikringsselskap (ansvarsforsikring)
	Byggebolig	Eier av diskusjonsforum på nett

Vedlegg B

Detaljerte vurderinger av datakilder

Kategorier av dataeiere	Type informasjon og aktuelle indikatorer	Aktuelle indikatorer	Relevante kildesystemer	Vurdering av nytte	Vurdering av datakvalitet og tilgjengelighet	Data som er vurdert i forprosjektet
Entreprenører/utbyggere	Registrerte skader, feil og mangler – primært fra byggeprosess, overtakelse og 1-årsbefaring.	Totalt antall registrerte skader, feil og mangler Antall registrerte skader, feil og mangler i forhold til antall prosjekter og/eller antall m2 som er bygget Fordeling av skader, feil og mangler på type bygg, fordeling per type skade/feil/mangel, geografisk fordeling	CRM-systemer/sakshåndterings-systemer Underliggende database i løsninger for registrering og oppfølging av skader, feil og mangler (f.eks. Byggeweb fra Vianova) for aktører som har dette.	👉	👉	Uttrekk fra utbyggers/entreprenørs system for kundeforhold og sakshåndtering

Kategorier av dataeiere	Type informasjon og aktuelle indikatorer	Aktuelle indikatorer	Relevante kildesystemer	Vurdering av nytte	Vurdering av datakvalitet og tilgjengelighet	Data som er vurdert i forprosjektet
Eiendomsforvaltere	Registrerte avvik (skader, feil og mangler) i bruksfase	Totalt antall registrerte avvik Totalt antall registrerte avvik i forhold til antall prosjekter og/eller antall m2 som forvaltes Fordeling av antall og type avvik på ulike byggtyper, på geografi, på utbygger	FDV-systemer og/eller egne systemer for avvikshåndtering	👍	N/A	Ingen data fra eiendomsforvaltere er testet i forprosjektet.
Rådgivnings- og takseringsfirma	Vurdering av skader, feil og mangler i skadetakster og rapporter fra byggfaglig spesialrådgivning	Antall rådgivningsoppdrag. Merk at dette er en indikator på boligeieres/-forvalteres villighet til å klage heller enn et objektivt mål på byggkvalitet.	En del nøkkelinformasjon om rapportene er lagret i database. Fullstendige rapporter er tilgjengelig som tekstdokument.	👉	👉	Eksempler på rapporter fra byggfaglige rådgivningsoppdrag

Kategorier av dataeiere	Type informasjon og aktuelle indikatorer	Aktuelle indikatorer	Relevante kildesystemer	Vurdering av nytte	Vurdering av datakvalitet og tilgjengelighet	Data som er vurdert i forprosjektet
Tilsynsmyndigheter (Plan- og bygningsetaten)	Tilsynssaker, hvor feil og mangler kan være avdekket. Byggesaker	Antall tilsynssaker Antall saker som gjelder spesifikke feil/mangler.	Tilsynsarkiv Byggesaksarkiv	N/A	N/A	Ingen data fra plan- og bygningsmyndighetene er analysert.
Forsikringsbransjen	Skadeomfang Regelverkseffekter		Finans Norges vannskade- og brannskaderegister (VASK og BRASK) Forsikringsselskaperenes skadearkiver	👍	👎 Det er særlig mangelfull registrering av byggeår, hvilket begrenser troverdigheten til analyser på skader i nye bygg.	VASK BRASK

Kategorier av dataeiere	Type informasjon og aktuelle indikatorer	Aktuelle indikatorer	Relevante kildesystemer	Vurdering av nytte	Vurdering av datakvalitet og tilgjengelighet	Data som er vurdert i forprosjektet
Tvistnemder og domstoler	Oppsummering av klagesaker, samt domsavsigelser. Gir indikasjon på omfanget av saker som forbruker mener er feil og/eller mangler.	<p>Antall tvistesaker</p> <p>Antall domsavsigelser hvor boligkjøper har fått medhold.</p> <p>Utvikling i erstatningsbeløp fra domsavsigelser.</p>	<p>Arkiv på webside med oppsummering av alle behandlede saker i boligvistnemnda.</p> <p>Arkiv med alle domsavsigelser i arkivet til domstol.no</p>	<p>👇</p> <p>Antall saker på boligvistnemnda er noe lavt. Hver sak gjelder mange klagepunkter.</p>	<p>👍</p>	<p>Oppsummering av boligvistnemndas saker i årene 2010-2015 (drøyt 200 saker)</p> <p>Domsavsigelsesdata er ikke vurdert.</p>
Eiere av diskusjonsfora på nett	Innlegg i diskusjoner vedrørende bygg.	<p>Analyser av diskusjoner vedrørende tolkning av regelverk</p> <p>Analyser av meninger om regelverk og deres effekter</p>	Database med alle innlegg på forumet.	<p>👍</p>	N/A	Data ikke mottatt for analyse, til tross for at Byggebolig AS var veldig interessert.

Vedlegg C

Detaljer vedrørende test/"Proof of concept"

Område 1: Kan DiBK få innsikt i utvikling og omfang av skader, feil og mangler gjennom mer automatisert datainnsamling og analyse av eksisterende informasjon?

Innenfor dette området testes det både hvilken innsikt som kan fås fra strukturerte data (test 1-1) og i hvilken grad analyse og modellering på tekstlig informasjon kan være nyttig (test 1-2).

