

Christoffer Aakervik Tangen & Therese Thorvaldsen

Masteroppgave

Kan bruk av master data hub bidra til økt datakvalitet?

Will the usage of master data hub improve data quality?

KDBA950 - Master i Digital Ledelse og Business Analytics

2022

Sammendrag

Formålet med denne studien har vært å utrede om bruk av master data hub kan bidra til økt datakvalitet. Studien består av en litteraturstudie og en casestudie. Vi har stegvis sett på teori som understøtter hvor viktig master data er som en av de viktigste eiendelene til virksomheter og hvordan dette skal forvaltes. For å samle inn empiri har vi i casestudien vår utført semistrukturerte intervju med målgrupper for kunde og leverandør av master data hub.

Litteraturstudien vår viser hvor viktig det er med god kvalitet i master data, og at dette henger sammen med god master data management. Det er et gap i tidligere forskning i sammenhengen mellom spesifikke verktøy innen master data management, og verktøyets direkte påvirkning på datakvalitet.

Denne forskningen omhandler større og komplekse virksomheter, da vi antok at slike virksomheter har et forhold til sine master data og master data management. Utvalget i studien vår inneholdt ikke kunder eller leverandører som bruker eller leverer master data management der *artificial intelligence* og *machine learning* er innebygd. Vi har studert hvilke faktorer som fører til implementering av master data hub, teknologien som ligger bak en master data hub og om datakvaliteten øker ved bruk av denne typen verktøy.

Vår empiri viser at fravær av klare og tydelige retningslinjer tilknyttet eierrollen for master data gjør det utfordrende for virksomheter å garantere og opprettholde god datakvalitet. Vår forskning viser at master data management må tilpasses den enkelte virksomheten. Empirien vår viser at en master data hub kan bidra til økt datakvalitet innenfor enkelte dimensjoner. Samtidig gjør den hverdagen til brukere mer kompleks. Videre skapes det avhengigheter i master data objekter om virksomheten ikke er nøye med hvilke master data de legger i samme hub. Dette er uavhengig av om en virksomhet er moden eller ikke innen master data management.

Nøkkelord: Master Data, Master Data Management, Master Data Hub, Datakvalitet, Data Quality Management

Abstract

The aim of this empirical study is to analyze through research to which extent, if any, the usage of a master data hub will improve data quality.

This thesis consists of an empirical study in which one has analyzed data that supports our theory and substantiates the importance of master data as a premium asset and how such assets are best managed. We will also present a case study whereas one has performed semi structured interviews targeting customers and master data hub providers.

The empirical study presented in this thesis will depict the importance of a certain level of quality in regard to master data and given management. Previous research showed some discrepancies regarding the correlation between specific tools within the usage of master data management and more specifically the tools' direct impact on data quality.

Presented research is centered around corporations of a certain magnitude, under the assumption that organizations as such are aware and conscious towards their master data and master data management. The selection is not centered around customer perspective nor providers of master data management with built-in AI and ML. This study will focus on which factors lead to a successful implementation of the master data hub, the underlying technology, and if this consequentially will increase the data quality in using such services.

Through empirical studies, one has found that a major challenge derives from the lack of strict guidelines, rules, and regulations from the providers. This lack of structure makes it difficult to ensure a certain level of quality.

Research shows that master data management has an unfulfilled potential if not customized by the provider for the specific client. It also proved that the master data hub did increase the

data quality in certain dimensions, and also complicates the customers' operations. One saw that it created additions within the hub if not aware and careful regarding which master data one introduces in a given hub without any correlation to a level of maturity within master data management

Keyword: Master Data, Master Data Management, Master Data Hub, Data Quality, Data Quality Management

Forord

Denne masteroppgaven er skrevet som en avslutning på vår treårige masterutdanning i økonomi og ledelse ved Handelshøyskolen Innlandet. Oppgaven er skrevet i vårsemesteret 2022 og er en del av hovedprofilen i digital ledelse og business analytics. Masteroppgaven har et omfang på 30 studiepoeng.

Fagene Digitalisering, endring og ledelse, Data management strategi og Digital strategi og beslutningsstøtte inspirerte oss til å forske videre innenfor aktuelt området. Spesielt har forelesningene til Bo Hjort Christensen vært svært interessante.

Våre forkunnskaper innenfor emnet har vært begrenset og vi har ingen praktisk erfaring innen området. Det har derfor vært tidkrevende og tidvis overveldende å tilegne oss nok kunnskap til å evne å diskutere rundt våre problemstillinger. Vi har brukt mye ressurser på å diskutere og sparre med fagfolk, noe som har vært utrolig lærerikt og gøy.

Vi må rette en stor takk til alle som har hjulpet oss gjennom prosessen; veileder, informanter, foreldre og venner.

Trondheim / Oslo, mai 2022

Christoffer Aakervik Tangen

Therese Thorvaldsen

Innholdsfortegnelse

Sammendrag	2
Abstract	3
Forord	5
Figuroversikt	8
1. Introduksjon	9
1.1 Nøkkelbegrep	10
1.2 Bakgrunn	11
1.3 Forskningsspørsmål	12
1.4 Posisjonering	14
1.5 Oversikt over oppgavens oppbygging	15
2. Teoretisk grunnlag	16
2.1 Litteraturstudie som design	16
2.2 Litteraturstudie	18
2.2.1 Master data	18
2.2.2 Hva er god datakvalitet?	19
2.2.3 Hvilke strategiske, taktiske og operasjonelle valg må en virksomhet ta for å ha god master data management?	21
2.2.4 Kan master data hub være en del av master data lifecycle management?	25
2.2.5 Hvilke organisatoriske-, forretningsmessige- og IT-system berøres av en master data hub?	26
2.2.6 Hvilke faktorer påvirker valg av master data hub?	31
2.2.7 Hvem leverer denne type teknologi, og hvilke tjenester/kvaliteter er bygget inn i master data hub?	33
3. Metode	36
3.1 Forskningsdesign	36
3.2 Casestudie	38
3.3 Utvalg	39
3.4 Datainnsamling	40
3.4.1 Intervju	42
3.4.2 Elektronisk innsamling	43
3.4.3 Analyse av datainnsamling	45
3.5 Vurdering av data	46
3.5.1 Reliabilitet	46
3.5.2 Validitet	46
3.5.3 Overførbarhet	48
3.5.4 Bekreftbarhet	49

4. Datafunn	51
4.1 Presentasjon av caseobjekt og informanter	51
4.2 Leverandør	52
4.3 Bedrift 1	60
4.4 Bedrift 2	67
5. Diskusjon	70
5.1 Hva er god datakvalitet?	70
5.2 Hvilke strategiske, taktiske og operasjonelle valg må en virksomhet ta for å ha god master data management?	76
5.3 Kan master data hub være en del av master data lifecycle management?	80
5.4 Hvilke organisatoriske-, forretningsmessige- og IT-systemer berøres av en master data hub?	83
5.5 Hvilke faktorer påvirker valg av master data hub?	85
5.6 Hvem leverer denne type teknologi og hvilke tjenester/kvalitet er bygget inn i disse, der man går ett steg lenger?	91
6. Konklusjon	94
Referanser	98
Vedlegg	102
Vedlegg 1 - Oversikt over vårt litteratursøk	102
Vedlegg 2 - Data Quality Dimensions, vurdering av datakvalitet	103
Vedlegg 3 - Informasjonsskriv	104
Vedlegg 4 - Første intervjuguide til virksomhetene i casestudien	108
Vedlegg 5 - Første intervjuguide til leverandøren i casestudien	109
Vedlegg 6 - Tidslinje for forskningsarbeidet	110
Vedlegg 7 - Andre intervjuguide til virksomhetene	112
Vedlegg 8 - Sentrale datafunn	113

Figuroversikt

Figur 1: Subjective and objective assessments	17
Figur 2: Master data lifecycle management (MDLM) map	22
Figur 3: Central Master Data System	24
Figur 4: Repository	25
Figur 5: Desentralisert distribusjon av master data	26
Figur 6: Overordnet arkitektur til en desentralisert master data hub	27
Figur 7: Best-Practice MDM Implementation Roadmap for Typical B2C Customer Data Scenario	30
Figur 8: Arkitekturen til Leverandør sin tjenesteplattform	49
Figur 9: Illustrasjon av arkitekturer, Bedrift 1	62
Figur 10: MDM Roadmap for Typical Product/B2B Customer Data	64

1. Introduksjon

Innledningen til vårt forskningsarbeid består av forklaring av nøkkelbegrep, bakgrunn, forskningsspørsmål, posisjonering og oppgavens oppbygging. Delkapittel “1.1 Nøkkelbegrep” er en viktig innledning til oppgaven, da forståelse rundt ord og begreper skal sørge for en god flyt i lesingen. I delkapittelet “1.2 Bakgrunn” skal vi gjøre rede for nåværende ståsted innen forskningen og hva det er vi skal forske videre på. I delkapittelet “1.3 Forskningsspørsmål” skal vi introdusere våre forskningsspørsmål med oppsummerende antakelser hentet fra litteraturstudien vår. Delkapittel “1.4 Posisjonering” skal gi innsikt i vår motivasjon for studien, hvor vi posisjonerer vårt forskningsarbeid og hva vi mener forskningen kan bidra til. Til slutt skal vi introdusere oppgavens videre struktur.

Noen engelske ord i denne studien lar seg ikke godt oversette til norsk, og vi har derfor valgt å stille disse i kursiv i oppgaveteksten vår. I tillegg er det lagt inn fotnoter på tekniske begreper der leser ikke nødvendigvis vet betydningen. Disse er ikke betydelig nok til å ansees som nøkkelbegreper.

1.1 Nøkkelbegrep

Nøkkelbegrep	Definisjon	Kapittel
Master data	Omfatter forretningsobjekter, definisjoner, klassifisering og terminologi som utgjør informasjon om forretningen. Typiske klasser innen master data er leverandører, kunder og artikler.	2.2.1
Datakvalitet	Et mål på om data er nøyaktig, konsis, tilgjengelig, tidsriktig, fullstendig og gyldig. Når data har høy kvalitet kan den brukes til prosessene som er tiltenkt, uten at det fører til operasjonelle feil og mangler.	2.2.2
<i>Data quality management</i>	Prosesser som skal bedre datakvalitet. Dette innebærer verktøy for dataintegrasjoner og master data management produkter (som <i>artificial intelligence</i> og <i>machine learning</i>)	2.2.2
<i>Data governance</i>	Virksomhetens regler som omfatter data, struktur og karakteristikker i master data, eierskap, tilgangsstyring og rettigheter	2.2.3
Master data management	Refererer til prosessene som omfatter administrering og vedlikehold av master data.	2.2.3
Master data hub (Kan også omtales som master data plattform)	En tjeneste innen master data management for håndtering av master data. I en master data hub har man muligheten til å samle master data på ett sted, etter behov, og man tilgjengeliggjør master data for digitale tjenester.	2.2.5
Informasjonssystem	Blir i oppgaven vår omtalt som <i>enterprise resource planning</i> (heretter referert til som ERP) og <i>customer relationship management</i> (heretter referert til som CRM) system. På fagspråket blir disse også omtalt som kjernesystemer.	-

1.2 Bakgrunn

I en raskt voksende digital verden er virksomheter avhengig av å henge med på teknologiske trender og utvikling for å være konkurransedyktige. Konkurransfordeler stammer fra at virksomhetene evner å gjøre gode eksterne og interne analyser for å kunne se forbedringspotensialer. Kvaliteten i master data er kritisk for all form for digitalisering og samhandling i virksomhetens verdikjede, samt rapportering og analyser. Vilminko-Heikkinen & Pekkola (2013) mener at master data management har blitt anerkjent etter å ha sett en kraftig vekst innen mengden data virksomheter håndterer. I dag er data ofte lagret i ulike informasjonssystem og databaser. Dette er spesielt problematisk når virksomheter lagrer samme data i flere forskjellige informasjonssystem, som en konsekvens av at data har blitt utviklet og lagret i siloer over de siste tiårene. Christensen (2021) mener virksomheter ikke kan lykkes med digitalisering dersom man ikke rydder opp i master data og databasene. En avkoblet tjenesteplattform kan være veien å gå for å oppnå dette.

Master data management kan løse problemer knyttet til dårlig datakvalitet forårsaket av data kommer fra ulike informasjonssystem, gjennom prosesser som dataprofilering¹ og standardisering (Hikmawati et al., 2021). Datakvalitet er et stort problem for virksomheter som har informasjonssystem fordelt på flere forskjellige enheter og avdelinger. Slik arkitektur legger til rette for problemer i datakvalitet som duplikater og inkonsistens. Gjennom å bruke master data management forventer man å se en reduksjon i disse problemene (Hikmawati et al., 2021; Lerche, 2014). Gjennom uavhengige analyse- og forskningsrapporter ser vi en tendens til at avansert master data management, som *artificial intelligence* (heretter referert til som AI) og *machine learning* (heretter referert til som ML), kan bidra til økt datakvalitet. Det er definert mange dimensjoner som skal måle datakvalitet, men det som går igjen flest ganger er variasjoner av nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet (Christen, 2012; DAMA, 2013; Hikmawati et al., 2021).

I midlertidig vet vi ikke om implementering av master data management, i form av en master data hub, vil kunne bidra til økt datakvalitet. Vi skal derfor studere dette nærmere i denne masteroppgaven.

¹ Dataprofilering er aktiviteten å lage et forståelig bilde av innholdet og strukturen til en datakilde

1.3 Forskningsspørsmål

For å kunne se nærmere på om implementering av en master data hub kan bidra til å øke datakvaliteten i en virksomhet har vi utredet følgende forskningsspørsmål:

1. Hva er god datakvalitet?

Dette forskningsspørsmålet har vi svart på gjennom en litteraturstudie om datakvalitet og *data quality management*. Litteraturen beskriver hvordan man vurderer datakvalitet (Pipino et al., 2002), hvor viktig *data quality management* er for virksomhetene (Hikmawati et al., 2021) og hvilke barrierer man kan knytte til god kvalitet i master data (Haug & Arlbjørn, 2011). I tillegg har vi gjort en casestudie der to virksomheter har bidratt med å belyse hvordan kvalitet i master data oppleves og vedlikeholdes i virksomhetene deres. Formålet her var å finne ut hvordan datakvalitet måles, hvem eller hva påvirker datakvalitet, og hvordan virksomheten oppnå høy datakvalitet.

2. Hvilke strategiske, taktiske og operasjonelle valg må en virksomhet ta for å ha god master data management?

Dette forskningsspørsmålet er besvart gjennom en litteraturstudie av master data management. Litteraturen beskriver hvordan virksomheter med flere informasjonssystem kan oppnå *single point of truth* (Davenport & Prusak, 1998; Spruit & Pietzka, 2015), hvordan de kan oppnå suksessfull master data management (Hikmawati et al., 2021; Vilminko-Heikkinen & Pekkola, 2013) og hvorfor *data governance* er viktig for å oppnå dette (Beasty, 2008). Hikmawati et al., (2021) beskriver hvorfor dette er viktig for fremtidig vekst. Forskningsspørsmålet er også besvart gjennom vår casestudie av virksomheter som ønsker å implementere, eller som allerede har implementert, en master data hub. Formålet her var å finne ut hvilke tiltak som fører til god eller manglende master data management.

3. Kan master data hub være en del av master data lifecycle management?

Dette forskningsspørsmålet blir besvart med bakgrunn i master data og en litteraturstudie av master data sin livssyklus, og da spesielt Ofner et al. (2013) sitt rammeverk for *master data lifecycle management*. I tillegg har vi brukt casestudien vår til å se på hvordan virksomhetene håndterer master data. Formålet med dette spørsmålet var å undersøke om master data hub kan håndtere hele livssyklusen til master data, eller ikke.

4. Hvilke faktorer påvirker valg av master data hub?

Dette forskningsspørsmålet henger tett sammen med kvalitet i master data (som omtales i forskningsspørsmål 1), samt at vi har sett på Spruit & Pietzka (2015) sin modell for modenhet innen master data management. Vår litteraturstudie viser at det er virksomheter som er modne innen master data management som ender opp med å velge å ta i bruk en master data hub. Vi har også sett på hvilke faktorer som ligger bak valg av master data hub blant virksomhetene i casestudien vår, samt sett på hvordan de forholder seg til modellen for modenhet. Formålet med dette var å finne ut hvorfor virksomheter velger å implementere master data hub, og om datakvalitet er en av årsakene.

5. Hvilke organisatoriske-, forretningsmessige- og IT-system berøres av en master data hub?

For å besvare dette forskningsspørsmålet har vi gjort en litteraturstudie av ulike arkitekturer for opprettelse, lagring, vedlikehold og distribusjon av master data. Hvordan virksomheten skal utveksle data må avgjøres på et organisatorisk nivå, og prosesser knyttet til harmonisering, *cleansing*², opprettelse, vedlikehold og arkivering av data må defineres (Loser et al., 2004). Dette forskningsspørsmålet blir også belyst i casestudien vår. Vi har studert et pågående prosjekt for artikkel master data og hvordan en master data hub fungerer i praksis. I tillegg har vi studert hvordan arkitekturen og teknologien til en leverandør av master data hub kan se ut. Formålet med dette forskningsspørsmålet var å studere arkitekturen til en master data hub, og få innsikt i hvilke system i en virksomhet som berøres av den.

² Cleansing er å oppdage og oppdatere/fjerne ukorrekte data

6. Hvem leverer denne type teknologi, og hvilke tjenester/kvaliteter er bygget inn i master data hub?

For å kunne svare på dette forskningsspørsmålet har vi sett på Yuhanna et al. (2021) og Parker et al. (2021) sine analyser av leverandører av master data management i 2021. Hvem som leverer denne typen teknologi og hvilke tjenester/kvaliteter som er bygget inn i deres master data hub er viktig å vite for å kunne svare på hvilke teknologier som kan være en faktor når det kommer til datakvalitet. Gjennom casestudien vår har vi sett nærmere på hvordan teknologien og kvalitetene tjenesteplattformen til leverandøren vår ser ut, og om denne sammenfaller med det som ansees å være best på markedet. Formålet med dette spørsmålet var å finne ut hvilken teknologi som ligger til grunn i en master data hub, og hva som eventuelt kan påvirke datakvalitet.