Test 1-1	Strukturerte data
Test skal besvare	Hvilken innsikt som kan gis ved visualisering av data som allerede er strukturert
Hypotese	Visualiseringsverktøy kan gi rask og nyttig innsikt i data om skader, feil og mangler som er registrert på strukturert format (for eksempel i en database)
Datagrunnlag	<ul style="list-style-type: none"> ▪ Test ble gjennomført på et uttrekk fra en utbyggers saksbehandlingssystem, som inneholder saker fra både befaringer og innkomne kundehenvendelser, ble visualisert i et analyse- og visualiseringsverktøy. Sakene gjaldt både skader, feil og mangler. Dataene var strukturerte (tabellform), men inneholdt også en kolonne med både faste kategorier og fritekstfelter. Sakene gjaldt et lite utvalg byggeprosjekter og inneholdt ikke datofelt. ▪ Visualiseringsverktøyet ble også testet på uttrekk fra Finans Norges databaser over skademeldinger, vannskader og brannskader, hhv. VASK og BRASK.
Gjennomføring av test	<p>Dataene ble visualisert på ulike måter:</p> <ol style="list-style-type: none"> 1. Gjennom tradisjonelle frekvensdiagrammer som viser forekomst av ulike feiltyper, ulike produkter, antall skader per byggtipe, mv 2. Gjennom ordskyer, som illustrerer hvilke tekststrenger som forekommer oftest i det valgte datagrunnlaget <ul style="list-style-type: none"> ▪ Verktøyet som ble brukt i denne testen er SAS Visual Analytics.
Funn	<ul style="list-style-type: none"> ▪ Visualisering av datagrunnlaget gir raskt oversikt over hva sakene gjelder, og man kan også se utvikling over tid dersom man har et mer komplett datagrunnlag fra en lengre tidsperiode.
Konklusjon	<ul style="list-style-type: none"> ▪ Visualiseringsverktøy gir meget nyttig innsikt i omfang av skader, feil og mangler for ulike typer bygg, basert på strukturerte data (både med kvantitativt innhold og tekstfelter). Automatiserte "kategoriseringsalgoritmer" kan benyttes for å angi sannsynlig kategori en skade/feil/mangel inneholder, basert på et fritekstfelt med beskrivelse av saken/kundens henvendelse.

Test 1-2	I hvilken grad analyse og modellering på tekstlig informasjon kan være nyttig
Test skal besvare	<p>Hvilken innsikt som kan gis ved visualisering og modellering av ustrukturerte data (gjennom ordskyer, "concept linking" og text mining).</p> <p>Visualisering av strukturerte data om skader, feil og mangler fra entreprenørs kunde- og klagehåndteringssystem. Dataene inneholdt både ordinære strukturerte data og ustrukturerte data (fritekstfelt).</p>
Hypotese	Text mining-verktøy kan benyttes til å utvikle modeller/regelsett som med rimelig høy grad av nøyaktighet automatisk kategoriserer en sak til en type skade/feil/mangel.
Datagrunnlag	Testen ble gjennomført på samme datagrunnlag som test 1-1.
Gjennomføring av test	<p>Hvilken innsikt kan DiBK (og øvrige aktører) få gjennom visualisering og analyse av ustrukturerte data, dvs. tekstdokumenter eller fritekstfelter i databaser.</p> <ul style="list-style-type: none"> ▪ Visualisering gjennom "ordskyer". ▪ Kan text mining-metoder benyttes til å kategorisere dokumenter og anslå f.eks. hvilken skadetype en sak representerer eller hvilke(n) bygningsdel(er) den gjelder? Testet på Skanskas saksbehandlingsdata og boligvistnemnda.
Tester	<p>Automatisert kategorisering av registrerte skader, feil og mangler basert på innhold i tekstfelt. Text Mining-metodikk testes for å få angitt sannsynlig skade/feil/mangel basert på strukturert og ustrukturert informasjon.</p> <p>Stegene for å utvikle og anvende "regelsettet" som klarer å klassifisere i de definerte skadetyper basert på fritekstfelt, er som følger:</p> <ol style="list-style-type: none"> a) Deler i trenings- og testdatasett b) Etablerer liste med ord som skal inkluderes i text miningen, basert på input fra DiBK samt ord og uttrykk fra regelverk. c) Verktøyet hjelper oss å filtrere bort uttrykk som ikke gir verdi d) Verktøyet hjelper oss med å lage et regelsett som indikerer hvilken kategori en tekststreng sannsynligvis tilhører basert på hvilke ord som forekommer i tekststrengen ("Text Rule Builder")

Test 1-2	I hvilken grad analyse og modellering på tekstlig informasjon kan være nyttig
Konklusjon	Analyse og modellering på tekstlig informasjon gir nytte; Ordskyer gir god innsikt i informasjonselementer som inneholder korte tekster. Text mining kan benyttes for å automatisk kategorisere dokumenter i mest sannsynlige type skade/feil/mangel.

Område 2: Kan DiBK få økt innsikt i effekt av regelverksendringer ved analyse av data i eksisterende datakilder?

Innenfor dette området gjennomføres det tester både relatert til regelverk om vannskadesikring (test 2-1) og brannsikring (test 2-2)

Test 2-1	Effekter av nytt krav til vannskadesikring (i praksis automatisk lekkasjestopper) på oppvaskmaskiner
Test skal besvare	Hvilke effekter har nytt krav til vannskadesikring (i praksis automatisk lekkasjestopper) på oppvaskmaskiner, hatt?
Hypotese	Ved å analysere utvikling i vannskader før og etter 2010 kan man få innsikt i effekter av nytt krav i 2010 til vannskadesikring (lekkasjestopper)
Datagrunnlag	Utdrag fra vannskaderegisteret ("VASK"), med skadedata. For testene er kun nye boliger (0 år i aktuelt skadeår) tatt med.
Gjennomføring av test	<ul style="list-style-type: none"> ■ Filtrer bort skader for boliger som er registrert med alder > 0 år ■ Visualiser skader med kilde "vanntilkoblet maskin" over tid i et histogram ■ Test om det har vært en nedgang etter 2010, og om en eventuell nedgang har vært signifikant.
Konklusjon	VASK kan benyttes til å gjøre analyser på utvikling i vannskader av ulike årsaker over tid, og analyser viser en signifikant nedgang i antall vannskader i nye bygg30 fra 2010. Merk imidlertid at mange skader er oppført uten alder på bygg, hvilket er en potensiell feilkilde.

³⁰ For en stor andel av registreringene er ikke byggets alder fylt ut. For bygg av typen "ny" eller av ukjent alder, er det ikke en signifikant nedgang. Det er derfor en viss usikkerhet knyttet til konklusjonen.

Test 2-2	Effekter av nytt krav om automatisk slokkeanlegg (i praksis sprinkleranlegg) i boligbygninger med krav om heis, dvs. boligbygg med mer enn to etasjer/boligblokker)
Test skal besvare	Hvilke effekter har krav om automatisk slokkeanlegg (i praksis sprinkleranlegg) i boligbygninger med krav om heis, dvs. boligbygg med mer enn to etasjer/boligblokker), hatt?
Hypotese	Nedgang i omfang av skader knyttet til branner i boligblokker, muligens med økning i vannskader knyttet til sprinkleranlegg.
Datagrunnlag	Skadedata fra brannskaderegisteret (BRASK)
Gjennomføring av test	<ul style="list-style-type: none"> ■ Forventet effekt: redusert antall brannskader (og skadebeløp) i boligbygninger). (Men også høyere antall vannskader pga tilløp til brann eller feilaktig aktivering av sprinkleranlegget? Kan disse sakene identifiseres?) ■ Søkeord (om nødvendig): brann, sprinkleranlegg, (automatisk) slokkeanlegg ■ Utfordringer: det er få branner hvert år, slik at det er vanskelig å oppdage endringer. Boligblokker er heller ikke mulig å skille ut direkte, da slik informasjon om bygningstype ikke er tilgjengelig i BRASK.
Konklusjon	<p>Datakvaliteten i BRASK er ikke like god som for VASK, og det er ikke mulig å skille skader som gjelder <i>boligblokker</i>.</p> <p>I VASK er omfanget av skader koblet til sprinkleranlegg minimalt, slik at det ikke er mulig å finne signifikant endring.</p>

Område 3: Kan DiBK få økt innsikt i den generelle kjennskapen til, og tolkningen av, regelverket gjennom analyse av informasjon på internett?

Innenfor dette området gjennomføres det tester både på data på Twitter (test 3-1) og data i diskusjonsfora på nett (test 3-2).

Test 3-1	Kan DiBK få økt innsikt i hvor godt kjent regelverket er, og hvor lett det er å forstå, gjennom analyse av innlegg i sosiale medier?
Test skal besvare	Kan DiBK få økt innsikt i hvor godt kjent regelverket er, og hvor lett det er å forstå, gjennom analyse av innlegg i sosiale medier?
Hypotese	Analyse av Twitter-meldinger med emneknagger #TEK10 og #TEK17 kan antyde hva befolkningen er opptatt av vedrørende disse regelverkene.
Datagrunnlag	Feeds fra Twitter, samlet inn via verktøyene SAS Visual Analytics og IBM Watson Analytics.

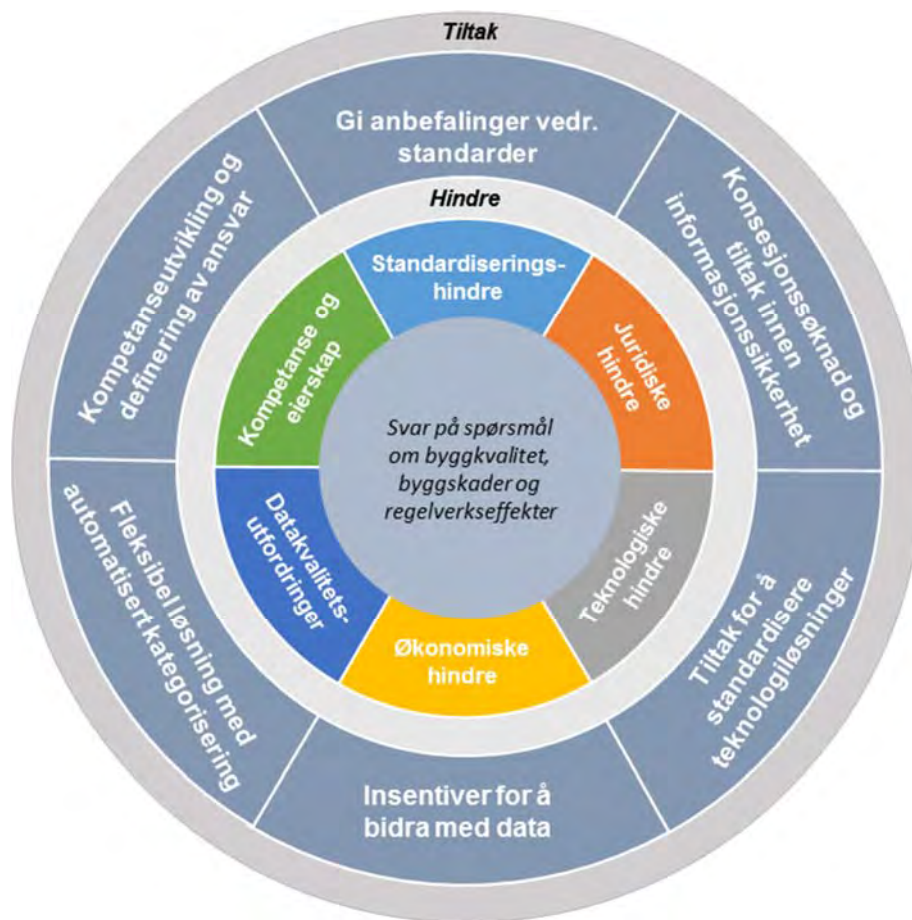
Test 3-1	Kan DiBK få økt innsikt i hvor godt kjent regelverket er, og hvor lett det er å forstå, gjennom analyse av innlegg i sosiale medier?
Gjennomføring av test	<ul style="list-style-type: none"> ■ Vi søker på emneordene #TEK10 og #TEK17 i verktøyene. ■ Twitter inneholder for få tweets med emneknagg #tek10. Det samme gjelder #byggutenaasoke ■ TEK10 har ca. 200 norske Tweets siden første kom i 2009-#byggutenåsøke har få treff (inkl. 2-3 fra DiBK selv). ■ #tek10 blir også benyttet som emneknagg for et japansk fenomen, som er omhandlet i svært mange tweets
Konklusjon	Per i dag blir emneknagger som TEK10 og TEK17 for lite omtalt til at det gir verdi å gjøre analyser av innlegg i Twitter basert på emneknagger.