Disse forskningsspørsmålene danner grunnlaget for følgende problemstilling:

Kan bruk av master data hub bidra til økt datakvalitet?

1.4 Posisjonering

Vi mener det er et gap i tidligere forskning når det kommer til å se på sammenhenger mellom spesifikke verktøy innen master data management, og dets direkte påvirkning på datakvalitet. Dette er motivasjonen for vår forskning. Vi ønsker derfor å studere om en master data hub kan være løsningen for å bidra til god kvalitet i master data ved å utføre en casestudie som skal beholde, utvikle eller skape ny teori. Denne masteroppgaven er et bidrag til forskningen innen master data management, og spesifikt master data hub. Vi ønsker å se på hvilke typer virksomheter som tar steget for å implementere master data hub, hvilke faktorer som fører til denne implementeringen og til slutt vil vi se om datakvaliteten øker ved bruk av denne typen verktøy.

1.5 Oversikt over oppgavens oppbygging

Denne masteroppgaven består av følgende kapittel; teori, metode, datafunn, diskusjon og konklusjon. I kapittel 2, teori, har vi gjennomgått en utdyping, diskusjon og argumentasjon for de forskningsspørsmålene som er satt. Kapitlet er strukturert etter spørsmålene med tilhørende relevant litteratur. Vi vil gå nærmere inn på hvordan vi har gått frem i litteraturstudien, og redegjøre for våre funn. I kapittel 3, metode, redegjør vi for vårt forskningsdesign, utvalgsstørrelse, utvalgsstrategi og rekruttering av informanter til vår casestudie. I tillegg vil vi redegjøre for vår datainnsamling, som består av semistrukturerte intervju og elektronisk kommunikasjon. En viktig del av kapitlet er vurdering av studiens reliabilitet, validitet, overførbarhet og bekreftbarhet av data vi har samlet inn.

I kapittel 4, datafunn, presenterer vi data som er samlet inn på en strukturert måte ved å introdusere informantene og deres virksomhet, sammen med tilhørende datafunn. I kapittel 5, diskusjon, vil vi sammenstille datafunn med antakelser fra vår litteraturstudie, der kapitlet er strukturert etter forskningsspørsmålene slik at det er tydelig hva vi diskuterer. Vi vil også kommentere om teorien beholdes, modifiseres og utvikles, eller må bygges på nytt. I kapittel 6, konklusjon, svarer vi på vår problemstilling med utgangspunkt i diskusjonene som er utført. Konklusjonen vår inneholder praktiske implikasjoner knyttet til oppgaven, samt forslag til videre forskning.

2. Teoretisk grunnlag

Det teoretiske grunnlaget i denne oppgaven består av en utdypning, diskusjon og argumentasjon for relevansen av forskningsspørsmålene. Vi mener det er vesentlig å definere hva master data er, før vi går videre med forskningsspørsmålene våre. Vi skal se på hvorfor master data management er viktig for virksomheter, hvorfor master data hub er en del av master data management, og hva som definerer god datakvalitet. Vi har også sett på litteratur innen teknologisk modenhet, noe vi mener må være til stede i en virksomhet som skal vurdere å ta i bruk en master data hub. En viktig del av denne studien har vært å se på hvilken teknologi og tjenester som ligger bak master data management, som master data hub.

2.1 Litteraturstudie som design

For å strukturere vårt litteratursøk har vi brukt Brocke et al. (2009) sitt rammeverk for litteraturstudie. Dette rammeverket består av 5 viktige faser; definere objektet man skal studere, konseptualisere emnet, søk i litteraturen, analyse av litteraturen og forskningsagenda. Undersøkelsesobjektene for våre litteratursøk har vært master data, master data management, datakvalitet, *data quality management*, og master data hub/master data plattform. For å finne relevant litteratur om disse nøkkelordene har vi i hovedsak bruk Google Scholar som søkemotor, og all litteratur har stort sett vært på engelsk. For å finne artikler som kan benyttes i forskningen vår studerte vi abstraktene til artiklene som dukket opp i de spesifikke søkene, for deretter å lese introduksjonen og konklusjonen. I de tilfellene der vi kom over interessante artikler, ble artiklene gjennomgått i sin helhet. Videre brukte vi Google Scholar sin funksjon “*cited*” for å se hvem og hvor artiklene har blitt brukt som referanser. Dette refererer Brocke et al. (2009) til som *forward search*. På denne måten har klart å grave oss dypere ned i litteraturen, ved å finne artikler som ikke ellers har dukket opp i våre søk.

I vår litteraturstudie fant vi svært mye informasjon om master data, datakvalitet, og ikke minst master data management (se vedlegg 1). Mengden informasjon gjorde det krevende å sile ut data som var relevant for våre forskningsspørsmål, og det ble tidvis noe overveldende. Det har derimot vært vanskelig å finne relevante artikler, studier og bøker som tar for seg

master data hub eller master data plattform, men det er noe av årsaken til at vi ønsket å skrive om dette temaet. Mange artikler skriver i stor grad om datavarehus og data lakes når man omtaler oppbevaring av data. Vi har ikke funnet noen artikler eller studier som sier noe om en master data hub forbedrer data kvaliteten. Vi mener dette er et gap i litteraturen, og det er vår motivasjon til å skrive denne oppgaven.

Vi endte til slutt opp med 38 kilder til vår litteraturstudie. Inkluderingskriteriene våre var at kildene våre inneholdt ett eller flere nøkkelord, hvor ofte de er referert til, da vi tolker dette som at kildene er relevant og troverdige, og vi har sett på publiseringsdato for å se om definisjoner eller litteraturen har endret seg over tid, samt at nyere artikler kan være lite sitert, men fortsatt aktuelle for vår forskning. Vi startet litteratursøket vårt med “*master data hub*” og “*master data platform*”. Artiklene som kom opp da viste i stor grad til “*master data management*”, noe vi derfor søkte videre på. I artiklene som handlet om “*master data management*” var det også skrevet en del om datakvalitet, som igjen førte til at vi søkte på etter dette nøkkelordet. Funn i litteraturstudien gjorde at vi etter hvert også søkte på “*data quality management*”. Ekskluderingskriteriene bestod i stor grad av tilgjengelighet, der artikler som ikke ligger ute som PDF-filer eller er gratis gjennom Google Scholar har blitt ekskludert. Bøker som vi ikke har funnet i norsk bokhandel eller via “Bookis” har også blitt ekskludert. I tillegg ble kilder som ikke var nordiske eller engelsk ekskludert. Litteraturstudien vår består dermed av fagartikler, fagbøker og uavhengige rapporter/analyser fra forsknings- og rådgivningsselskap. En av utfordringene ved å finne relevant litteratur er at nøkkelordene våre er like, noe som gjør det vanskelig å finne spisset litteratur knyttet til de spesifikke nøkkelordene. Basert på denne litteraturstudien har vi gjort oss noen antakelser som vi tar med oss videre inn i en casestudie og diskusjonskapittelet.

2.2 Litteraturstudie

2.2.1 Master data

For å kunne besvare forskningsspørsmålene våre er det vesentlig at vi har en forståelse for hva master data er og hvilken rolle den spiller i virksomheten. Hva master data betyr for den enkelte virksomheten varierer noe, men gjennom litteraturstudien har vi funnet det vi mener er gode beskrivelser av hva master data er, og hvorfor god kvalitet i master data er viktig for en virksomhet.

Master data omfatter forretningsobjekter, definisjoner, klassifisering og terminologi som utgjør informasjon om forretningen (Vilminko-Heikkinen & Pekkola, 2013). Typiske klasser innen master data er leverandører, kunder og artikler. Disse klassene spiller en vesentlig rolle i forretningens automatisering av prosesser og dataanalyse. Dermed er høy kvalitet i master data veldig viktig for virksomheter. Situasjonen i dag er at virksomheter bruker flere applikasjoner med egne datasett for å understøtte ulike forretningsprosesser. Som regel innebærer dette heterogene informasjonssystemer hvor master data verken er tidsriktige eller konsistente mellom systemene (Loser et al., 2004). Master data er en effekt av transaksjonsdata, der transaksjonsdata beskrives som relevante hendelser i en virksomhet som ordre, faktura, betalinger og leveringer. Transaksjonsdata vil ikke kunne utføre sin funksjon dersom master data ikke er korrekt, fordi transaksjonsdata bruker master data.

Feil i master data kan resultere i signifikante kostnader eller eventuelt inntektsfall for virksomheter. Årsaker til feil i, eller inkonsistent, master data kan være systemarkitektur, utilstrekkelig koordinering med forretningsprosesser, utilstrekkelig implementering av programvare eller uoppmerksom brukeradferd (Haug & Arlbjørn, 2011). Med korrekt *data governance* og oversikt vil, data i master data systemet (eller oppbevaringssted/register) kvalifiseres som en enhetlig og konsistent dataressurs, som alle applikasjoner kan stole på for informasjon av høy kvalitet (Loshin, 2009). Master data med høy kvalitet er en forutsetning for at organisasjoner skal kunne oppnå strategiske forretningsmål, som bedre beslutningstaking (Ofner et al., 2013). De viktigste dimensjonene når man skal se på datakvalitet er aktualitet, konsistens, nøyaktighet og integritet (Morbey, 2013).

2.2.2 Hva er god datakvalitet?

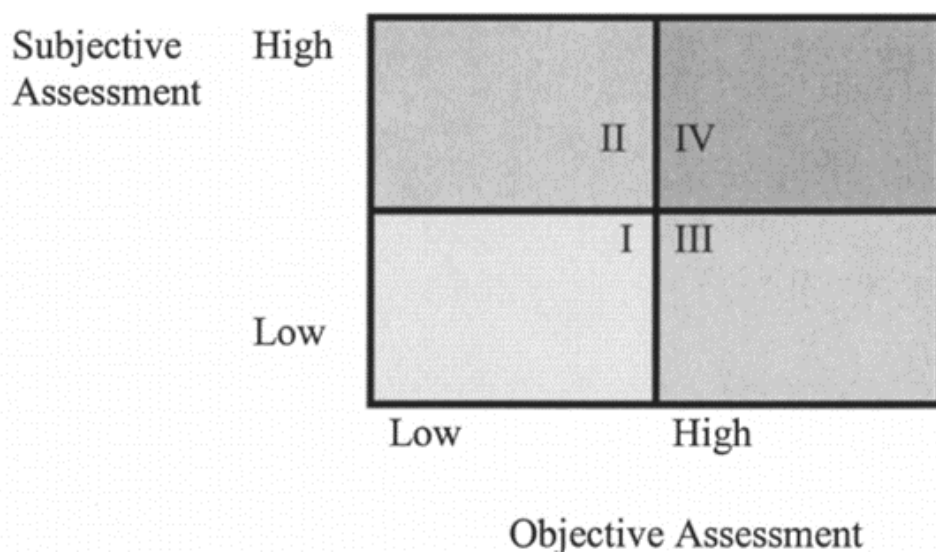
Dette forskningsspørsmålet vil hjelpe oss å danne et bilde av hva datakvalitet er, og på den måten gi oss grunnlaget til å besvare om en master data hub kan bidra til å øke datakvalitet. Litteraturstudien av datakvalitet avdekker at kvalitet kan være både subjektivt og objektivt. Noe som kommer tydelig frem i litteraturen er at alle er enige om at jo høyere kvalitet data har, jo bedre er det for virksomhetene. Master data management kan løse problemer knyttet til datakvalitet forårsaket av data som kommer fra ulike kilder gjennom prosesser som dataprofilering og standardisering (Hikmawati et al., 2021).

“Tradisjonelt handlet datakvalitet om overholdelse og styring for å redusere operasjonell risiko og kostnader, mens fokuset nå er at datakvalitet er en nødvendighet når man skal forsterke analyser for å bedre innsikt og ta pålitelige, datadrevne beslutninger” (Chien & Jain, 2021, Market/Definition/Description, avsn.1, vår oversettelse). De mener datakvalitet kan føre til konkurransefortrinn hos virksomheter. Informasjonssystem må være troverdig og pålitelig for å sikre kvalitet, konsistens og korrekthet i data for å støtte forretningsbeslutninger, og som et grunnlag for fremtidig vekst i virksomheten (Hikmawati et al., 2021). Datakvalitet er et stort problem for virksomheter der informasjonssystem brukes av flere forskjellige enheter og avdelinger. Slik arkitektur legger til rette for problemer i datakvalitet som duplikater og inkonsistens. “Gjennom å bruke master data management forventer man å se en reduksjon i disse problemene” (Hikmawati et al., 2021, s. 90; Lerche, 2014, vår oversettelse).

Datakvalitet er et mål på om data er tilstrekkelig for den enkelte prosessen data skal brukes til. Datakvalitet er et multidimensjonalt konsept som ikke kan beskrives av enkelte trekk, men ut ifra ulike dimensjoner og mål (Schäffer & Leyh, 2017). Det er definert mange dimensjoner som skal måle datakvalitet, men det som går igjen flest ganger er variasjoner av nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet (Christen, 2012; DAMA, 2013; Hikmawati et al., 2021). Schäffer & Leyh (2017) skriver at master data kvalitet ikke er direkte beskrevet i litteraturen, men kan sees på som en sammensetning av master data og datakvalitet, der master data består av attributter som beskriver essensielle

forretningsobjekter som danner grunnlaget for operasjonell verdiskaping i prosesser og analytiske beslutningsprosesser i virksomheten.

Hvor god datakvalitet en virksomhet har, og hvordan den måles, er det vanskelig å finne et entydig svar på. Datakvalitet er et flerdimensjonalt konsept hvor virksomhetene må forholde seg til både den subjektive oppfatningen til individene som er involvert i data og prosessene, samt de objektive målingene basert på spøringer i datasett (Pipino et al., 2002). Videre foreslår de tre steg for å gjøre vurderinger i praksis basert på subjektive og objektive vurderinger for å kunne forbedre datakvaliteten til virksomheten. Først må virksomheten analysere de subjektive og objektive vurderingene av datakvalitet. Deretter må de sammenligne resultatet av disse, for å kunne identifisere avvik og årsaker. Til slutt må virksomheten iverksette tiltak for forbedringer. Når virksomhetene skal utføre analyser basert på de subjektive og objektive vurderingene må de ulike dimensjonene sammenlignes, der resultat av analysene faller inn under en av fire kvadranter som vist i matrisen nedenfor (Pipino et al., 2002). Vurdering av datakvalitet baserer de på egendefinerte datadimensjoner (se vedlegg 2). Målet for datakvalitet er å ha høy subjektiv vurdering og høy objektiv vurdering. Skulle virksomhetene havne under en av de andre kvadrantene må man finne årsaker og gjøre korrigerende tiltak.



Figur 1: Subjective and objective assessments (Pipino et al., 2002, s. 217)

Det finnes ikke en *one size fits all* løsning, da vurdering av datakvalitet er en kontinuerlig prosess som krever bevissthet om grunnleggende prinsipper for subjektive og objektive målinger av datakvalitet (Pipino et al., 2002).

“*Data quality management* har modnet de siste årene, der leverandørene i dag fokuserer på å møte kravene i markedet som omhandler *workflow*³, roller, samarbeid og prosesser som skal overvåke, rapportere og utbedre problemer for datakvalitet” (Chien & Jain, 2021, Market Definition/Description, avsn.2, vår oversettelse). Videre skriver de at et resultat av dette er at man integrerer løsninger for datakvalitet gjennom blant annet verktøy for dataintegrasjon, metadata management løsninger og master data management produkter. Løsningene som skal sikre god datakvalitet innebærer prosesser og teknologier som skal identifisere, forstå og korrigere feil i data på tvers av operasjonelle forretningsprosesser og beslutningstaking. Dette utføres gjennom kritiske funksjoner som profilering, reformatering, standardisering, *cleansing*, *matching*⁴ og monitoring, samt oppretting av regler og analyser, innebygd *workflow*, kunnskapsbaser og samarbeid (Chien & Jain, 2021) I master data management hensyntas og ivaretas datakvalitet gjennom dokumenterte roller og ansvar som ligger inn under *data governance* (Buffenoir & Bourdon, 2013; Hikmawati et al., 2021).

2.2.3 Hvilke strategiske, taktiske og operasjonelle valg må en virksomhet ta for å ha god master data management?

Master data management er en sentral del av oppgaven vår, da master data management skal sikre bedre datakvalitet. Master data hub kan være en tjeneste for å oppnå dette. Bruk av master data hub er et resultat av god master data management. Derfor blir dette forskningsspørsmålet vesentlig for å kunne diskutere virksomheters modenhet for å ta i bruk master data hub, samt at de har et bevisst forhold til hvilken kvalitet master data i virksomheten har. Gjennom dette vil vi kunne se på om en master data hub vil kan forbedre datakvaliteten. Litteraturstudien vår viser at virksomheter må innarbeide master data management i sin strategi. Master data management er en taktisk beslutning virksomheter tar der IT-ressurser skal kunne garantere og opprettholde datakvalitet. Det er viktig at en

³ Workflow er en serie aktiviteter som er nødvendig for å gjennomføre en oppgave

⁴ Matching skal sjekke om datasett inneholder samme referanser og eventuelt duplikater

virksomhet anerkjenner at data er en viktig ressurs for forretningen, og som et resultat av dette iverksetter master data management.

Master data management refererer til prosessene som omfatter administrering og vedlikehold av master data. Vilminko-Heikkinen & Pekkola (2013) skriver at master data management har blitt anerkjent etter man så en kraftig vekst innen mengden data virksomheter håndterer. På nåværende tidspunkt er data ofte lagret i ulike informasjonssystem og databaser. Videre skriver de at det er spesielt problematisk når virksomheter lagrer samme data i flere ulike informasjonssystem, som en konsekvens av at data har blitt utviklet og lagret i siloer over de siste tiårene. Master data management og bruk av master data hub er en løsning til dette problemet, og skal hjelpe med å avdekke data som ligger lagret i siloene.