Test 3-2	Kan analyse av innlegg i utvalgte diskusjonsgrunnlag gi innsikt i hvilke (deler av) regelverk som oppfattes som uklare?
Test skal besvare	Kan analyse av innlegg i utvalgte diskusjonsgrunnlag gi innsikt i hvilke (deler av) regelverk som oppfattes som uklare?
Hypotese	<p>"Crawling" og analyser av innlegg fra byggebolig.no og evt. kvinneguiden.no kan gi innsikt i hva befolkningen generelt omtaler #TEK10 sammen med.</p> <ul style="list-style-type: none"> ■ Se hva folk snakker om sammen med ordet "tek10" eller "garasje" for eksempel. ■ utfordringer: man får ikke hentet ned kun ett og ett innlegg.
Datagrunnlag	Nettsiden "byggebolig.no" ble benyttet som datagrunnlag for testen.
Gjennomføring av test	Legge inn søkeord i verktøyene, og vurdere resultatene.
Konklusjon	Det er krevende å hente ut innleggene på en "ren" nok form ved bruk av en ordinær web crawler. Det kreves spesialiserte verktøy for å hente ut informasjon fra diskusjonsforaene på en tilstrekkelig strukturert form. Alternativt kan slike analyser gjennomføres ved å få tilgang på databasen som forvalteren av diskusjonsforumet sitter på.

Vedlegg D

Hindre og tiltak for å overkomme identifiserte hindre

Identifiserte hindre og aktuelle tiltak er oppsummert i figuren nedenfor, og detaljeres i påfølgende tabell.



Figur 21 - Hindre og tiltak

Område/hinder	Observasjon	Anbefaling/tiltaksområder	Aktiviteter
Standardiseringshindre	Aktørene benytter i liten grad standarder som for eksempel bygningsdelstabellen eller standardiserte skade- og feiltyper.	Utvikle og implementere en "motor" for å "oversette" til standardkategorier ved import av data til løsningen	Utvikle algoritme/motor for automatisert kategorisering.
		Styrke standardiseringsarbeidet	Anbefal aktørene å benytte gitte standarder, som for eksempel bygningsdelstabellen fra byggforskserien fra Sintef Byggforsk.
	Aktørene synes i liten grad å ha standardiserte systemer/infrastruktur.	Arbeide for å standardisere/oppmuntre til å bruke felles løsninger og drifts- og forvaltningspartner på IKT for statlige forvaltningsselskaper som Boligbygg, Statsbygg, Undervisningsbygg og Forsvarsbygg?	Gå i dialog med store og/eller viktige verktøy-/teknologileverandører (f.eks. Vianova) og oppfordre til samarbeid rundt standardisering.
Juridiske hindre	Mange av datakildene inneholder personopplysninger, og kobling av for eksempel adresse og en klage kan regnes som sensitiv personopplysning. Da adresse eller gårds- og bruksnummer er nødvendig for å kunne koble på informasjon om byggtipe, byggeår etc. fra Matrikkelen, er DiBK avhengig av konsesjon for å kunne innhente og behandle disse dataene.	Øke/knytte til seg juridisk kompetanse vedr. relevant lovgivning (POL og personopplysningsforskriften) og informasjonssikkerhet	
		Sikre akseptabelt nivå på informasjonssikkerhet	
	Konsesjon krever at organisasjonen har kontroll på informasjonssikkerhet	Konsesjonssøknad	Etablere databehandleravtaler Sikre at dataeiere har innhentet samtykke som dekker denne bruken
Teknologiske hindre	Høy grad av digitalisering og modernisering i bransjen, særlig blant store og middels store selskaper. Høy endringstakt på systemer skaper utfordringer både med tanke på integrasjon mot stadig nye kilder, samt at innholdet i grunnlaget for indikatorer og analyser endres over tid.	Still krav om at løsningen må være en fleksibel, med lave kostnader og lav kompleksitet ved å koble på nye systemer og håndtere endringer i datakilder og formater. Ved vurdering av drifts-/forvaltningsmodell må det stilles krav til at forvalter må ha tilstrekkelig kapasitet	