Master data management er en samling av de beste data management praksisene som orkestrerer interessenter, deltakere og forretningspartnere med å innlemme forretningsapplikasjoner, metoder for ledelse av informasjon og verktøy for ledelse av data. Dette brukes videre til å implementere *policies*, prosedyrer, tjenester og infrastruktur som støtter innhenting og integrasjon av data som er nøyaktig, betimelig, konsistent og komplett master data (Loshin, 2009). Det er en generell enighet, og sunn fornuft tilsier at korrekte, tilgjengelige og tidsriktige data er av stor betydning som kan gi et konkurransefortrinn (Borghoff & Pareschi, 1998; Kahn et al., 2002; Spruit & Pietzka, 2015). Det som imidlertid er en stor utfordring, er at mange selskaper ikke har tilstrekkelige data management strategier. De store selskapene sliter med at de er i besittelse av enorm mengde data, hvor strategien for å håndtere og utnytte disse er fraværende (Davenport & Prusak, 1998; Spruit & Pietzka, 2015). Master data management legger til rette for at virksomheter kan bygge en *single point of truth* som skal sikre at data er komplett, nøyaktig, tidsriktig og sikker. I tillegg kan master data management løse problemer knyttet til datakvalitet som duplikater, unøyaktigheter og inkonsistens.

Master data management innebærer fire prosesser; profilering av master data, som vil si å gjøre en vurdering av kvaliteten til ulike kilder; konsolidere master data til et *repository*⁵ og

⁵ Repository er et oppbevaringssted for master data

knytte master data til ulike eksisterende applikasjoner; rense master data og berike informasjon; synkronisere master data med virksomhetens forretningsprosesser gjennom tilknyttede applikasjoner som støtter *business intelligence* (heretter referert til som BI) og rapporteringssystemer (Hikmawati et al., 2021; Joshi, 2007).

Beasty (2008) mener det er viktig at virksomheter klarer å skille mellom *data governance* og *data management*, noe som kan resultere i fundamentale feil og svikt i kommunikasjonen mellom teknologiresurser og ledelsen. *Data governance* er virksomhetens regler som omfatter data, struktur og karakteristikker i master data, eierskap, tilgangsstyring og rettigheter. *Data management* er den taktiske utførelsen av virksomhetens *data governance* slik at IT-ressurser kan garantere og opprettholde datakvalitet, GDPR og korrigerer feil. Effektiv *data governance* kan forbedre kvalitet, tilgjengelighet og integritet til virksomhetens data ved å strukturere samarbeid på tvers av virksomheten (Hikmawati et al., 2021). Virksomheter forbedrer sin master data management gjennom *data governance*.

En vanlig utfordring i master data management er manglende kunnskap om livssyklusen til master data på tvers av forretningsenheter, avdelinger, forretningsprosesser, IT-systemer og flere ERP-system (Ofner et al., 2013). De mener problemet ligger i at de fleste forretningsdomenene i en virksomhet er interessert i å bruke master data for å dekke sine egne funksjonelle behov, noe som kan resultere i at master data blir tvetydig på tvers av virksomheten. Dette fører ofte til ukoordinerte master data ressurser, og inkonsistent og unøyaktig master data. Forståelse, planlegging og overvåking av master data sin livssyklus sees på som en hjørnestein innen master data management.

For å ha en suksessfull master data management prosess mener (Hikmawati et al., 2021; Vilminko-Heikkinen & Pekkola, 2013) at virksomheten må gjennom følgende steg;

- Identifisere hva som er virksomhetens behov for master data og hvilke system som bruker den
- Definere styring av master data management, prosesser for vedlikehold og standarder

- Fremtidige forbedringer for nåværende datakvalitet, planlegge arkitektur for master data management. Dette gjelder applikasjoner, dataflyt, datasikkerhet og personvernregler.
- Opplæring og kommunikasjon med alle interessenter
- Ha en strategi for utvikling av master data management
- Definere karakteristikk for master data management. Dette innebærer grensesnitt, *workflow*, endring av data, konsolidering og integrasjon.

For å kunne implementere master data management er virksomheter avhengig av å anerkjenne at data er en viktig forretningsressurs, som må behandles og administreres deretter (Vilminko-Heikkinen & Pekkola, 2013). Loshin (2009) mener det er flere fordeler knyttet til master data management, blant annet konsistent rapportering, forbedret konkurransevne, forbedret beslutninger, og bedre kostnadsanalyser og kostnadsplanlegging. Videre poengterer han at selv om master data management ikke kan sees på som en teknologi, er det helt klart at man ikke klarer å oppnå god master data management uten å utnytte seg av verktøy og teknologi. Det tekniske aspektet inneholder blant annet profilering av data, komplekse dataanalyser, metadata management, data modellering, integrasjon av data, standardisering, kobling og sammenstilling av *records*, *cleansing*, tjenestestyrt arkitektur, tilgangsstyring og levering av data. Master data management er en direkte årsak av tjenestestyrt arkitektur fordi det har åpnet for muligheten til å synkronisere data på tvers av systemer (Beasty, 2008).

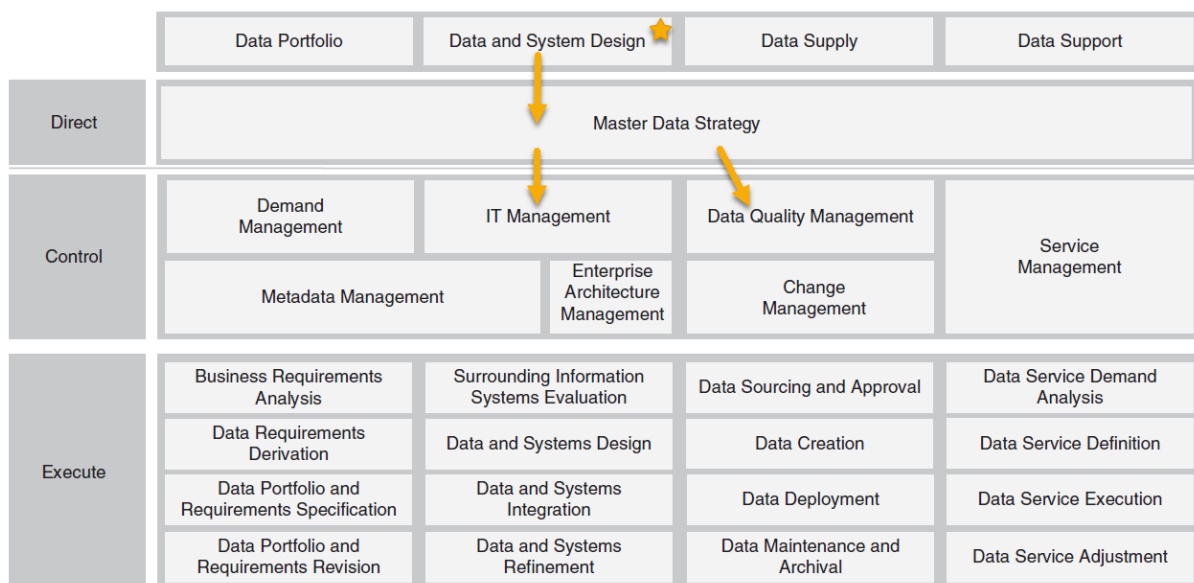
Virksomheter må definere roller, funksjoner og ansvar innen *data governance* slik at datakvaliteten blir vedlikeholdt (DAMA, 2013; Hikmawati et al., 2021; Loshin, 2009; Otto, 2011). Uklare roller for eiere av data kan resultere i dårlig datakvalitet ettersom eieren er ansvarlig for at data er korrekt (DAMA, 2013; Haug & Arlbjørn, 2011; Hikmawati et al., 2021; Loshin, 2009; Silvola et al., 2011; Vilminko-Heikkinen & Pekkola, 2012, 2017, 2019). Dårlig dokumentasjon for roller og ansvar til de ansatte som skal vedlikeholde datakvaliteten, som *data stewards*, fører til dårlig oppfølging og eventuelt overføring av ansvar til andre ansatte (DAMA, 2013; Hikmawati et al., 2021; Loshin, 2009; Silvola et al., 2011; Smith & McKeen, 2008; Vilminko-Heikkinen & Pekkola, 2019). I praksis har ikke virksomheter noen klare retningslinjer til de som håndterer data når det kommer til dataformat, flyten til registrering av data, hvordan oppdatere data, eller hvem som har ansvaret når det oppstår

problemer knyttet til data (Haug & Arlbjørn, 2011; Hikmawati et al., 2021; Silvola et al., 2011; Smith & McKeen, 2008; Vilminko-Heikkinen & Pekkola, 2017). Dette kan føre til at registrering av data blir unøyaktig, overflødig og ikke enhetlig.

2.2.4 Kan master data hub være en del av master data lifecycle management?

Da master data sin livssyklus sees på som en hjørnestein innen master data management har vi tatt med dette forskningsspørsmålet for å kunne se på om en master data hub kan være en del av denne livssyklusen. Gjennom litteraturstudien vår fant vi at en master data hub kan være en del av master data sin livssyklus, og mer spesifikt har vi sett på dette gjennom Ofner et al. (2013) sitt rammeverk for *master data lifecycle management map*.

Master data management innebærer all aktivitet som opprettelse, endring og sletting av master data. Livssyklusen til master data må behandles på samme måte som andre eiendeler (Wang et al., 1998; Ofner et al., 2013). *Master data lifecycle management* defineres som master data krav og -portefølje, design og implementering av en master data modell, -prosess, og –arkitektur, samt hvordan man skal forsyne brukerne med master data, og støtte brukerne av master data (Ofner et al., 2013).



Figur 2: Master data lifecycle management (MDLM) map (Ofner et al., 2013, s. 480)

Master data hub vil være en strategisk konsekvens av *master data strategy* og *data quality management* ifølge Ofner et al. (2013) sitt *master data lifecycle management map*. Master data management kan løse problemer knyttet til datakvalitet forårsaket av data som kommer fra ulike kilder gjennom prosesser som data profiling og standardisering (Hikmawati et al., 2021).

Master data strategy skal spesifisere de viktigste prinsippene og retningslinjene som har effekt på beslutninger og aktiviteter når det kommer til *data portfolio*, *data and system design*, *data supply chain* og *data support* (Ofner et al., 2013, s. 480, vår oversettelse). *Data portfolio* handler om å forstå forretningskrav, slik at man kan etablere retningslinjer og spesifikasjoner for master data portefølje og krav til data. *Data and system design* er en viktig del av livssyklusen til master data arkitektur fordi det setter retningslinjer for hvordan master data bør lagres og distribueres, overvåker implementeringen eller forbedringen av applikasjoner og databaser, og vedlikehold av teknisk metadata. Herunder ligger også design av master data modeller, -prosesser, -arkitektur og IT-systemer for å sikre at IT-systemene er teknisk klare. *Data supply chain* skal lage retningslinjer for opprettelse og vedlikehold av master data, overvåke kvaliteten på master data, samt administrere forespørsler, samle inn datakomponenter, og opprette og distribuere master data. *Data support* skal definere støtte som kreves av organisasjonen, overvåke mengden hendelser, problemløsning, og støtte master data brukere.

2.2.5 Hvilke organisatoriske-, forretningsmessige- og IT-system berøres av en master data hub?

Dette forskningsspørsmålet grunner i hvilken type arkitektur en master data hub er bygget på. Dette er viktig å få klarhet i for å kunne si noe om hvordan master data skal opprettes, lagres, vedlikeholdes og distribueres. Rutiner og prosesser rundt opprettelse, lagring, vedlikehold og distribusjon av master data skal være med på å sikre god kvalitet. Alle forretningsenhetene til virksomheten skal bruke master data som ligger i huben til å utføre sine forretningsprosesser, rapporter og analyser. Det er viktig at virksomheter har en god forståelse for arkitektur og en strategi før de begynner å bevege seg inn i implementeringen av en master data hub.

Loser et al. (2004) legger vekt på at det finnes fire ulike arkitekturer som kan være med å skape konsistent data; *Central Master Data System*, *Leading System*, *Master Data Harmonization via Standards* og *Repository*. De skiller mellom to dimensjoner; globale master data attributter, og opprettelse og vedlikehold av data og/eller distribusjon. Globale master data attributter avgjør hvilke attributter som skal være standardiserte på tvers av systemer. Ved opprettelse og vedlikehold av data er det viktig å avgjøre om master data skal opprettes og vedlikeholdes sentralt eller desentralt. Forskjellen på disse arkitekturene er hvor master data opprettes, lagres og vedlikeholdes, integrasjonen mellom systemene og hvordan master data distribueres, og om systemene krever mapping for å kunne bruke master data i sine prosesser.

Ved en arkitektur som *Central Master Data System* vil alle systemene bruke samme master data. Master data vil i en slik løsning alltid opprettes i det sentrale systemet, noe som sikrer at data opprettes likt og er unik. Overføringen av master data blir initiert i det sentrale systemet, og skjer asynkront. Dette fører til at data som applikasjonene skal bruke aldri vil være tidsriktige, da de er tilgjengelige etter en viss forsinkelse.

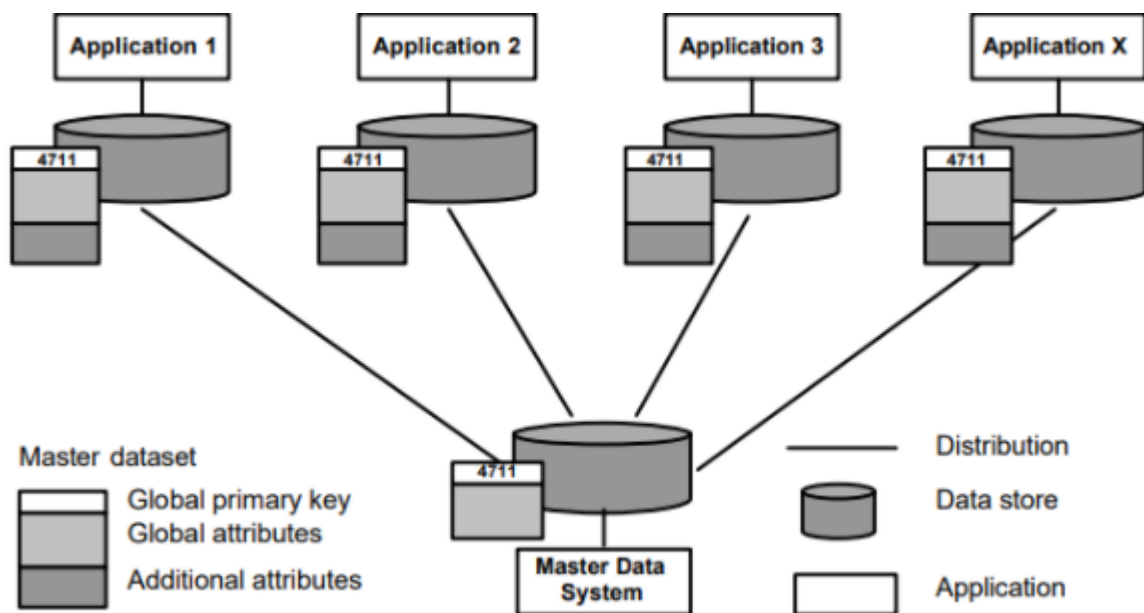


Figure 2: Central Master Data System

Figur 3: Central Master Data System (Loser et al., 2004, s. 2)

Med tilnærmingen ved arkitekturen *Repository* oppretter virksomheter et felles sentralt oppbevaringssted for alle systemer og applikasjoner med data som er involvert. Det blir sendt en spørring til oppbevaringsstedet om hvor (hvilket system/applikasjon) data man trenger ligger. Deretter mapper⁶ systemet sine egne data med master data som er etterspurt, og det er applikasjonen som etterspør master data, som er ansvarlig for å mappe mottakende data (Loser et al., 2004). Master data opprettes, vedlikeholdes og lagres desentralt med denne tilnærmingen. Oppbevaringsstedet har en global nøkkel som brukes som standard på tvers av systemene, og sørger for at alle applikasjoner bruker samme master data. Endringer som er gjort i master data vil umiddelbart være tilgjengelig når man etterspør data. Fordelen med denne tilnærmingen er at applikasjonenes autonomi opprettholdes, og man har kun en liten avhengighet til et sentralt system. Bakdelen er at master data opprettes og endres desentralt ut ifra ulike prosesser.

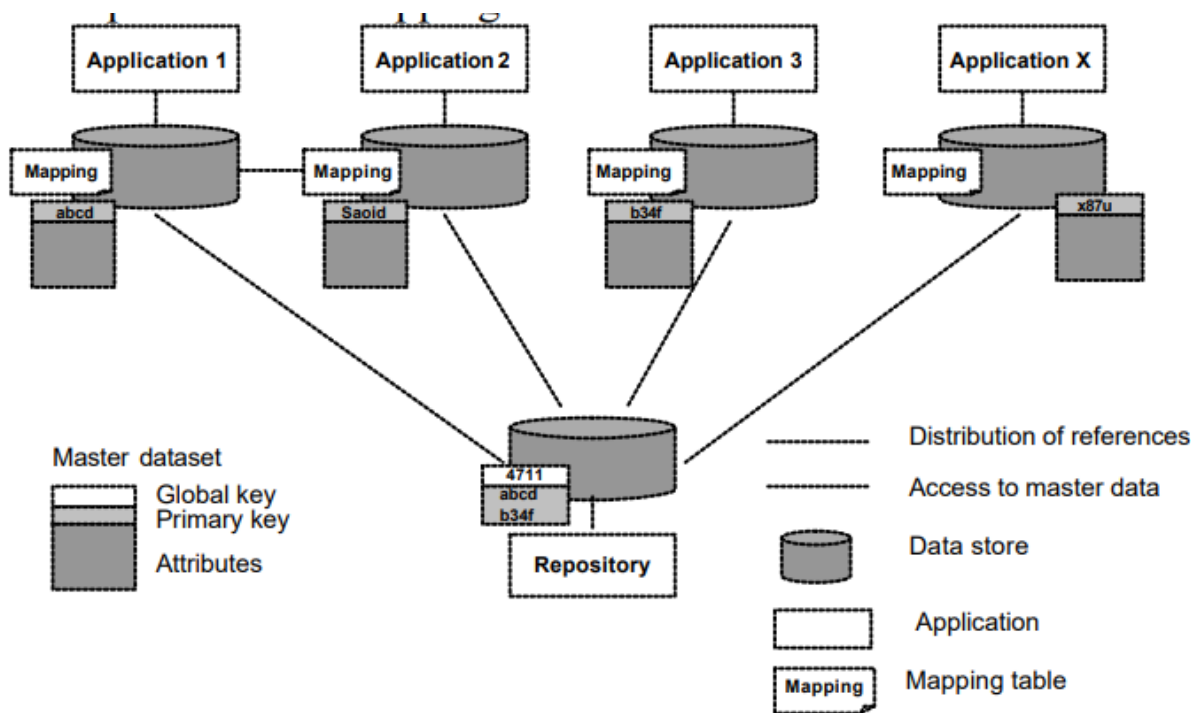
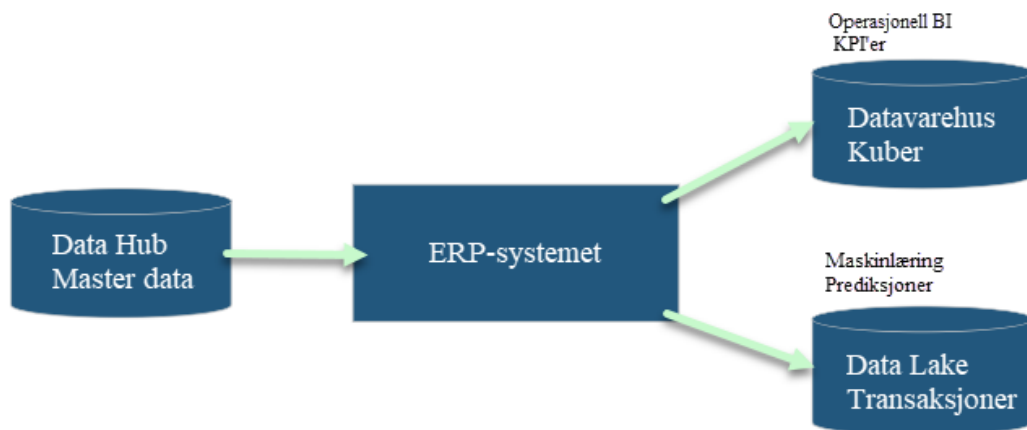


Figure 5: Repository

Figur 4: Repository (Loser et al., 2004, s. 4)

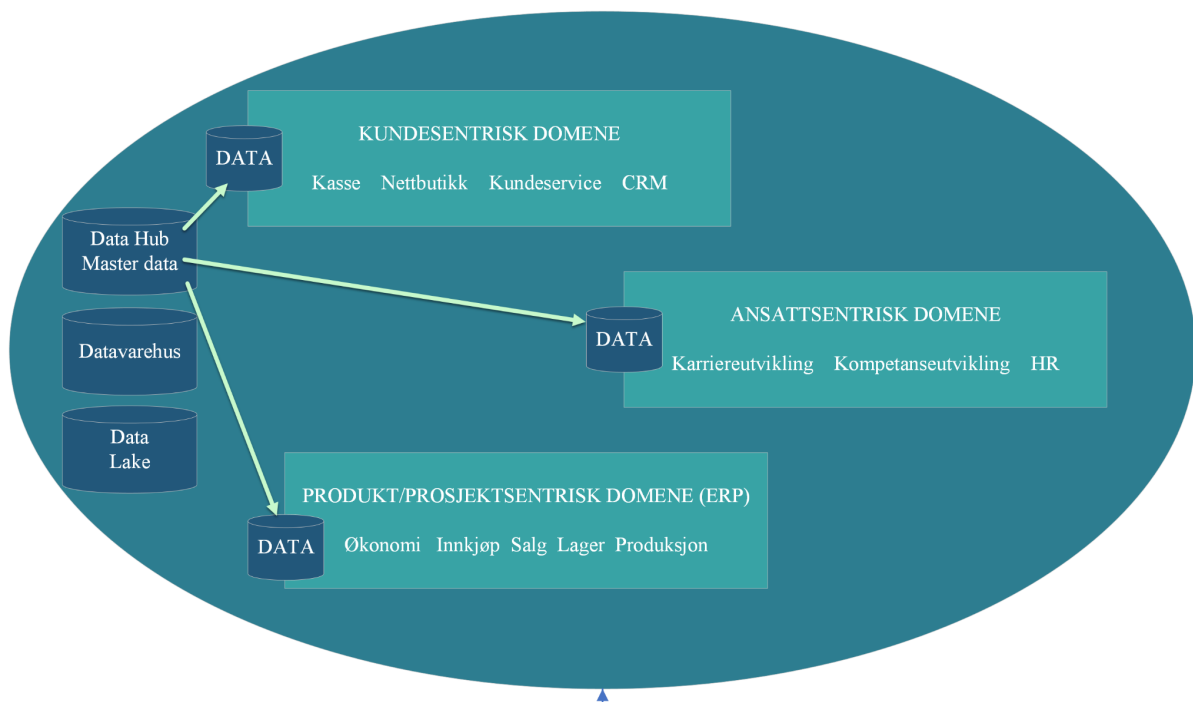
⁶ Mappe vil si å sammenstille felt fra flere datasett

Christensen (2021) mener at en master data hub skal være desentralisert og ligge frakoblet fra ERP-systemet. Videre skrive han at en av de viktigste oppgavene til et ERP-system er å administrere master data. Disse dataene skal nødvendigvis ikke kun brukes av ERP-systemet selv, men kan også brukes av integrerte nisjeløsninger til systemer hos kunder og leverandører (ekstern integrasjon). Kvaliteten i master data blir dermed ansett å være kritisk suksessfaktor for nesten all form for digitalisering og digital samhandling i verdikjeder. Videre illustrerer han hvordan en master data hub distribuerer master data til kjernesystemet, som videre distribuerer data til datavarehuset og data lake for ulike formål. Master data distribueres ved ekstraher, transformer og last (heretter referert til som ETL) eller ved ekstraher, last og transformer (ELT):



Figur 5: Desentralisert distribusjon av master data (egen figur)

Det kundesentriske domenet vil i et slikt tilfelle kalle master data fra master data hub, og sende data videre til applikasjoner og systemer for kundebehandling som markedsføring og salg. Kjernesystemene, som ERP og CRM, vil kalle master data fra master data huben og distribuere denne videre til BI for analyser i Power BI, eller inn i AI og ML for prediksjoner. På denne måten har man en master data hub som håndterer alle master data uavhengig om den skapes i HR-domenet, kundedomenet eller ERP-domenet:



Figur 6: Overordnet arkitektur til en desentralisert master data hub (egen figur)

Beasty (2008) beskriver en master data hub som en plattform med sentral arkitektur som utnytter at plattformen har en *golden record*⁷ med definisjon av data og attributter, som migrerer data ut til virksomhetens applikasjoner og sluttbruker på bestilling. Han beskriver en master data hub som transaksjonsbasert og som fungerer for sentraliserte, IT-drevne forretninger som har et begrenset antall domener og forretningslinjer. Han er også enig i at en master data hub kan være basert på *registry* som distribuerer data fra en *repository*. En slik sentralisert master data hub vil være transparent og fungerer best for organisasjoner som opererer med flere databaser og systemer. Det viktigste for virksomheten er å velge en arkitektur som passer best til organisasjonens *data governance* og *workflow*.

Selv om det i dag finnes ulike tilnærminger til arkitektur og software løsninger som skal hjelpe virksomheter å ivareta master data, er ikke disse verktøyene nok i seg selv for å oppnå en bekymringsløs utveksling av master data objekter mellom ulike systemer og applikasjoner. Hvordan virksomheten skal utveksle data må avgjøres på et organisatorisk nivå der prosesser knyttet til harmonisering, *cleansing*, opprettelse, vedlikehold og arkivering av data må defineres (Loser et al., 2004).

⁷ Global nøkkel

2.2.6 Hvilke faktorer påvirker valg av master data hub?

Gjennom vår litteraturstudie har vi kommet frem til at det er virksomheter som er modne innen master data management som ender opp med å ta i bruk master data hub. Formålet med dette forskningsspørsmålet er å finne ut hva som fører til at virksomheter velger å implementere en master data hub.

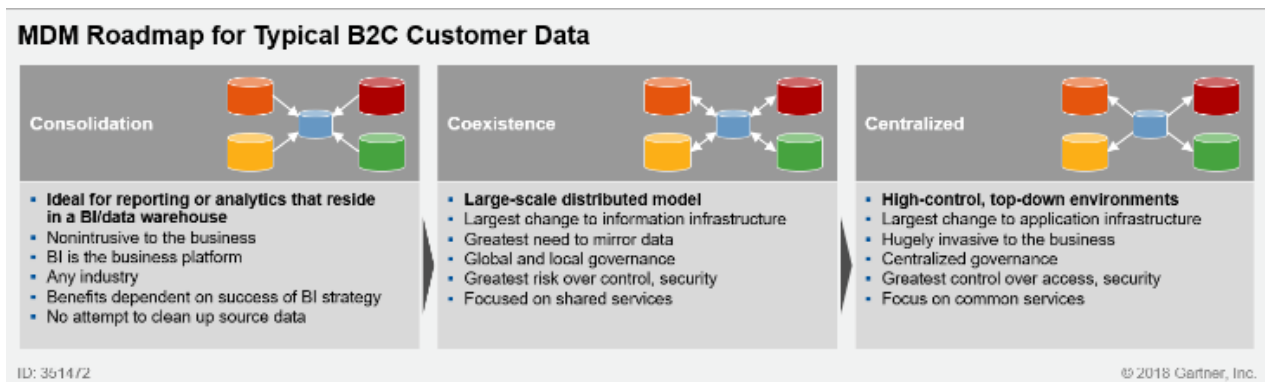
Uavhengig av om en virksomhet har vokst organisk eller via oppkjøp og fusjoner er det nærliggende å tro at forretningssystemene ikke er i stand til å gi en konsistent oversikt eller kilde til master data (Destein, 2007). Videre skriver han at virksomheter vil støte på utfordringer knyttet til å sammenstille flyktige data som distribueres på tvers av flere ulike datamodeller, system og eksterne datakilder. Det å opprette sentralisert master data som integreres, ledes og deles på tvers av organisasjonen kan høres ut som et enkelt konsept, men det er vanskelig å få til uten en master data hub. En master data hub sammenstiller både interne og eksterne data referanser med kontoer og transaksjoner, og gjør det mulig å gi pålitelig og enhetlig syn på tvers av forretningsfunksjoner. Uten dette vil ikke virksomhetens ulike forretningslinjer effektivt møter virksomhetens forretningsmål, eller kunne utnytte seg av muligheter for vekst og økt profitt (Destein, 2007).

Spruit & Pietzka (2015) har laget med modell som skal se på organisasjonens modenhet når det kommer til master data management (heretter referert til som MD3M). Denne modellen ser på en virksomhets *data model*, *data quality*, *usage & ownership*, *data protection* og *maintenance* for å måle grad av modenhet. *Data model* ser på hvilke data virksomheten definerer som master data, hvordan disse er strukturert, hvilke systemer som bruker master data og hvor master data ligger lagret. *Data quality* går ut på om virksomheten har et forhold til hvilken kvalitet deres master data har, hvordan kvaliteten kan påvirke forretningen, hvor kilden til dårlig kvalitet i master data ligger og hvordan virksomheten kan forbedre kvaliteten på sin master data. Dette er viktig da det er datakvaliteten som til syvende og sist gjør det mulig for virksomheter å oppnå konkurransefordeler og redusere feil i forretningsprosesser. *Usage & ownership* gir virksomheten muligheten til å definere hvem som bruker data i hvilke systemer, hvem har lese/skrive tilganger og om man har klare retningslinjer knyttet til hvem som har disse autorisasjonene. *Data protection* handler om hvor og hvordan data er sikret mot

mulige forhold som svikt i komponentene, software bugs eller spesifikke angrep. *Maintenance* handler om fysisk lagring og livssyklusen til master data. For å kunne dra nytte av tekniske innovasjoner er man avhengig av at system er oppdatert og vedlikeholdt korrekt, samt at data som er i bruk må holdes ren ved å arkivere eller slette utdaterte data i henhold til livssyklusen (Bowker et al., 2009; Spruit & Pietzka, 2015).

Forstyrrelser, kompleksitet og kostnader knyttet til master data management setter taktiske og strategiske fordeler i fare. Implementering av master data management skal være utviklet som en del av virksomhetens visjon og strategi (O’Kane et al., 2018). Gartner sine klienter har implementert master data management ved å bruke konseptet *Master Data Management-implementeringsstiler* som skal bidra å øke hastigheten for levering av forretningsverdi. De viser til flere stiler for implementering av en master data hub, og at virksomheter må finne områder hvor en eller flere av de mindre inngripende stilene kan benyttes. Virksomheter velger stiler for implementering etter beste praksis bestående av to variabler; harmonisering og kontroll, og graden av forstyrrelser og hvor inngripende tiltak virksomheten er villig til å ta på tvers av både IT og forretningen.

Hvilken implementeringsstil som skal benyttes kan være avgjørende for hvilket brukstilfelle de er tiltenkt. (O’Kane et al., 2018) viser til to brukstilfeller; B2C (harmonisering) og B2B (kontroll). B2C refereres til som *downstream harmonization* og B2B refereres til som *upstream control*. Figuren nedenfor viser beste praksis for B2C-kundedata der en mindre inngripende stil er første steg i implementeringen av master data hub. Målet er en vellykket kartlegging og lasting av kildedatasett, og implementering og justering av regler for *cleansing*, transformasjon og harmonisering for å oppfylle forretningskrav. Ved å bruke beste praksis vil virksomheter kunne avdekke hvor god datakvaliteten er, samt om det finnes kompatibilitetsproblemer tidlig i fasen slik at man slipper å gjøre om arbeidet som allerede er utført.



Figur 7: Best-Practice MDM Implementation Roadmap for Typical B2C Customer Data Scenario (O’Kane et al., 2018, s. 8)

2.2.7 Hvem leverer denne type teknologi, og hvilke tjenester/kvaliteter er bygget inn i master data hub?

Hvem som leverer denne typen teknologi og hvilke tjenester/kvaliteter som er bygget inn i deres master data hub er viktig å vite for å kunne svare på hvilke teknologier som kan påvirke datakvalitet. For å finne ut av dette har vi sett på ulike rapporter som tar for seg de største leverandørene i markedet, basert på inntekt og antall kunder. Rapportene inneholder ulike løsninger for data management, der det sees på nåværende produkt og strategi. Leverandørene deles inn i *leaders*, *strong performers*, *contenders* og *challengers* (Yuhanna et al., 2021).

Yuhanna et al. (2021) sin rapport om master data management baserer seg på en score som er fordelt 50/50 mellom nåværende produkt og strategi. For nåværende produkt er det scoren for datakvalitet, sammen med dataintegrasjon og datasikkerhet, som vektlegges sterkest. Master data management produktet de leverer er ikke knyttet til en spesifikk applikasjon, BI, prediktive analyser eller integrasjonsmønstre. Produktet må kjøres i et frittstående miljø som støtter ulike typer arbeidsbelastninger. Derfor er det viktig å se på hvilke kvaliteter leverandørene som scorer høyt på datakvalitet har i sitt master data management verktøy. Det som ser ut til å ha sterkest korrelasjon med datakvalitet, ut fra rapporten, er innebygde AI og ML aktiviteter. Innen 2022 vil 60% av virksomheter bruke *ML-enabled data quality* teknologi for å bedre datakvaliteten sin (Chien & Jain, 2021). Egenskaper som går igjen for

kategorien “*leaders*” innen master data management er bruken av AI og ML for automatisering og distribusjon av master data, de støtter flere domener og leverer tjenestene sine i skyen. I tillegg har de strategier som baserer seg på å komme enda lenger med AI og ML. For å sikre datakvalitet integrerer leverandører innsikt for å optimalisere og automatisere master data management og forretningsens *data governance* (Parker et al., 2021).

Virksomheter må være veldig nøye i valg av leverandør av master data hub, og de må velge en leverandør som kjenner bransjen man jobber i. Mange master data management løsninger er spesialiserte på de horisontale kravene man har, men ikke nødvendigvis de vertikale (Beasty, 2008). Yuhanna et al. (2021) definerer de horisontale kravene som leverandørens strategi når det kommer til master data management og de vertikale kravene som leverandørens nåværende produkt. Styrker i en leverandørs strategi vil være at de har *roadmap* for innovasjon, partnerskap, visjon for produktet, *roadmap* for utførelse, teknisk support og profesjonelle tjenester. Styrker til en leverandørs produkt vurderes av *matching*, kobling og oppløsningen til *entities*; datakvalitet, data kontekst, samsvar/sammenstilling, *data governance*, dataintegrasjon, maler til forretningen, distribusjon, avstamning, sikkerhet, *high availability/disaster recovery* (heretter referert til som HA/DR), skalerbarhet, domener og støtte for flere domener.

Det er viktig å huske på at virksomheter velger leverandører som samsvarer med egne ambisjoner om master data management, og ikke velger ut ifra markedsledere. Leverandører som er trukket frem i Yuhanna et al. (2021) er blant annet leverandører som lager løsninger for:

- Omfattende, skybaserte master data management løsninger på tvers av flere domener
- Kunder som støtter SAP-applikasjoner og -tjenester
- Virksomheter som søker bredere ende-til-ende databehandling som inkluderer master data management i kjernen
- Virksomheter som har store og komplekse master data management krav på tvers av hybride skyløsninger
- Virksomheter som spesifikt ser etter en master data management løsning i Azure med tettere integrasjoner

- Skybasert master data management plattform som støtter *governance*, kvalitet og integrasjon på tvers av flere domener.

Det er altså mange faktorer som spiller inn når man skal finne en leverandør som passer virksomheten strategi og *roadmap*.

3. Metode

I dette kapittelet skal vi se på den metodiske tilnærmingen vi har valgt å bruke i vårt forskningsprosjekt. Dette vil vi gjøre gjennom å introdusere utvalget til vår casestudie, argumentere og utdype for hvilken forskningsmetode vi har benyttet oss av, og generelt hvilket forskningsdesign som er grunnlaget for våre datafunn. Vi vil gå nøye gjennom hvordan datainnsamlingen har funnet sted og kommer til slutt til å gjøre en vurdering av disse dataene. Metode handler ifølge Johannessen et al. (2011) om hvordan man skal samle inn, analysere og tolke data. Hellevik (2002) mener man gjennom metodelæren blir presentert for ulike fremgangsmåter, samt drar nytte av tidligere forskeres erfaring, og ikke blir henvist til å lære gjennom egen prøving og feiling. Da motstår man fristelser for å bruke fremgangsmåter som gir de resultatene man ønsker.