Område/hinder	Observasjon	Anbefaling/tiltaksområder	Aktiviteter
		til å videreutvikle/justere løsningen (dataintegrasjonsdelen)	
		Løsningen må ha mulighet for "regelsett"/automatisert kategorisering som tar input på ulike formater og tilpasser til samme struktur.	
Økonomiske hindre / forretningsmessige forhold	Dataeierne har i utgangspunktet varierende evne og motivasjon til å dele data med DiBK. Fagmiljøet må ha tilstrekkelig tid til involvering da de er "bestillere" og "sponsorer" for prosjektet, og IT/tekniske ressurser vil måtte bruke betydelige ressurser på å sette opp uttrekk og overføring av data til DiBK. Motivasjon blir derfor meget viktig.	Tiltak som forenkler arbeidet hos aktørene, for eksempel at dataoverføringsløsning som velges skal være så enkel som mulig å "koble seg på"/bruke for dataeierne.	Kartlegge motivasjon (gjøres i fase 0), kartlegge investeringsvilje (både hos brukere og dataeiere), identifisere aktuelle insentiver, etablere aktuelle insentiver
		Tiltak for å gi aktørene insentiver for å bidra med data, f.eks. tilgang til analyseløsning for benchmarking og analyse av egne data. Forretningsmodeller som sørger for at de som har særlig verdifulle data å dele, kan få betaling for å dele sine data, bør vurderes.	
		Tiltak for å holde kostnadene nede: etablere samarbeid med aktører som kan ha nytte av å drifte/forvalte løsningen. Vurdere ulike finansieringsmodeller (spleiselag? Søke spesifikke midler? Koble seg på initiativ hos andre direktorater/etater?)	Kartlegge aktører som kan drifte/forvalte. SSB? Infoland? Sintef Byggforsk?
Datakvalitetsutfordringer	Kvaliteten i informasjon om type skade og bygningsinformasjon varierer.	Koble på informasjon fra matrikkelen for å få god datakvalitet på bygningsinformasjonen tilknyttet den enkelte skade/feil/mangel.	Gå i dialog med Finans Norge for å få adresse eller gårds- og bruksnummer på data i VASK og BRASK, for å muliggjøre oppslag mot Matrikkelen.
	Kvaliteten kan antas å variere over tid hos hver og én aktør på grunn av endringer i systemer og krav til registreringer.	Løsningen bør fokusere på de data som er kvalitetssikret og tilgjengelig i digital form. Inkludering av skannede og håndskrevne	Informasjon om hvilke større endringer som har funnet sted i systemer og registreringsrutiner, bør være tilgjengelig for brukerne av løsningen.

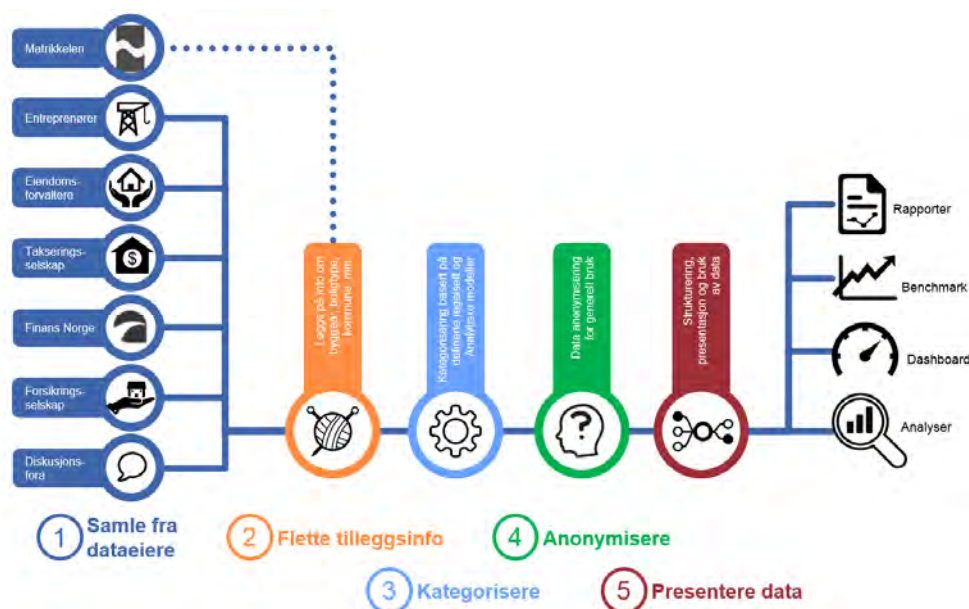
Område/hinder	Observasjon	Anbefaling/tiltaksområder	Aktiviteter
		dokumenter kan eventuelt gjennomføres som et videreutviklingsprosjekt etter at løsningen er etablert.	
		Datakvalitetsvurderinger gjennomføres for alle datakilder som tenkes inkludert i løsningen.	Vurder datakvalitet med tanke på nøyaktighet (fritt for feil), fullstendighet (all relevant info er med) og konsistens (likt over tid)
Kompetanse og eierskap	DiBK har liten kompetanse på området databehandling, datalagring og bruk av analyse- og rapporteringsverktøy.	Utvikling, drift og forvaltning bør vurderes satt ut.	Kartlegg hvilke drifts- og forvaltningspartnere som benyttes i søsterdirektorater og relaterte offentlige etater.
	Bransjen er relativt umoden på bruk av analyse- og rapporteringsverktøy		

Vedlegg E

Løsningsskisse for målbildet

Funksjonell beskrivelse av løsningsskisse

Figuren nedenfor beskriver foreslått overordnet løsningsskisse for en fremtidig løsning som dekker DiBK og andre aktørers behov og mulig bruk. Videre beskrives de ulike delene av den tenkte løsningen.



Figur 22 - Funksjonell løsningsskisse

1. Data, med info om bygningsnr/adresse/gårds- og bruksnummer hentes fra dataeierne.

Aktuelle dataeiere/datatyper, er følgende:

- Forvaltere
 - ❖ Avvik (Skader, feil og mangler som er innmeldt i drifts- og forvaltningsfasen)
 - ❖ På sikt: sensordata som angir luftfuktighet, temperatur, røykutvikling mv.
 - ❖ Særlig for Statsbygg: database med status på universell utforming på alle offentlige bygg i Norge (for eksempel dørbredder)
- Takseringsselskap
 - ❖ Skadetakster

- Finans Norge
 - ❖ VASK og BRASK
 - ❖ Sprinkleranlegg-database
- Forsikringsselskap
 - ❖ Skadedetaljer
- Drifere av diskusjonsfora
 - ❖ Innlegg i diskusjonsfora
- Ikke vurdert i forprosjektet, men bør være aktuelt
 - ❖ Data fra selskapsregistre: selskapsinformasjon, inkl. nøkkelpersoner, etableringsår, geografi etc. Når koblet til boligprosjekter kan man kunne se mønstre i hvilke typer selskaper og evt. nettverk av nøkkelpersoner som går igjen i prosjekter med mange eller spesielle typer skader, feil og mangler.
 - ❖ På sikt vil data fra sensorer i bygg kunne være aktuelt.

2. Oppslag i eiendomsregister for å legge på info om byggeår, boligtype, kommune/bydel, mv.

Dette oppslaget er nødvendig da mange av de kartlagte datakildene ikke har pålitelig informasjon om byggeår, boligtype, område, mv, og/eller de har ulike måter å kategorisere dette på. Ved å gjøre oppslag mot matrikkelen, sikrer vi at data fra alle kilder blir tildelt de samme kategoritypene.

3. Skader, feil og mangler blir automatisk tilordnet standardkategorier for type skade/feil/mangel, bygningsdel, mv, basert på definerte regelsett og analytiske modeller som tolker tekstinformasjon fra fritekstfelter og dokumenter. Dette gjøres for å etablere felles kategorier som kan benyttes til analyse- og statistikkformål.

4. Anonymisering av data

Dataene anonymiseres før de blir tilgjengeliggjort for generell bruk. Visse data kan også beholdes med adresse og/eller andre personopplysninger på, så lenge det kun er aktørene selv som har tilgang på dataene.

5. Presentasjon og bruk av dataene

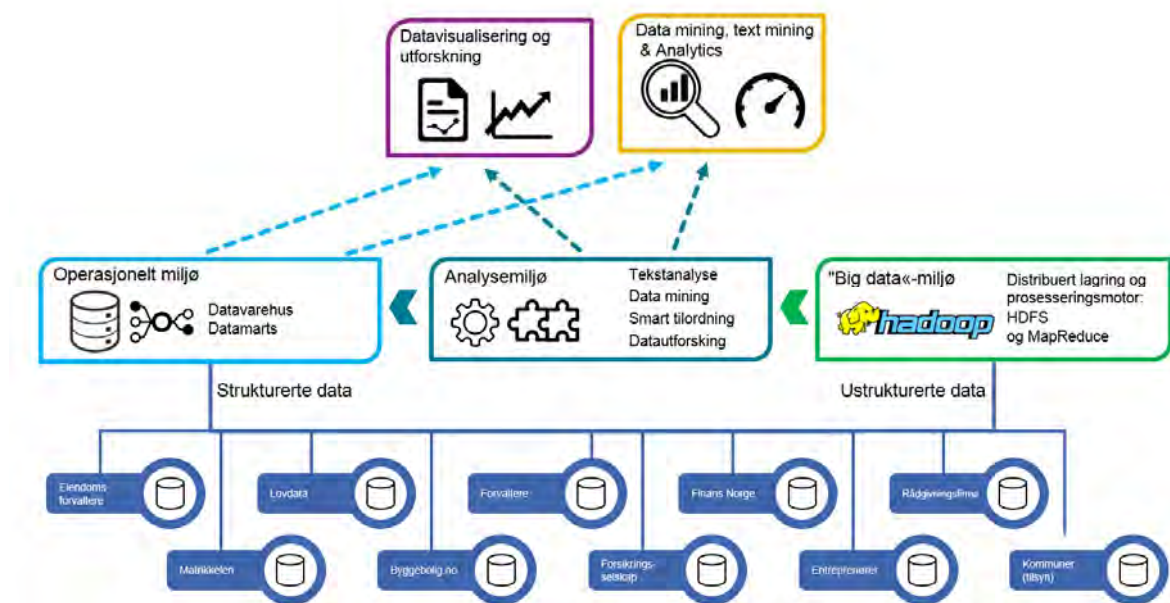
- ❖ Indikatorer presenteres i et "dashboard". Indikatorer/analyser/trender fra ulike kilder kan vektas til en "total indikator" og/eller separate indikatorer presenteres separat, men kan samlet si

noe om utvikling. F.eks. antall saker behandlet hos Boligtvistnemnda, antall saker meldt til Forbrukerrådet, antall vannskader i VASK, antallet skaderapporter, andel avvik i ulike bygningsdeler, mv.

- ❖ Grensesnittet/verktøyet kan benyttes av aktørene for å benchmarke seg mot andre aktører.
- ❖ Aktørene kan analysere sine egne data på skader, feil og mangler.
- ❖ DiBKs ansatte kan gjøre grundigere/mer omfattende analyser av utvalgte data i forbindelse med prosjekter (for eksempel at VASK og BRASK gjøres tilgjengelig).
- ❖ DiBKs ansatte (og andre aktører?) kan på "ad hoc"-basis laste opp data fra andre kilder for å gjøre enkeltstående analyser.

Arkitekturskisse/innspill til teknisk løsning

Den foreslåtte løsningen krever en arkitektur som støtter både tradisjonell lagring i datamarter og lagring av ustrukturerte data i et Big Data-miljø, for eksempel med såkalt Hadoop. Løsningen krever også verktøy og miljø for utvikling og produksjonssetting/scoring med aktuelle statistiske modeller og regelsett. Applikasjonslaget støtter datavisualisering, utforskende analyse, data mining og text mining.