3.1 Forskningsdesign

Forskningsdesign handler om alt som er knyttet til en undersøkelse (Johannessen et al., 2011), og vi vil i denne delen begrunne hvilket forskningsdesign vi har valgt for vår studie. For å finne ut hvilket forskningsdesign som passer vår forskning starter vi med å se på forskningsspørsmålene våre, og hvilke praktiske problemer som må løses for å kunne svare på dem.

1. Hva er god datakvalitet?

Datakvalitet kan måles kvantitativt med tanke på antall feil det fører til (eksempelvis prediksjon og accuracy), men da trenger man et datasett. Eller det kan måles kvalitativt (subjektivt eller objektivt) ved å hente inn empiri gjennom intervjuer og spørreundersøkelser. For å hente inn så mye informasjon som mulig om dette, og ikke bare predikere datakvalitet, mener vi forskningsspørsmålet best kan besvares med en kvalitativ metode.

2. Hvilke strategiske, taktiske og operasjonelle valg må en bedrift ta for å ha god master data management?

Med et stort nok utvalg kan dette forskningsspørsmålet løses med en kvantitativ tilnærming. Forskningsspørsmålet lar seg også løse kvalitativt. Siden vi har begrensninger i form av tid og ressurser mener vi dette løses best med en kvalitativ tilnærming fordi utvalget vårt er lite.

3. Kan master data hub være en del av master data lifecycle?

Dette forskningsspørsmålet må løses med en kvalitativ metode, da vi må studere hvordan master data sin livssyklus er hos virksomheter som har en master data hub.

4. Hvilke organisatoriske-, forretningsmessige- og IT-systemer berøres av en master data hub?

Dette forskningsspørsmålet må løses med en kvalitativ metode, da vi gjennom teorien har sett at alle applikasjoner og systemer kan berøres av en master data hub. Vi må finne ut om dette stemmer.

5. Hvilke faktorer påvirker valg av master data hub?

Dette forskningsspørsmålet må løses med en kvalitativ metode. Vi oppfatter at det kan være ulike variabler som påvirker valg av master data hub og hvis forskningsspørsmålet uansett skulle vært løst med en kvantitativ metode, mener vi at det burde vært løst med en kvalitativ metode i første omgang for å innhente de ulike variablene.

6. Hvem leverer denne type teknologi, og hvilke tjenester/kvaliteter er bygget inn i master data hub?

Dette må besvares med en kvalitativ metode. Fordi vi må få informasjon om hvilke teknologier som ligger bak en master data hub.

Når vi ser på de praktiske problemene som er knyttet til våre forskningsspørsmål mener vi at de trolig løses best med en kvalitativ tilnærming. I starten, da vi kun hadde en problemstilling, hadde vi en antakelse om at forskningen vår kom til å ha en kvantitativ tilnærming, som kunne måle datakvalitet. Etter hvert som forskningsprosjektet pågikk så vi fort at vi hadde manglende kunnskap om master data hub, og innhenting av informasjon ble derfor vesentlig for å kunne komme noen vei. En kvalitativ tilnærming kjennetegner fraværet av en analytisk hovedretning, og det er viktig at man velger hvordan data skal samles inn tidlig i prosessen (Johannessen et al., 2011). Ulike emner forskes på forskjellig og det er viktig å ha et konkret forskningsdesign. For å kunne besvare forskningsspørsmålene våre på best mulig måte, mente vi at en casestudie ville være riktig design for vår forskning. En casestudie undersøker ett eller flere tilfeller som studeres inngående. Forskningsobjektet vårt er et master data management verktøy, hvor vi studerer tre ulike kontekster ved hjelp av informanter. I tillegg er dette forskningsprosjektet over en begrenset periode som igjen setter begrensninger for forskningsdesign.

3.2 Casestudie

En klassisk casestudie består ifølge Yin (2013) av en grundig undersøkelse av et spesifikt og komplekst fenomen, «en case», som er satt innenfor en virkelig kontekst. Det finnes mange tilnærminger til en casestudie, og vi har valgt å bruke Yin (2007) sin metodiske tilnærming. Caseundersøkelser kan gjennomføres med å bruke forskjellige metode for å skaffe seg mye og detaljert data for å gi en mest mulig inngående beskrivelse av casen. Han mener det er 5 komponenter som er spesielt viktig i gjennomføringen av en casestudie; problemstilling, teoretiske antakelser, analyseenheter, den logiske sammenhengen mellom data og antakelsene, og kriterier for å tolke funnene. Vårt design er en forklarende casestudie. Johannessen et al. (2011) skriver at kjernesporsmål for en forklarende undersøkelse er hvilke faktorer x som er årsaken til y , eller hvilke fenomener y som er konsekvenser av x . Metoden er altså årsaksforklarende og skal se på hvilke omstendigheter som gjør at x utløser y . I vårt tilfelle er x = master data hub og y = datakvalitet.

Gjennom litteraturstudien gjør forskeren seg noen antakelser etter å ha stilt seg noen grunnleggende spørsmål. Antakelsen kan være hva man tror er svaret. Når problemstillingen er definert er det naturlig å avgrense enheten som skal studeres. Casen kan være et individ, et program, institusjon osv. Programmet er analyseenheten som er knyttet til spørsmålene. Når man skal se på den logiske sammenhengen mellom data og antakelsene opererer man med to analysestrategier; teoretiske antakelser som er teoristyrte, og en forklarende casestudie. Teoretiske antakelser er å foretrekke, og man bruker kun beskrivende casestudie dersom man ikke har noen antakelser på forhånd. Yin (2007) mener man skal tolke funnene opp mot allerede eksisterende teori på området. Det er viktig at man har en foreløpig teori før selve datainnsamlingen, dette er de fire første komponentene i casestudien. Forskeren skal relatere funnene sine til eksisterende teori, og i rapporteringen enten beholde, modifisere og utvikle, eller bygge helt ny teori.

3.3 Utvalg

Det å velge ut hvem som skal delta i forskning er en vesentlig del av forskningsprosjektet. Her mener Johannessen et al. (2011) det er viktig at man tenker på utvalgsstørrelse, utvalgsstrategi og rekruttering av deltakere. Når det kommer til utvalgsstørrelsen må man ta hensyn til at man skal hente ut mye informasjon fra et begrenset antall personer som betegnes som informanter. En informant er et intervjuobjekt, og ikke en deltaker, i en studie dersom det gjennomføres gjentatte intervjuer med den spesifikke personen (Yin, 2014).

Hvor mange informanter man trenger til sin casestudie er avhengig av problemstillingen og måter data samles inn på. Mange forskere mener at det under studiet bør gjennomføres intervju helt til forskeren ikke får mer informasjon fra sine informanter (Johannessen et al., 2011). Videre skriver de at det er begrenset med tid og penger i et studentprosjekt, noe som gjerne resulterer i 10 eller færre intervju. Dersom man merker man ikke får tilstrekkelig med informasjon fra sine informanter, kan man søke informasjon fra andre.

Det er også viktig at man har en strategi for rekrutteringen av informanter. Informantene i vår studie ble ikke rekruttert tilfeldig, da vi skulle ha mest mulig informasjon om master data hub og virksomhetsarkitektur. Rekrutteringen hadde i utgangspunktet én målgruppe som bestod av kunder som trenger eller bruker master data hub. Dette var for å studere påstanden om at bruk av master data management kan redusere problemer knyttet til datakvalitet (Hikmawati et al., 2021; Lerche, 2014). Etter hvert i forskningsprosessen fant vi ut at vi trengte enda en målgruppe som bestod av de som leverer en master data hub. Da fikk vi i tillegg studert hvilken teknologi og arkitektur som skal bidra til å sikre god datakvalitet. Dette kalles strategisk utvelgelse, eller *purposeful sampling*, av informanter i metodelitteraturen (Patton, 1990). Ved strategisk utvelgelse tenker forsker gjennom hvilken målgruppe som skal delta for å få samlet nødvendig data. Det neste steget er å velge ut personer fra målgruppen som skal delta i studien. Når det kommer til antall informanter som deltar i en studie skriver Johannessen et al. (2011) at det er hensiktsmessighet som styrer dette, og ikke representativitet. Vi tror flere informanter innenfor målgruppene våre kunne gitt oss bredere og mer informasjon. I tillegg kunne det økt bekreftbarheten til studien. Begrensninger i tid og ressurser har hindret dette, men vi er fornøyde med å ha informanter innenfor våre målgrupper som dekker hensiktsmessigheten.

3.4 Datainnsamling

For at leseren skal kunne følge og vurdere forskningsprosessen mener Johannessen et al. (2011) at det er viktig at det legges vekt på å beskrive alle beslutninger gjennom hele forskningsprosessen. Dette oppnår man gjennom å være selvkritisk til hvordan prosjektet er gjennomført og kommentere hvordan fordommer og oppfatninger kan påvirke fortolkningen og tilnærmingen til prosjektet. “Det er viktig at funnene i kvalitative undersøkelser er et resultat av forskningen, og ikke et resultat av forskerens subjektive holdninger” (Johannessen et al., 2011, s. 249). Vår metode for innsamling av data er basert på en casestudie med et caseobjekt og flere informanter. Caseobjektet er master data hub. Informantene representerer tre forskjellige kontekster, der én er bruker av master data hub og har gitt oss innsikt i hvordan denne fungerer, én er leverandør av master data hub som en del av en tjenesteplattform og har gitt oss innsikt i teknologien og arkitekturen som ligger bak sitt

produkt, og den siste har gitt oss innsikt i hvordan man går fra kjernesystem til master data hub.

I dette prosjektet har vi ivaretatt vårt etiske og juridiske ansvar. Prosjektet ivaretar informantenes privatliv og rett til å trekke seg fra prosjektet, dette gjennom godkjenning fra Norsk senter for forskningsdata (heretter referert til som NSD). For å få studiet godkjent av NSD er forskerne lovpålagt å laste opp informasjonsskriv (se vedlegg 3). Informantene og deres virksomheter er anonymisert i denne studien. Data vi har etterspurt i målgruppen “kunde” vil være sensitiv i form av teknologi- og forretningsstrategi, pågående prosjekt og datakvalitet. Informanten for målgruppen “leverandør” presenterte i stor grad sine egne meninger om sitt produkt og sin virksomhet, og det var derfor hensiktsmessig å anonymisere denne målgruppen også. “Samtykke som deltaker i en undersøkelse skal være en frivillig, uttrykkelig og informert erklæring fra de opplysningene gjelder om at den godtar behandling av opplysninger om seg selv og virksomheten” (Johannessen et al., 2011, s. 100). Samtykket kan gis muntlig eller skriftlig, elektronisk eller på papir. Samtykke til deltakelse fra alle informantene ble gjort skriftlig, og samtykke til opptak av samtaler på Microsoft Teams ble gjort muntlig. Informantene har også fått lov til å lese gjennom og godkjenne det vi har skrevet i datainnsamlingen som omhandler dem selv.

Litteraturstudien og forskningsspørsmålene våre resulterte i at vi hadde behov for to målgrupper for vår rekruttering til informanter. Gjennom litteraturstudien vår hadde vi en antakelse om at informanter i store virksomheter kunne gi oss best innsikt i håndtering av master data, master data management og datakvalitet. Vi tok derfor kontakt med flere virksomheter der alle til å begynne med var interesserte i å samarbeide om vår studie. På grunn av frafall underveis og manglende oppfølging fra enkelte virksomheter, endte vi opp med utvalget vårt på tre informanter.

Fordelt på de tre informantene er det utført tre semistrukturert intervjuer á 1,5 timer. Kvalitative intervjuer kan være mer eller mindre strukturert, og Johannessen et al. (2011) viser til tre ulike strukturerer på intervju; ustrukturert, semistrukturert og strukturert intervju. Intervju er ifølge Yin (2014) en av de viktigste kildene for bevis i casestudier.

Semistrukturerte intervju vil si at vi har hatt en strukturert form med en overordnet intervjuguide som utgangspunkt, men spørsmål, temaer og rekkefølge kan variere underveis. Selv om man da følger konsekvent spørsmål og tema som er satt i forkant av intervjuet, vil samtalen være mer flytende enn rigide (Rubin & Rubin, 2011). Dette har lagt til rette for at vi kunne stille oppfølgingsspørsmål direkte, hatt interessante diskusjoner og fått informasjon om tema og spørsmål vi ikke hadde tenkt ut på forhånd.

I og med at utvalget vårt er lite, så vi stor verdi av å ha semistrukturerte intervju slik at vi kunne hente inn så mye informasjon som mulig. Intervjuene vil ligne mer på veiledende samtaler i stedet for strukturerte samtaler. En intervjuguide er ikke et spørreskjema, men en liste over temaer og spørsmål som skal gjennomgås i selve intervjuet (Johannessen et al., 2011, s.147). På denne måten unngikk vi også at relasjonen mellom oss og informantene var avgjørende for hvilken informasjon som kom frem i intervjuene, noe som kan være et problem ved ustrukturerte intervjuer, fordi det allerede var satt en agenda. Vi merket raskt at dette var en metode som fungerte fint med våre informanter. Vi opplevde at informantene også likte det semistrukturerte formatet, og at det var svært givende å kunne stille spørsmål og sparre med dem fortløpende i forskningsprosessen. Det var heller aldri noe problem å få intervju og samtaler med virksomhetene vi endte opp å samarbeide med. Vår opplevelse av denne strukturen er at vi klarte å hente ut mer informasjon enn vi ville fått gjennom strukturerte intervju eller spørreundersøkelse.

3.4.1 Intervju

Innledningsvis ble alle informanter gjort kjent med problemstillingen og forskningsspørsmålene våre elektronisk, og vi stilte åpent spørsmål om hva de mente de kunne bidra med. Deretter ble det satt opp intervjuguider basert på våre antakelser fra litteraturstudien (se vedlegg 4). I litteraturstudien kom det fram at det mest sannsynlig var en korrelasjon mellom datakvalitet og AI/ML. Vi hadde derfor en antakelse om at leverandører som leverer master data management med innebygd AI/ML vil kunne sikre god datakvalitet. I målgruppen for leverandør var intervjuguiden (se vedlegg 5) derfor spesifikt rettet mot master data hub, hvilken teknologi de har bygget inn i sin hub og hva som gjør at kundene velger dem. Yin (2014) mener det er viktig at forsker ikke blir for avhengig av en nøkkelinformant

på grunn av påvirkningen informanten kan ha. Dette må håndteres ved å stole på andre beviskilder, samt det å bekrefte innsikt og søke etter motstridende bevis.

Vi hadde også en antakelse om at det virksomheter som er modne innen master data management er de som velger å implementere verktøy som master data hub. Intervjuguiden til vår målgruppe for kunder bestod derfor av Spruit & Pietzka (2015) sin MD3M modell. Denne modellen hjelper oss med å belyse målene på datakvalitet. Vi hadde også en antakelse om at virksomheter som sliter med dårlig datakvalitet ville være mer interessert i å implementere en form for master data management, og stilte derfor spørsmål til informantene om deres opplevde datakvalitet i virksomheten. På denne måten fikk vi innsikt i datakvaliteten til en virksomhet med master data hub, og en uten. Dimensjonene for datakvalitet som går igjen mest er nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet (Christen, 2012; DAMA, 2013; Hikmawati et al., 2021). I intervju med målgruppen “kunde” kom det frem at det var viktig å ha et forhold til hvilke master data man flytter over til en master data hub. Disse utsagnene ville vi ha bekreftelse fra målgruppen “leverandør”, om at dette også var tilfellet blant deres kunder. I etterkant av intervjuene satte vi av tid til å gå gjennom intervjuene for å finne ut om vi hadde fått den informasjonen vi søkte i forkant.

3.4.2 Elektronisk innsamling

Etter de første intervjuene fikk vi beskjed fra samtlige informanter at de kunne stille til nye intervju ved behov. I tillegg kunne de svare på oppfølgingsspørsmål elektronisk. Videre var kommunikasjonen vår elektronisk basert på oppfølgingsspørsmål fra litteraturstudien og informasjon fra målgruppene. Fordelen med dette er at man kan sende intervju til flere informanter samtidig (Johannessen et al., 2011). Ryen (2002) mener at dette kan by på utfordringer med å skape tillit når kommunikasjonen utelukkende er skriftlig, og at mail ikke gir kontekstinformasjon som ikke-verbal kommunikasjon. Videre sier han at informanter, lett kan miste interessen om de blir neddyngtet med mail der forskeren ber om inngående svar. Vi opplevde ikke at dette var tilfellet i vår datainnsamling, da vi fikk rask respons fra alle informanter på vår elektroniske kommunikasjon (se vedlegg 5). Dette var svært nyttig for oss

i vår forskning da vi fikk mulighet til å sparre med informantene våre om funn i litteraturstudien fortløpende. Den elektroniske kommunikasjonen var mer strukturert i den forstand at spørsmålene var satt. Vi var påpasselige med å stille åpne spørsmål som ga rom for utdypende svar fra informantene.

Etter en gjennomgang med målgruppen “kunde” sendte vi mail til målgruppe “leverandør” for å starte diskusjon rundt hvilke master data det er gunstig å tilgjengeliggjøre i en master data hub. I målgruppen “leverandør” var vi interessert i å komme i kontakt med noen som hadde vært med å sette opp tjenesteplattformen til en av kundene deres. Planen var da å stille de samme spørsmålene. I tillegg ønsket vi å få svar på om master data management var en del av kundens strategi, hvilke faktorer påvirket valget av master data hub, modenhet innen master data, hvor mye forarbeid krevde implementeringen, hvilke master data de la i master data hub, bruker de master data hub til rapportering og analyser, hvilken *master data governance* de har, og om de er fornøyde med den nye arkitekturen sin. Dessverre fikk vi ikke kontakt med denne kunden eller de som hadde jobbet med prosjektet.