Figur 23 - Arkitekturskisse

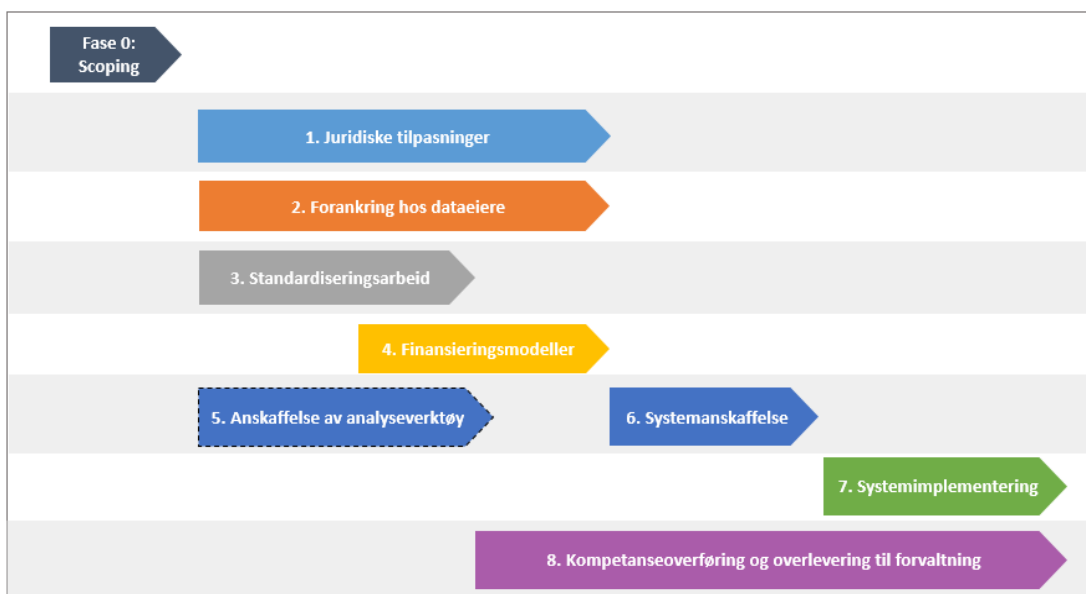
Vedlegg F

Prosjektplan for foreslått hovedprosjekt

Hovedprosjekt har til hensikt å sørge for at Direktoratet for Byggkvalitet når målbildet om en analyseplattform for byggkvalitet og byggskaadeomfang. Før selve hovedprosjektet settes i gang for fullt anbefaler vi å kjøre en fase 0 hvor mandat for hovedprosjektet blir definert og det legges detaljerte planer for gjennomføringen av hovedprosjektet.

Selve hovedprosjektet kan bestå av arbeidsstrømmer som sikrer forankring, juridiske tilpasninger, standardisering og overlevering til systemeier, samt evaluering av finansieringsmodeller. I tillegg kan det være egne faser for henholdsvis anskaffelse og implementering av systemløsning/verktøy. Anskaffelsen av analyse- og visualiseringsverktøy kan gjennomføres sammen med anskaffelsen av resten av systemet, eller den kan gjennomføres som et eget delprosjekt som raskt gir verdi for DiBK.

Nedenfor finnes en tentativ skisse til overordnet prosjektplan, samt forslag til omfang, milepæler, leveranser og aktiviteter for hver arbeidsstrøm.



Figur 24 - Overordnet prosjektplan med arbeidsstrømmer

Fase 0: Scoping av hovedprosjekt	
Formål	Konkretisere og etablere prosjektmandat for hovedprosjekt
Milepæler	Mandat utarbeidet og godkjent
Leveranser	Prosjektmandat for hovedprosjektet, inklusiv detaljert prosjektplan og kostnadsestimater
Foreslåtte aktiviteter³¹	<ul style="list-style-type: none"> ■ Prioritering av bruksområder og prioritering og definering av indikatorer ■ Identifisering av aktuelle organisasjoner innenfor hver gruppe av dataeiere ■ Kartlegging av dataeieres data og teknologi (inkl. status datakvalitet) samt motivasjon/ønsker for løsning ■ Vurdere finansierings- og driftsmodeller, inklusive vurdering av alternative forvaltere ■ Konkretisering og prioritering av delprosjekter i hovedprosjekt ■ Utarbeide prosjektplan for hovedprosjektet ■ Detaljere kostnadsestimater ■ Utarbeide prosjektmandat ■ Samle krav som identifiseres underveis, for bruk i kravspesifikasjon for anskaffelse av løsning
Arbeidsstrøm 1: Juridiske tilpasninger	
Formål	Sikre at nødvendige tiltak blir iverksatt for å møte juridiske krav, særlig knyttet til personopplysningsloven og informasjonssikkerhet
Milepæler	<ul style="list-style-type: none"> ■ Konesjon innvilget ■ Databehandleravtaler er etablert
Leveranser	<ul style="list-style-type: none"> ■ Konesjonssøknad ■ Liste med systemkrav knyttet til informasjonssikkerhet
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere arbeidsgruppe for arbeidsstrømmen ■ "Gap-analyse" for krav til konesjonssøknad ■ Igangsetting og oppfølging av tiltak for å tilfredsstille krav ■ Definere krav knyttet til informasjonssikkerhet for systemløsning
Arbeidsstrøm 2: Forankring hos dataeiere	
Formål	Sikre motivasjon og deltakelse fra aktuelle dataeiere
Milepæler	<p>Aktuelle dataeiere er kategorisert og prioritert</p> <p>Prioriterte dataeiere har forpliktet seg til å delta i prosjektet</p>

³¹ Listene over foreslåtte aktiviteter er ment som innspill og er ikke uttømmende