Resterende samarbeid med målgruppen “leverandør” gikk ut på en teknologisk gjennomgang av deres tjenesteplattform, som er grunnlaget for deres master data hub. Vi ønsket mer innsikt i integrasjoner og integrasjonsmønster, hvilke fordeler de ser ved å levere en tjenesteplattform som inkluderer både transaksjonsdata og master data, hvordan foregår datamodellering i plattformen og er formatet/strukturen tilpasningsdyktig til kundene, samt hvor formateringen skjer. Etter den tekniske gjennomgangen sendte vi oppfølgingsspørsmål knyttet til Yuhanna et al. (2021) sin master data management rapport, der det kommer frem at kunder bør velge leverandører som:

- Støtter flere domener (eksempelvis kunde og leverandører) via en enkelt instans
- Forenkler master data management gjennom AI / ML og støtter innebygde analyser
- Kan sikre datakvalitet gjennom profesjonelle tjeneste- og teknologileverandører (*Accenture, BearingPoint, Capgemini, Cognizant, Deloitte, Infosys, Wipro*).
- Har et *roadmap* med like store ambisjoner innen master data management

- Har et *roadmap* som fokuserer på AI/ML automatisering, utvidet integrasjon til nye datakilder, aktivert datasikkerhet som standard, integrasjon med andre verktøy for databehandling og –teknologier

Vi fulgte også opp målgruppen “kunde” elektronisk. Kunden som allerede har implementert master data hub fikk spørsmål om valg av leverandør, hvilken virksomhetsarkitektur de hadde før implementering av master data hub, og hvordan de opplever datakvaliteten før og nå. Målgruppen fikk også spørsmål om barrierer knyttet til kvalitet i master data. Vi hadde en antakelse om at (English, 1999; Haug & Arlbjørn, 2011; Lee et al., 2006; Umar et al., 1999; Xu et al., 2002) sine barrierer til god kvalitet i master data stemmer (se vedlegg 7).

Fra kunden som ikke har implementert master data hub fikk vi uoppfordret en gjennomgang av deres strategi for master data management. Dette var veldig nyttig, og vi ville derfor vite om kunde som har implementert master data hub har noen fremtidig strategi for master data management. (Hikmawati et al., 2021; Vilminko-Heikkinen & Pekkola, 2013) mener dette er et viktig steg mot å ha suksessfull master data management.

3.4.3 Analyse av datainnsamling

Analyse av datainnsamling i en casestudie kan ifølge Yin (2007) være basert på teoretiske antakelser eller en beskrivende casestudie. Vår analysestrategi er teoristyrte basert på antakelser fra vår litteraturstudie. I selve analysen skal forskeren danne seg et helhetsinntrykk for deretter å plukke ut viktige funn i undersøkelsens hovedområder. I en slik analyse vil kriteriene for å tolke datafunn være at de tolkes opp mot allerede eksisterende teori, basert på forskerens antakelser. Teksten omskrives til et profesjonelt språk der intervjuene sammenfattes, og det utledes en generell struktur basert på alle intervjuene som er gjort. I rapporteringen kan man beholde eksisterende, modifisere og utvikle, eller bygge ny teori.

Intervjuene og den elektroniske innsamlingen ble i sin helhet transkribert, og utgjorde til sammen 27 sider med informasjon. Vi benyttet oss av Microsoft Teams ved lydopptak og

opplevde at kvaliteten på lyden var god. Dette gjorde transkriberingen lett å gjennomføre. Våre forskningsspørsmål og litteraturstudie var førende for datainnsamlingen, og dermed var mye av arbeidet relatert til koding allerede gjort. Det var derfor enkelt for oss å selekere hvilke data som var relevant til å besvare vår problemstilling. Data som ble ekskludert var informasjon om tredjeparter som ikke lot seg bekrefte, og diskusjon rundt tema for å øke egen forståelse. Det vi brukte mest tid på var å anonymisere informasjonen vi hadde fått, og skrive den om til en profesjonell tekst.

3.5 Vurdering av data

Forskning basert på casestudie har ifølge Yin (2014) blitt kritisert for at forskere unnlater å utvikle et operativt sett med tiltak og at subjektive vurderinger ofte har en tendens til å bekrefte forskernes forutinntatte meninger. Han viser til tre teknikker som brukes når man skal bedømme den konstruerte validiteten ved casestudier; flere bevis kilder, kjede av bevis og at utkastet til studien blir gjennomgått av nøkkelpersoner. I tillegg må data vurderes med grunnlag i reliabilitet, validitet, overførbarhet og bekræftbarhet.

3.5.1 Reliabilitet

Johannessen et al. (2011) mener det er lite hensiktsmessige å teste datas reliabilitet, da man normalt i kvalitative undersøkelser benytter ustrukturerte datainnsamlingsteknikker og at observasjoner er kontekstavhengige. Ustrukturerte data gjør det vanskelig å sammenligne data man har innhentet og å duplisere forskningen som er gjort. I tillegg vil det at vi har anonymisert informantene og deres virksomheter føre til at andre forskere ikke vil kunne gjennomføre samme studien som vi har utført.

3.5.2 Validitet

Validitet handler om hvorvidt undersøkelsen er egnet til å gi gyldige svar på problemstillingen eller forskningsspørsmålene som er satt for undersøkelsen (Johannessen et

al., 2011). Dette mener vi er oppnådd da intervjuene har basert seg på antakelser fra teorien. Validitet i kvalitative undersøkelser dreier seg om i hvilken grad forskernes fremgangsmåte og funn reflekterer formålet med studien og representerer virkeligheten (Johannessen et al., 2011, s. 244). Videre skriver de at teoretiske funn, klarhet i begreper og metodiske vurderinger inngår i slike diskusjoner. En redegjørelse av validitet skal ifølge Postholm (2010) inneholde opplysninger om metoder som er brukt i datainnsamlingen, intervjumetoder og analyse av transkripsjoner. Analyse av transkripsjoner skjer gjennom datareduksjon og hvordan data er kategorisert.

All informasjon siles gjennom forskeren, og det er derfor viktig å være klar over de ulike skjevhetene som kan påvirke vår forskning. Vi mener skjevheter innen intervju, frafall og kognitivitet kan treffe vår forskning. Johannessen et al. (2011) har med grunnlag i forskningen til (Cohen et al.2008; Moyles 2002; Robson 2002; Shaugnessy et al.2003; Wilkinson 2000) definert intervjuksjevhet, frafallsskjevheter og kognitiv skjevhet. Intervjuksjevhet, også kalt intervju-effekt, handler om at intervjueren påvirker informanten enten bevisst eller ubevisst gjennom spørsmål, kroppsspråk, antrekk og holdninger etc. (Johannessen et al. 2011, s. 245). Forskere kan styrke troverdigheten til informantene sine ved å formidle og få bekreftet resultatene. "Frafallsskjevheter er en konsekvens av at enkelte som i utgangspunktet er rekruttert til en studie ikke fullfører. Dette får spesielt store konsekvenser i kvalitative undersøkelser" (Johannessen et al., 2011, s.246). Vi ville heller hente så mye informasjon som mulig, enn å bruke tid på å bekrefte informasjonen ved å hente inn flere informanter. Vi mener hensiktsmessigheten er til stede i vår forskning, men flere informanter kunne styrket studien. Vi har nok informasjon til å svare på forskningsspørsmålene våre. Videre skriver de at kognitiv skjevhet kan være en tendens til å fortelle om egne positive trekk. Vi mener dette gjelder spesielt målgruppen "leverandør" som promoterer sitt produkt. Det kan også være at målgruppen "kunde" ikke ønsker å formidle virkelig status på datakvalitet i sin virksomhet.

Intervjuene ble tatt opp i sin helhet, med video og lyd, noe som resulterte i at vi som forskere kunne fokusere på selve intervjuet og ikke på å notere underveis. På denne måten fikk vi en fullstendig dialog da alle parter var investert i samtalen, i tillegg til at ingen informasjon rundt

intervjuet “gikk til spille”. De semistrukturerte intervjuene er tatt opp, med video og lyd, etter samtykke fra informantene. Vi har hørt gjennom intervjuene og transkribert hver for oss, og på denne måten satt sammen våre tolkninger av intervjuene. Fordelen med dette er at vi kan diskutere funnene og se om vi har tolket intervjuene likt. Med denne metoden kan hukommelseskjevhet elimineres. I kapitlet datafunn har vi skrevet datafunnene som fullstendig tekst. Årsaken til dette er at intervjuene er semistrukturerte og mye av dialogen består av et muntlig språk. Muntlig språk hører ikke hjemme i en akademisk tekst, men vi har valgt å ta med direkte sitater for å øke troverdigheten i våre datafunn. Johannessen et al. (2011) mener forskere bør lage seg en teknikk for å skille mellom egne tolkninger, notater og sitater fra informanten.

Vi har skrevet under konfidensialitetserklæring med den ene virksomheten. Dette mener vi har vært med på å legge til rette for tillit, og at informanten derfor har følt seg komfortabel med å dele konfidensiell informasjon som anonymiseres i denne forskningen.

3.5.3 Overførbarhet

Johannessen et al. (2011) mener at overførbarhet handler om hvorvidt man har lyktes med å etablere beskrivelser, begreper, fortolkninger og forklaringer som skal være nyttige for andre områder enn hva som faktisk studeres. Altså en generalisering av det man forsker på. Vi mener at vår forskning ikke gir grunnlag for generalisering, og dermed blir temaet overførbarhet overflødig. Punktet er med for å understreke dette. Forskningen lar seg ikke generalisere da det er mange variabler for datakvalitet, arkitektur, størrelse og behov for de ulike virksomhetene. Forutsetninger for virksomhetene er såpass ulike at generalisering ikke er hensiktsmessig.

Lincoln & Guba (1985) mener at kvalitative undersøkelser må vurderes på en annen måte enn kvantitative, og at det da er viktig at man ser på overførbarhet og bekreftbarhet som mål på kvaliteten i kvalitative undersøkelser. Vi velger derfor å fokusere på den interne validiteten i vår forskning. I den forbindelse er det vår oppgave å sørge for gode beskrivelser av

målgruppene slik at leser kan vurdere om de ønsker å overføre våre funn til sin egen forskning.

3.5.4 Bekreftbarhet

For at leseren skal kunne følge og vurdere forskningsprosessen mener Johannessen et al. (2011) at det er viktig at det legges vekt på å beskrive alle beslutninger som er tatt gjennom hele forskningsprosessen. Dette oppnår man gjennom å være selvkritisk til hvordan prosjektet er gjennomført, og kommentere hvordan fordommer og oppfatninger kan påvirke fortolkningen og tilnærmingen til prosjektet. De poengterer også at forskere kan styrke bekræftbarheten ved å støtte seg på annen litteratur og få bekræftelse fra informantene at fortolkningene som er gjort i studien støttes. Vi har forsøkt å søke flere beviskilder gjennom å bruke informasjon vi har hentet inn i målgruppen “kunde” inn i intervju og diskusjon med målgruppen “leverandør”, og motsatt. Kjede av bevis har vært vanskelig å ta stilling til, da vi opplever at datakvalitet og strategi knyttet til master data management varierer fra virksomhet til virksomhet. Vi har søkt bekræftelse på våre funn ved å kryssjekke informasjonen vi har samlet inn fra én informant med de andre informantene i studien vår. Våre informanter har fått muligheten til å gjennomgå datainnsamlingen vi har gjort om dem, og komme med kommentarer og korreksjoner der de måtte ønske. All informasjonen vi har samlet inn bunner i litteraturstudien vår, og antakelser fra denne. Litteraturstudien blir derfor bekreftet eller avkreftet av vår empiri.

Vi har opplevd frafallsskjevhet, og mangel på tid og ressurser setter begrensninger for mulige samarbeid. Frafallet av samarbeidspartner skjedde i februar 2022, og manglende tid og ressurser gjorde det utfordrende å hente inn informasjon fra nye samarbeid. Dette har ført til begrenset bekræftelse i studien vår.

Forskeres egne oppfatninger kan være med på påvirke tolkning og tilnærming til studier. Vi mener manglende teknisk erfaring hos oss, om blant annet virksomhetsarkitektur og master data management, har ført til at vi i stor grad er avhengig av erfaringen vi har tilegnet oss

gjennom litteraturstudien og datainnsamling. På tross av at informantene har lest gjennom og godkjent informasjon de har gitt oss, kan det hende bransjespesifikke begreper og uttrykk mistolkes av oss.

4. Datafunn

I dette kapitlet skal vi overordnet presentere vårt caseobjekt og våre informanter. Dette skal bidra til å skape kontekst for datafunnene som blir presentert i kapitlet. Kapitlet er delt inn i delkapittel etter informantene, og vi starter med å presentere målgruppen “Leverandør” slik at leser får teknisk innsikt i hva en master data hub kan være. Deretter presenterer vi våre funn fra målgruppen “Kunde”. Terminologien for system er ulike i virksomhetene, og vi har valgt å gjengi informantenes egne ord.

4.1 Presentasjon av caseobjekt og informanter

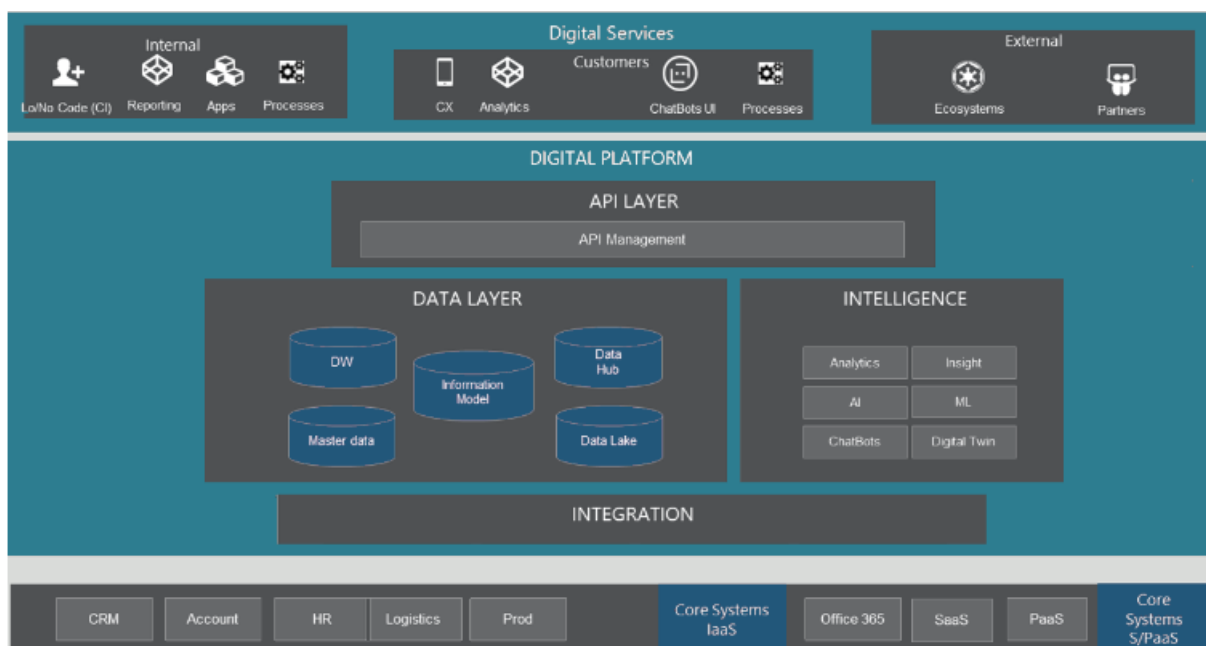
I teorikapitlet har vi gjennom en litteraturstudie gjort en antakelse om at virksomheter som velger å ta i bruk master data hub er modne innen master data management. Det vil si at de har et forhold til hva som er master data i sin virksomhet, hvor master data er lagret og hvordan de er strukturert. I tillegg er det viktig at virksomhetene har et forhold til kvalitet i master data, hvordan kvaliteten påvirker forretningen, hvor kilden til dårlig kvalitet ligger og hvordan de forbedrer den. Vi har derfor hentet inn data fra to virksomheter der den ene har implementert master data hub og den andre ønsker å implementere master data hub. En master data hub kan være både sentral og desentral. En sentral master data hub eier, oppretter, vedlikeholder og sletter/arkiverer master data. Master data vil i en slik løsning alltid opprettes i huben, noe som sikrer at data opprettes likt og er unik. Ved å samle master data sentralt sikrer virksomheten at man overholder forhåndsdefinerte prosesser, men dette går på bekostning av fleksibiliteten. En desentral master data hub distribuerer master data som er opprettet og vedlikeholdt i kjernesystemene. Data opprettes, lagres og vedlikeholdes desentralt med denne tilnærmingen. Endringer som er gjort i master data vil umiddelbart være tilgjengelig når system etterspør data.

Virksomheten som ønsker å implementere master data hub har vi valgt å kalle “Bedrift 1”, og den andre kaller vi “Bedrift 2”. Bedrift 1 er en stor virksomhet som opererer i ulike marked, der arkitekturen består av flere kjernesystem som en konsekvens av dette. Det samme gjelder Bedrift 2, men denne virksomheten har besluttet å tilgjengeliggjøre sine master data i en master data hub, i stedet for å distribuere data fra sitt fagsystem. I tillegg har vi hentet inn data fra virksomheten vi har valgt å kalle “Leverandør”. Leverandør tilbyr en

tjenesteplattform til sine kunder. Teknologien i denne plattformen danner grunnlaget for deres master data hub, som skal tilrettelegge for distribusjon av master data til de digitale tjenestene kunden måtte ønske å benytte seg av.

4.2 Leverandør

For å gjøre datafunnene i denne delen av oppgaven mer forståelig for leser presenterer vi under vår egen, forenklete arkitektur av tjenesteplattformen til kilden vår.



Figur 8: Arkitekturen til Leverandør sin tjenesteplattform (egen figur)

Hovedessensen i Leverandør sin tjenesteplattform er API⁸-laget. Her tilgjengeliggjør de funksjonalitet og data til de digitale tjenestene som ligger på toppen. De digitale tjenestene består av spesifikke formål eller applikasjoner som skal integreres i tjenesteplattformen. Datalaget består av ulike typer teknologier for lagring av data som skal tilgjengeliggjøre data på en enkel måte på tvers av virksomhetens applikasjoner, ved hjelp av API-laget. I datalaget ligger det altså persisterte data⁹ til ulike formål. Master data huben er designet for å lagre master data. Her ser vi at Leverandør har både master data og transaksjonsdata, datavarehus

⁸ API står for *Application Programming Interface*, og er en kode for utveksling av data mellom system

⁹ Persisterte data er data som forkastes av databasen etter bruk

og data lake, i sin tjenesteplattform. Dette mener informanten er en fordel for kundene, da de lett kan sammenstille data til ulike formål, som analyser.

Integrasjonslaget ekstraherer hovedsakelig data fra kjernesystem til datalaget, enten i *real time* eller via pakker (batch). Her finner man et sett med integrasjonsmekanismer, som er leverandørens egne produkter, og disse implementeres etter kundens behov. Her ligger det også en logging- og monitoreringspakke som gjør det mulig å følge data gjennom arkitekturen (avstamning). Dette legger til rette for at kunden kan få tilbakemeldinger om hvor data stopper i *workflow*, dersom det skulle være tilfellet. I integrasjonslaget ligger også det Leverandør kaller for “Akseleratoren”. Denne inneholder script som setter opp API-management, master data hub og noen utvalgte standardkomponenter ved implementering av tjenesten hos kunden. På denne måten vil implementeringen gå raskere.

I bunnen vil kundens kjernesystem ligge. Det er disse som refereres til som siloer. Kjernesystem kobles sammen med tjenesteplattformen gjennom et sett med protokoller eller adaptere, avhengig av hvordan systemene snakker sammen. For eksempel vil nye CRM-løsninger kunne koble seg rett opp på API’er, men gamle ERP-løsninger må koble seg til databasen. Hvert kjernesystem har eller kan ha sin egen datamodell, og bestemmer selv hva som skal defineres som konsistente data. I store systemer er datamodellene gjerne svært komplekse. “Når dette er tilfellet vil kunden kunne modellere en forenklet datamodell av kjernesystemet i datalaget, som dermed er tilpasset stammespråket kunden benytter seg av i sine system”. Kunden må altså selv modellere sitt perspektiv på master data, da Leverandør ikke leverer noen modeller eller funksjonaliteter for dette. Informanten mener dette er noe som gjør tjenesteplattformen tilpasningsdyktig.

Leverandør sin tilnærming til datamodellering er at dette utføres i datalaget slik at det blir enkelt for de digitale tjenestene å benytte seg av master data. Det er integrasjonslaget som transformerer datamodellen fra kjernesystem, til en noe enklere og mer generisk datamodell i master data hub. Informasjonslaget har dermed som oppgave å mappe data fra en semantikk til en annen. Informasjonsmodellen er plassert i midten for å illustrere at kunden selv er nødt til å modellere og ha et forhold til, hvordan man modellerer data. Det vil si at kunde selv må

modellere kapabiliteter som *cleansing* og duplikatsjekker gjennom koding i integrasjonslaget, eller de kan legge til mikro-tjenester som gjør dette for dem. Det som er sentralt i informasjonsmodellering er at alt skal henge sammen; man skal ha de samme begrepene på master data, transaksjonsdata og rapportering. Kunden skal hele tiden kunne kjenne seg igjen i terminologien uavhengig av hvor man er i dataflyten. Master data hub kan også være en toveis integrasjon der CRM distribuerer data til master data hub, samtidig som en applikasjon i tjenestelaget gjør det samme. “Dette er komplisert og skaper konflikter med hvilket system det er som sitter på *single point of truth*”. Dette kan løses av master data management løsninger, som kan si noe om hvilket system som sitter med preferanse data, samt logikk for å kunne beregne sannsynligheten for hvilket felt som stemmer. Dette oppnås ved å bruke AI og ML.

“Datakvaliteten skal sikres gjennom *system of entry*, som vil si at problemer fanges opp der data kommer inn”. Det er kjernesystemet som eier og vedlikeholder master data som huben lagrer og distribuerer. Derfor er det heller ikke innebygd noe master data management i huben. Dette mener informanten kunne vært aktuelt dersom man har flere *system of entries*, for eksempel flere CRM-system etter oppkjøp som alle skal sende sine master data til huben. “Master data huben vil da fort kunne inneholde duplikater som kan kreve funksjonalitet som gjenkjenner duplikater, og som kanskje anerkjenner en *golden record* på tvers av CRM-system”.

Leverandør av master data hub ser på tjenesten som et element i sin helhetlige tjenesteplattform, og at denne er en naturlig del av en virksomhetsarkitektur som har et plattformperspektiv. Deres master data hub står seg ikke alene, og kan ikke isolert sett fungere som komponent i en virksomhetsarkitektur. Virksomheten posisjonerer sin master data hub på lik linje som en integrasjonskomponent, AI/ML-funksjonalitet, og hvordan man organiserer data i datamodeller og datadefinisjoner. Dette henger tett sammen med hvordan man får tak i data enten det er via hybride koblinger, skyen, *on-premise*¹⁰ eller *streams*¹¹.

¹⁰ Programvare som er installert og kjører på datamaskiner i virksomhetens eget IT-miljø

¹¹ Streams er kontinuerlig tilføring av data i real time

Leverandør skiller ikke mellom master data og transaksjonsdata i sin data hub, men trekker frem at kundene gjerne starter med å flytte master data ut i huben som en overgang til tjenesteplattformen. “Kjernesystemene til kunden forer datahubene og data lake med *real time* data, som benytter seg av data i forkant av analyseverktøyene som ligger i datavarehuset”. Den tekniske implementasjonen som Leverandør tilbyr i sin arkitektur legger til rette for å kunne supplere data (*internet-of-things; IOT*), ekstrahere data og legge til rette for funksjonalitet som skal sikre datakvalitet (integrasjon og datamodeller). I integrasjonslaget kan virksomheter oppdage duplikater eller dårlig fyllingsgrad i master data, slik at de må applisere master data med ny data gjennom funksjonaliteter som ligger i *Orchestration (Akseleratoren)*. Det er også mulig å koble på AI og ML for å sikre enda bedre kvalitet, men dette må kunden selv orkestre, og er ikke noe Leverandør leverer med sin tjenesteplattform.

Leverandør har to klare formål med sin master data hub; den skal gi muligheten til å effektivisere og skape nye arbeidsprosesser, og den skal tilgjengeliggjøre data for analyser og beslutningsstøtte i *real time*. Det å effektivisere og skape nye arbeidsprosesser, oppnår de gjennom å tilgjengeliggjøre data og funksjonalitet for nye digitale tjenester, ved gjenbruk på tvers av eksisterende IT-landskap. Tilgjengeliggjøring av data for analyser og beslutningsstøtte skal oppnås ved å ha fokus på informasjonen som ligger i data og sammenhengen mellom dem, samt automatisering av dataflyt og logikk. Videre beskriver informanten at master data hub inngår som en komponent i en moderne integrasjonsarkitektur, hvor man ønsker å gå vekk fra punkt til punkt og tette koblinger i en låst arkitektur, og over til en mer løst koblet, fleksibel arkitektur.

Informanten tror veldig på at man kan konsolidere master data og transaksjonsdata i en tjenesteplattform. Det skal ikke være nødvendig å hente samme data flere ganger, selv om de anvendes ulikt. På denne måten skal Leverandør hjelpe kundene sine med å tenke mer helhetlig gjennom å fjerne løse koblinger i sin arkitektur, og øke innovasjonstakten og selvbetjeningsgraden. Tjenesteplattformen legger til rette for å publisere og tilgjengeliggjøre informasjonen som ligger i kundens siloer, gjennom selvbetjening. Når data allerede er tilgjengelig i datalaget til tjenesteplattformen åpner det for mer eksperimentering, prøving og feiling. Dette er en kontrast til at virksomhetene må utvikle en komplett løsning med

integrasjoner hver gang de skal teste noe nytt innen digitale tjenester. Leverandør vet ikke hva fremtiden bringer, derfor er filosofien deres er at en tjenesteplattform med rik tilgang til data vil sikre at kundene deres er klare for det meste. Suksessen til Leverandør ble at de klarte å koble sammen master data og transaksjonsdata i en hub, og gjøre den tilgjengelig gjennom rapporteringsverktøy som Power BI som fører til at man kommer unna graving etter data i siloene.

Tjenesteplattformen er en skybasert løsning. Fordelene med dette mener informanten er at man reduserer nedetiden. Kjernesystemet kan tas opp og ned for vedlikehold og oppdateringer, men data vil alltid være tilgjengelig i tjenesteplattformen. Data vil da naturlig nok ikke være tilgjengelig i *real time*. Det fører også til redusert belastning på kjernesystemet gjennom API kall mot master data huben som tilgjengeliggjør data for applikasjonene, i stedet for at applikasjonene skal være koblet til kjernesystemet. Informanten sier at “når man kommer i en situasjon hvor applikasjoner skal integreres mer og mer, vil man få flere kall inn mot kjernesystemene, som vil påvirke ytelsen for systemene”. Fordeler som trekkes frem ved dette er at man slipper å investere i infrastruktur og sterkere servere, og virksomheter med *cloud*-løsninger kan betale for skalerbarhet. I en slik sammenheng vil virksomheter unngå å gjøre ytelsesessøkende tiltak på infrastrukturen.

Informanten forteller at deres tjenesteplattform skal bidra til at virksomheter kan frigjøre data fra siloer, transformere data, redusere nedetid, redusere belastning på kjernesystem, og skal fungere som en komponent i en moderne integrasjonsarkitektur. Informanten sier:

Dette er en løsning for å spare tid og kostnader knyttet til det å hente ut et dataelement fra komplekse ERP-suiter. Dette er kunnskap virksomheter sjeldent sitter på selv, slik at de må leie inn folk selv for de enkleste tingene ved å eksempelvis tilgjengeliggjøre data.

Når det kommer til transformasjon av data forteller informanten at datamodellene i ulike applikasjoner gjerne er spesifikke til applikasjonen. Datamodellene spisser seg inn mot analyse, og skal tilrettelegge for modeller som gjør brukernes hverdag enklere. Dette har for eksempel med feltbeskrivelser å gjøre, at et felt skal hete det samme i alle systemer helt fra kilde til sluttbruker, slik at man ikke skaper forvirring om hva det er man analyserer. Dette er

noe informanten merker at kunden setter pris på og som legger til rette for et godt brukergrensesnitt, men som de nødvendigvis ikke etterspør. Gjennom frigjøring av data, skal data tilgjengeliggjøres til bruker både teknisk og fysisk på en form som er førende.

Informanten opplever at forvaltningen av data ved bruk av en hub kontra data som ligger i siloer ikke nødvendigvis vil bli lettere. Informanten sier:

På den ene siden har man bedrifter som er modne innenfor forvaltning, de vil nok ikke oppleve det å flytte data ut i en hub betyr noe for deres arbeidsdag. Dette kan være typiske SAP-miljø der man har tekniske ressurser på SAP, og ansatte som har eierskap til prosesser og data.

Dersom man flytter data opp i en data hub må virksomheter forstå at data forvaltes på vegne av hele organisasjonen, og ikke kun for den enkelte prosessen. Dermed vil de som forvalter master data kunne oppleve at hverdagen blir mer kompleks, sammensatt og krevende fordi master data brukes av flere prosesser og brukere. På den andre siden vil verdien av master data øke fordi den vil være konsis, tidsriktig og korrekt.

Leverandør forteller at de har kunder der virksomheten ikke har noe forhold til forvaltning eller master data fra tidligere. “Disse kundene vil kunne oppleve en verdi av at man faktisk modellerer data, der man sier at dette er master data, dette er eierskapet og slik skal det flyte”. For denne typen kunde vil master data hub vær en ny start. “Enkelte kunder lever i en verden som er svært standardisert, noe som gjør at datamodellene også er styrt av standarder. Andre kunder bruker generisk industristandard, og noen ønsker å skape egen selskapsstandard”. Informanten sier at de som leverandør alltid vil gi råd om at det å bare flytte data ut i en hub, uten å tenke på forvaltning og eierskap, er noe som vil resultere i liten verdi for kunden. Leverandør opplever at flere av deres kunder søker inspirasjon hos dem for hvordan de skal håndtere sine master data. Dette gjelder kunder som ikke er spesielt drevet av detaljer i datalaget, men heller av hva det er de får tilgjengeliggjort i de digitale tjenestene. Større kunder har i noen tilfeller egne *information management team*, og eventuelt datavarehus team, som har jobbet med datamodellering i mange år.

Det kunden må huske er at data rives løs fra sin autorisasjonsmodell når den flyttes til datalaget eller en master data hub. “Derfor er det lurt for kunden å tenke på hva som skal være allment tilgjengelig og hva som skal ligge i kjernesystemet, spesielt viktig er dette i CRM-systemet som sitter på mye kundedata som skal være beskyttet gjennom GDPR”. Dersom kunden ønsker at ulike brukergrupper skal ha ulike tilganger til data, må de modellere dette i databasen eller i API-laget. I datalaget legger kunden til autentiseringsmodell og autorisasjonsmodell for de dataene brukerne skal ha tilgang til. “Et interessant aspekt med dette er at man kan unngå for eksempel lisenser til CRM-system, for å vise data til andre grupper med behov i virksomheten ved å tilgjengeliggjøre data i tjenesteplattformen”. Relevant kundedata kan derfor bli tilgjengelig for flere ansatte.

Videre forklarer informanten at det finnes flere ulike metoder for å lagre data i en slik tjenesteplattform. Når man løfter master data ut i en master data hub blir data duplisert, noe som går ut over lagringskapasiteten til virksomheten. Informanten mener at beste praksis er å tilgjengeliggjøre data som er nødvendig for de scenariene virksomheten mener vil oppstå gjennom å gjøre seg noen tanker om hva man tror man kommer til å trenge i fremtiden. Det er alltid en diskusjon i implementeringen, også med tanke på integrasjoner, om hvilke type data man trenger, hva betyr data og hvordan skal de flyte. Hvilke data som ligger lagret i master data huben mener informanten bør være *default* siste versjon av sannheten. Vedlikehold og opprettelse skjer derfor ikke i datalaget, men datalaget abonnerer på data fra kjernesystemene.

Leverandør sin tilnærming til en master data hub er at den skal ha lav kompleksitet. Huben inneholder ikke noen master data management funksjonalitet ut over de reglene kunden selv modellerer. Kundene står fritt til å koble på dedikerte master data management produkter med algoritmer til å kjøre analyser for dem. “De mer avanserte master data management systemene er gjerne komplekse og dyre”.

Informanten oppgir at det er både modne og umodne kunder som oppsøker dem for å se på muligheter rundt tjenesteplattformen. Leverandør har vært kjent for å ha et sterkt fagmiljø rundt systemintegrasjon med fokus på Microsoft og Oracle teknologi, og tjenester innen

elektronisk samhandling mellom virksomheter, i markedet de opererer i. “Kundene som tar kontakt er ofte kunder som ønsker å videreutvikle integrasjoner, som ofte har gammel teknologi som Biztalk, eller som ønsker å bytte ut gammel integrasjon med en skybasert løsning”. Disse utgjør brorparten av leverandørens kundebase. I de fleste tilfeller har ikke kundene noen klar ide eller strategi om en tjenesteplattform, men dette kommer som følger av integrasjonsproblemer. Informanten sier:

Det er noen virksomheter som har et fokus på digitalisering og digital transformasjon som tar kontakt med oss, med formålet om at de tenker at plattformen deres må bygges på annen måte, men dette er sjeldent og disse anses å ligge i kategorien “*leading edge*”. Da er man helt i front når det gjelder forskning fra MIT og Gartner.

Leverandør opplever de fleste selskapene som relativt konservative, med mindre det er små tech-selskaper som er *born in cloud*. De mest modne kundene har begynt å tenke på selvbetjening ved at de tilgjengeliggjør data slik at de er mer generisk tilgjengelig for forretningsbrukere, eventuelt superbrukere som har noe IT- eller teknisk kompetanse. Dette vil bidra til at brukerne kan lage rapporter og kuber selv med Power BI eller andre verktøy, der det tidligere var IT-avdelingen som eide hele verdikjeden fra ETL til kjernesystemet, og til rapportene lå klare dashbordet. “Dette gjør det også mulig for kunden å gå bort fra forretningsspesifikke dashbord”.

Kjernesystem har en lang endringstakt fordi en endring ofte krever involvering fra en systemleverandør eller en tredjepart. Koding og versjonsoppgraderinger er hinder for endringer i kjernesystemene. Informanten sier:

Koding gjør det komplisert å gjøre endringer, og oppgraderinger er krevende. I digitale tjenester vil dette vil være lettbeint. Når det kommer til *master data lifecycle* er det ikke kun snakk om teknologi, det omfatter også informasjonsmodellering, forretningsforståelse og hvordan data anvendes i ulike prosesser. Dette er et veldig komplekst fagområde.

Informanten sier at kriteriene for valg av leverandør av master data management, presentert i Forrester sin rapport, virker fornuftige. “Rapporten er nok tilpasset relativt store kunder med avansert behov for master data management og mange vil nok klare seg med mindre avanserte løsninger enn som skissert via Forrester-rapporten”. En kompleks organisasjon med

flere kjernesystemer og IT-arkitektur som er et resultat av oppkjøp, uten særlig form for konsolidering, kan ha behov for en slik *best of breed* tilnærming. Informanten trekker også frem at noen egenvalgte evalueringskriterier er; pris- og supportmodell, integrasjonsvennlighet og hvilke integrasjonsmønster som støttes (som ETL, *real time*, eventer), og brukervennlighet. Ofte håndteres ikke master data management av tekniske personer, men av personer med kompetanse innen forretning og data, som *data stewards*. “Harde egenskaper som sikkerhet, skalerbarhet, HA/DR osv. støttes av teknologien som ligger i datalaget og infrastrukturen i skyløsningen, altså tjenestene som master data huben er bygget på”.

4.3 Bedrift 1

Master data er et enormt område som informanten i Bedrift 1 jobber med til daglig. For å kunne vite om det er gunstig å flytte data til en master data hub, mener informanten at det er viktig å ha et forhold til hva master data er i sin egen virksomhet, men også hva det ikke er. I Bedrift 1 er det 10-12 master data områder som ikke har noe med hverandre å gjøre, der master data objektet “kunde” er det virksomheten har størst fokus på. Grunnen til dette er at det er ulike typer kunder som er svært forskjellige, ligger samlet under dette objektet.