Leveranser	-
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere arbeidsgruppe og prosjektorganisasjon for arbeidsstrømmen ■ Kategorisere dataeiere etter a) hvor viktig dataeieren/datakilden er for prosjektet og b) hvor viktig prosjektet er for dataeieren ■ Definere tiltak for å få skape insentiver og motivasjon hos de prioriterte dataeierne ■ Iverksette definerte tiltak for å få dataeierne til å bidra
Arbeidsstrøm 3: Standardiseringsarbeid	
Formål	Sikre en utvikling med stadig større konsistens i registreringer/datagrunnlag
Milepæler	<ul style="list-style-type: none"> ■ Konkretiseres i scoping-fase
Leveranser	<ul style="list-style-type: none"> ■ Konkretiseres i scoping-fase
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere prosjektorganisasjon for arbeidsstrømmen ■ Iverksette initiativ sammen med systemleverandører for å sikre at disse benytter samme standarder ■ Oppfordre bransjen til bruk av standarder, for eksempel felles begrepsapparat i tråd med anbefalinger i "Byggforskserien" fra Sintef Byggforsk
Arbeidsstrøm 4: Finansieringsmodeller for systemløsning	
Formål	Sikre at kostnader for løsningen dekkes på en hensiktsmessig måte
Milepæler	Finansieringsmodell er valgt
Leveranser	<ul style="list-style-type: none"> ■ Forslag til finansieringsmodell
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere prosjektorganisasjon for arbeidsstrømmen ■ Kartlegge betalings- og investeringsvillighet blant aktuelle brukere ■ Evaluere ulike finansieringsmodeller ■ Anbefale finansieringsmodell
Arbeidsstrøm 5: Anskaffelse av analyseverktøy	
Formål	Anskaffe og implementere analyse- og visualiseringsverktøy som kan benyttes til "ad hoc"-analyser og analyser på filer som er semi-manuelt overført. Denne anskaffelsen kan kjøres separat eller som del av arbeidsstrøm 6 Systemanskaffelse.
Milepæler	<ul style="list-style-type: none"> ■ Anskaffelsesprosedyren besluttet ■ Anskaffelse utlyst ■ Leverandør og verktøy valgt ■ Verktøyet er implementert
Leveranser	<ul style="list-style-type: none"> ■ Kravspesifikasjon

	<ul style="list-style-type: none"> ▪ Kvalifikasjons- og konkurransegrunnlag
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ▪ Etablere prosjektorganisasjon for arbeidsstrømmen ▪ Vurdere og beslutte anskaffelsesprosedyren ▪ Spesifisere krav/tildelingskriterier for verktøy (funksjonelle, tekniske, sikkerhetsmessige, mv) og til leverandør ▪ Gjennomføre anskaffelsesprosessen og vurdere leverandører. (Vurder å be leverandør demonstrere verktøyets funksjonalitet på testdata fra DiBK, for eksempel VASK-data og/eller data fra entreprenør eller eiendomsforvalter) ▪ Implementere verktøy
Arbeidsstrøm 6: Systemanskaffelse	
Formål	Anskaffe teknisk løsning/verktøy for å sammenstille og lagre data fra dataleverandører, koble på informasjon fra Matrikkelen, automatisk kategorisering, datakvalitetskontroller og tilgjengeliggjøring av data for analyse- og visualiseringsverktøy. Inkluderer analyse- og visualiseringsverktøy dersom dette ikke anskaffes separat i tidligere fase/forprosjekt.
Milepæler	<ul style="list-style-type: none"> ▪ Anskaffelsesprosedyren besluttet ▪ Anskaffelse utlyst ▪ Leverandør og verktøy valgt ▪ Verktøyet er implementert
Leveranser	<ul style="list-style-type: none"> ▪ Arkitekturskisse ▪ Kravspesifikasjon ▪ Kvalifikasjons- og konkurransegrunnlag
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ▪ Etablere prosjektorganisasjon for delprosjektet ▪ Vurdere "in-house" vs. skybasert løsning ▪ Vurdere alternative datautvekslingsteknologier ▪ Utarbeide arkitekturskisse ▪ Vurdere og beslutte anskaffelsesprosedyren ▪ Spesifisere krav/tildelingskriterier for verktøy (funksjonelle, tekniske, sikkerhetsmessige mv) og til leverandør ▪ Gjennomføre anskaffelsesprosessen og vurdere leverandører/løsninger ▪ Implementere løsning
Arbeidsstrøm 7: Implementering av system for hovedløsning	
Formål	Implementere teknisk løsning/verktøy for å sammenstille og lagre data fra dataleverandører, koble på informasjon fra Matrikkelen, automatisk kategorisering, datakvalitetskontroller og tilgjengeliggjøring av data for analyse- og visualiseringsverktøy
Milepæler	<ul style="list-style-type: none"> ▪ Verktøy er implementert ▪ Infrastruktur mot dataeiere er satt opp ▪ Brukerne har tilgang til løsningen
Leveranser	<ul style="list-style-type: none"> ▪ Spesifisert databehov

Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere prosjektorganisasjon for arbeidsstrømmen ■ Spesifisere databehov ■ Implementere løsning ■ Implementere dataflyt
Arbeidsstrøm 8: Kompetanseoverføring og overlevering til forvaltning	
Formål	Sikre at DiBK er klare, både teknisk, organisasjons- og kompetansemessig, for å eie og bruke løsningen.
Milepæler	<ul style="list-style-type: none"> ■ Kompetansegap er dokumentert ■ Prioriterte tiltak er gjennomført ■ Roller og ansvar er dokumentert og fordelt
Leveranser	<ul style="list-style-type: none"> ■ Tiltaksplan for kompetanseoverføring ■ Overlevering av system
Foreslåtte aktiviteter	<ul style="list-style-type: none"> ■ Etablere prosjektorganisasjon for arbeidsstrømmen ■ Opplæring i verktøy og datakilder

Kontakt oss

Bjørn Lillebekk

Director

T: +47 40 63 96 52

E: bjorn.lillebekk@kpmg.no

Cathrine Pihl Lyngstad

Senior Manager

T: +47 40 63 93 19

E: cathrine.pihl.lyngstad@kpmg.no

© 2015, KPMG AS, a Norwegian limited liability company and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative ("KPMG International"), a Swiss entity. All rights reserved.

This proposal is made by KPMG AS, a member firm of the KPMG network of independent firms affiliated with KPMG International, a Swiss cooperative, and is in all respects subject to the negotiation, agreement, and signing of a specific engagement letter or contract. KPMG International provides no client services. No member firm has any authority to obligate or bind KPMG International or any other member firm vis-à-vis third parties, nor does KPMG International have any such authority to obligate or bind any member firm.

The KPMG name, logo and "cutting through complexity" are registered trademarks or trademarks of KPMG International Cooperative ("KPMG International").