Bedrift 1 anser data som en del av virksomhetens ressurser. Dette er blant annet tydeliggjort i deres forretningsstrategi og teknologistrategi. “Det kan diskuteres om virksomheten opptrer deretter”. Bedrift 1 kan ikke svare på generell basis om hva de mener er god kvalitet i master data, da dette varierer mellom ulike master data. Generelt er det fokus på god kvalitet på alle virksomhetens master data objekt. Virksomheten har et forhold til hvordan datakvaliteten i master data påvirker forretningen, og informanten trekker spesielt frem at mangelfull artikkel-, kunde-, og brukerdata oppleves som et hinder i virksomhetens strategiske måloppnåelse. Informanten svarer på vegne av seg selv når det kommer til hva som er kilden til dårlig datakvalitet og hvordan virksomheten kan forbedre denne. “Det er mangelfulle data, eller datakvalitet, i de fleste master data objektene bortsett fra leverandørdatabaser”. Artikkel master data sliter med mangelfulle data fra de eksterne databasene. Virksomheten sliter også med enkelte inn- og utmeldingsprosesser og noen datastrukturer. Ansatt master data har

manglende sentral *governance* innen HR og ansvaret ligger fordelt på flere enheter. For markedssegmentene er det manuelt vedlikehold som er den største feilkilden.

Det er delvis klare retningslinjer om hvem som har tilgang til å lese og skrive master data, og hvem som skal bruke master data i de ulike systemene og applikasjonene. Retningslinjene er tydelig når man snakker om brukere, da disse er knyttet til autorisasjoner. De er mindre tydelige når virksomheten skal se på hvilke deler av organisasjonen, eller hvilke løsninger, som har eierskap til data. Data sikres gjennom tilgangsstyring på brukernivå, redundante løsninger og generell sikring av den tekniske infrastrukturen. Det er ingen generelle retningslinjer når det kommer til fysisk lagring og livssyklusen til master data. Hvert master data objekt har hver sin *policy* knyttet til håndtering av livssyklusen. Utgåtte artikler vil eksempelvis arkiveres i datavarehuset.

Når man snakker om en master data hub fra et systemløsningsperspektiv mener informanten det finnes mange løsninger som støtter *multi domain*, altså en master data management løsning som støtter flere domener. Informanten sier:

Domener som kunde, artikkel og leverandør kan gi mening å ha i samme master data hub, fordi de kan være knyttet til hverandre, som at en artikkel er knyttet til en spesifikk leverandør. Men det er også mange master data som ikke har noen forretningsmessig verdi eller operasjonell verdi ved å være samlet i en master data hub.

Fra virksomhetens ståsted må man ha et forhold til hvilke master data det gir mening å holde sammen gjennom å samle de i en hub. Det gir ikke mening å flytte master data objekter i samme hub kun for løsningen sin del. Det vil ikke gi mening å flytte artikkel master data og kunde master data i samme hub, da de er to separate domener og koblingen mellom disse er transaksjonsdata.

Informanten oppfatter sin virksomhet som en moden organisasjon når det gjelder flere typer master data, da disse er uttalt kritiske for virksomheten. Modenheten er særlig god knyttet til artikkeldata, kundedata og til dels leverandørdato. "En del andre master data områder står det

nok betydelig dårligere til med, men det jobbes aktivt med å heve kvaliteten, eksempelvis på ansattdata og markedssegment data”. Komplexiteten innen master data i virksomheten kommer i stor grad av størrelsen på organisasjonen, og mange komplekse og sammensatte prosesser. Derfor oppstår det en del hull i håndteringen av master data i enkelte segment. “På enkelte områder har vi for eksempel dramatisk mye dårligere master data knyttet til artikkel enn vi burde”. Informanten oppfatter at prosedyrene i virksomheten er gode angående master data, men de er i ulik grad rettet mot selve kvalitetsdimensjonen.

Når det kommer til barrierer til kvalitet innen master data skriver informanten at rollene og ansvaret i teorien er ganske godt definert, men det er krevende når flere deler på ansvaret rundt master data. Videre er informanten veldig enig i at manglende eierskap til datakvalitet er en barriere. Informanten sier:

Jeg tror mye av problemet hos oss ligger i at vi ikke følger opp kvaliteten spesifikt. Vi er mer opptatt av eierskap til data generelt, og kvalitet blir dermed bare en implisitt del av dette ansvaret, og dermed med litt «tilfeldig» fokus.

Når det kommer til manglende teknologi og verktøy mener informanten at dette er sammensatt, og ulikt fra domene til domene. Artikkel master data har pr. tidspunkt dårlige løsninger som skaper sårbare og omstendelige prosesser. For kunde master data ligger det krevende utdatert arkitektur rundt nye løsninger som implementeres. Det finnes ingen felles løsning for håndtering av all ansatt master data fra de ulike segmentene, men det pågår prosjekt for å forbedre dette. I markedssegmentene er det ingen gode felles løsninger for å håndtere datasett, men også her planlegges det forbedringer. På manglende forståelse og involvering fra ledelsen skriver informanten følgende: “Ulikt fra domene til domene, og sikkert også på ulike ledelsesnivåer. Tror likevel dette ikke er et hovedproblem”.

Bedrift 1 jobber rettet mot en løs kobling av arkitekturen sin, som vil si hvordan de skal få system til å være minst mulig avhengige av hverandre. “Man kan si at et lagringssted for master data kan være et virkemiddel til det, men alle master data på ett sted gir den motsatte effekten gjennom å skape nye avhengigheter som ikke er nødvendige”. Bedrift 1 jobber med

å flytte artikkel master data til en master data hub. Dette fører til at virksomheten får en autoritativ kilde til master data som ikke er tett knyttet til et spesifikt fagsystem, som logistikk eller økonomi, men som et nøytralt master data system.

Årsaken til at Bedrift 1 søker en løsning for master data håndtering av artikkel master data er fordi dagens løsning har gått ut av support, samt at den ikke fungerer optimalt. Teknisk sett har de en løsning i dag som fungerer som et *repository* for artikkel, men i realiteten er det ERP-systemet deres som er den autoritative kilden til artikler. Informanten sier:

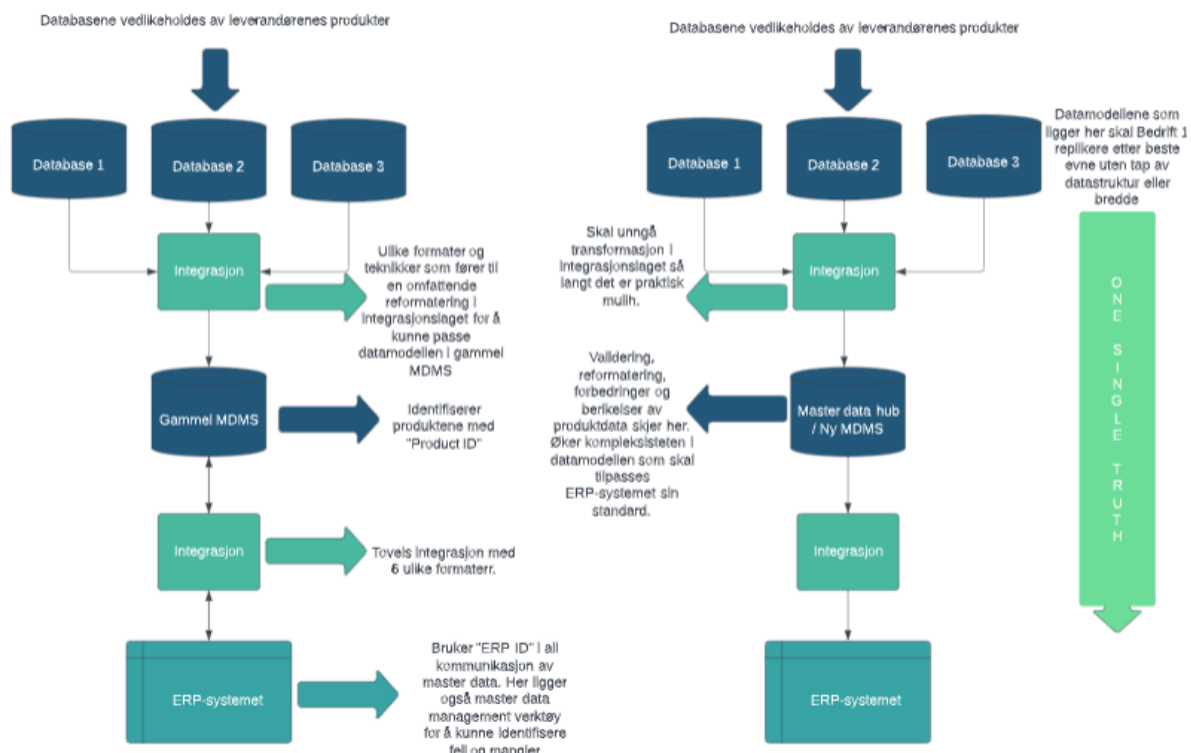
I tillegg har ikke dagens *repository* kapabilitetene virksomheten ønsker fra en master data management løsning; den er ikke fleksibel nok i distribusjon av master data og den mangler verktøy for å orkestre litt mer avanserte valideringer av mange datastrukturer samtidig.

Dette ser de på som et stort problem som må løses ved å bytte ut løsningen som brukes pr. dags dato. I tillegg må de fjerne interaksjonen som ligger i bakkant mellom ERP-systemet og de andre systemene og applikasjonene, slik at disse er koblet rett mot en ny master data management løsning. De har også utfordringer knyttet til kunde master data som ligger lagret i et CRM-system. Her er ikke kundedatabasen nøytralisert. Det vil si at master data er systemavhengig, og skaper direkte integrasjoner mellom CRM som holder master data og alle andre systemer og applikasjoner som bruker kunde master data. “Når systemet som holder master data også er et operativt system skaper det konflikter og virksomheten låser systemet veldig. Det er noe som gjør det vanskelig å gjøre endringer i arkitekturen og utvikling i systemene over tid”.

Informanten sier at dette er en tenkemåte virksomheten ikke hadde for 2 år siden, men at prosjektet er en direkte konsekvens av teknologistrategien virksomheten har utarbeidet. Virksomheten ønsker å gjøre dataprodukt (master data og transaksjonsdata) tilgjengelig gjennom åpne, og definerte grensesnitt. Fordi ERP-løsningen setter begrensninger for hvordan artikkel master data kan se ut (med en egen ID), er det ikke alltid like lett å distribuere denne direkte til andre systemer og applikasjoner som skal bruke artikler. For å kunne koble artikkel master data mot netthandelsløsninger, er virksomheten nødt til å finne omveier for å kunne omstrukturere master data slik at de kan benyttes i løsningen. Her ser

Bedrift 1 for seg at master data i stedet kan distribueres direkte fra en master data management løsning. Da vil virksomheten unngå å skape flere konkurrerende kilder, men heller ha en autoritativ kilde til artikkel master data.

Bedrift 1 startet arbeidet med prosjektet høsten 2021 som et resultat av virksomhetens reviderte teknologistrategi. De er nå i gang med å se etter leverandører som kan møte deres krav. I teknologistrategien kommer det frem flere teknologiske mål som understøtter implementering av en plattform for artikkel master data management; avkobling og modulisering av systemarkitekturen, redusere ERP-systemet sitt fotavtrykk, data som et produkt, agil utvikling og kontinuerlig implementering. *Upstream* (leverandør) skal endres, *downstream* (master data) skal være urørt. Bedrift 1 sin arkitektur vil også bli påvirket av den nye teknologistrategien, og da spesielt med tanke på integrasjoner. Artikkel master data management påvirker flere applikasjoner og applikasjonstjenester, både internt og eksternt, i Bedrift 1. Under er en oversikt over Bedrift 1 sin nåværende arkitektur (til høyre) og deres optimale arkitektur (til venstre):



Figur 9: Illustrasjon av arkitekturer, Bedrift 1 (egen figur)

Nåværende situasjon i Bedrift 1 er som følger: tre eksterne databaser, ett pr. markedssegment, sørger for mesteparten av inputen av artikkel data i virksomhetens verdikjede. Data fra disse kildene importeres ved bruk av ulike format og teknikker til et eget master data management system. For å gjøre videre lesing enklere har vi valgt å kalle dette systemet “*master data management system*” (heretter referert til som MDMS). Databasene vedlikeholdes av leverandørene til artiklene. Når master data importeres utføres det en omfattende reformatering i integrasjonslaget til Bedrift 1, slik at data passer datamodellen til MDMS. Dette systemet har en toveis integrasjon med ERP-systemet. På grunn av ulike scenarier og diversifisering av artikler, består integrasjonen av seks ulike format.

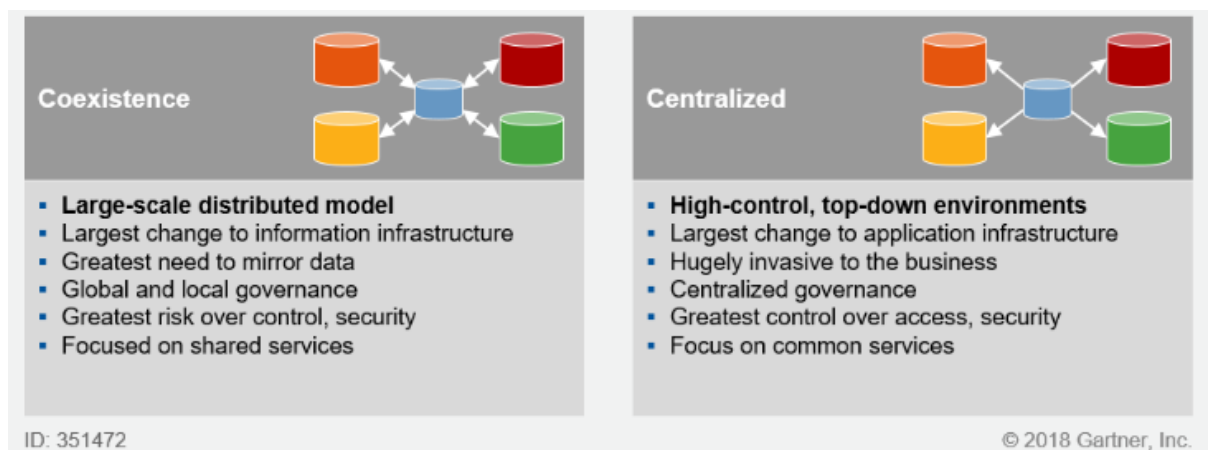
Identifikasjonen av artikler ligger i en egen “*Product ID*” til tross for at ERP-systemet bruker og utnytter en egen “*ERP ID*” til all kommunikasjon og distribusjon av artikkel master data. I tillegg til ERP-systemet brukes et eget master data management verktøy for å identifisere feil og manglende egenskaper i artiklene. Kritiske parameter og egenskaper håndteres manuelt i ERP-systemet, som derfor kan ansees som det faktiske systemet for master data management. Noen, men ikke alle, berikede data lastes så tilbake til MDMS. Prosesser for prising og sortiment håndteres også i ERP-systemet. Man kan dermed si at det er ERP-systemet som distribuerer relevant data videre til Bedrift 1 sin systemportefølje som består av system orientert rundt verdikjeden, *point-of-sale*, *e-commerce*, *product information management*, *digital asset management* systemer, og spesialiserte planleggingssystemer.

Bedrift 1 har foreslått en optimal arkitektur der MDMS først og fremst er erstattet og tatt ut av drift. De tre databasene skal reintegreres mot et nytt MDMS. “Datamodellene som brukes i disse databasene skal etter beste evne bli replisert på Bedrift 1 sin side, uten tap av datastruktur eller bredde. Dette krever sofistikert data modellering som skal sammenstille alt til *one single truth*”. Virksomheten forventer også å tilrettelegge for nye mønster innenfor integrasjonen, som skal utnytte eksisterende og fremtidige egenskaper i de tre databasene. De skal unngå transformasjon av data i integrasjonsnivået så langt det er praktisk mulig.

Validering, reformatering, forbedringer og berikelse av artikkel data skal skje på et så tidlig tidspunkt som mulig. Dette innebærer også at noe kan utføres allerede i databasene. Fra

Bedrift 1 sin side skal dette skje i MDMS, og ikke i integrasjonen mellom MDMS og ERP-systemet. Integrasjonen mellom MDMS og ERP-systemet skal være lik som den er i dag med samme midler og formater. Dette skal minimere risikoen, selv om det kanskje øker kompleksiteten til MDMS sin datamodell og integrasjon. Datamodellen skal passe ERP-systemet sine forventninger. ERP-systemet skal fremdeles håndtere pris- og sortiment prosessene, og vil derfor fremdeles fungere som distribusjonskanal av relevant data til Bedrift 1 sin systemportefølje.

Bildene under er med på å visualisere hvordan lagring og distribusjon av master data kan se ut i en virksomhet som Bedrift 1. Dette er ikke master data huber som samler all master data, men som samler master data til en autorativ kilde. Det finnes mange ulike varianter av disse arkitekturene, og mange ulike måter å gjøre og tenke master data på i virksomheter. “Ulike virksomheter vil ha veldig ulikt behov når det kommer til sine master data domener”. Som en aktør som har bygget sin virksomhet innenfra, er det disse to implementeringsstilene, *Coexistence* og *Centralized* det gir mening for informanten å se på i forbindelse med MDMS:



Figur 10: MDM Roadmap for Typical Product/B2B Customer Data (O’Kane et al., 2018, s. 10)

Informanten mener at arkitekturen kan være ulik for de forskjellige master data domenenene, da det ikke er sikkert at det som fungerer for artikkel master data også er det som fungerer for kunde master data. Leverandør master data har en sterk relasjon til artikkel master data

objektet. Når artikkel master data flyttes over til en ny MDMS løsning må virksomheten sørge for å håndtere denne relasjonen på en god måte, samtidig som de ivaretar den teknologiske strategien om en løsrevet arkitektur. “Da det kun er artikkel master data som skal flyttes ut i en nøytralisert løsning vil det føre til utfordringer og belastninger for integrasjonen til virksomheten, da master data objektene ligger lagret på ulike steder”. Bedrift 1 har valgt å fokusere på å sammenstille objektene ved behov, i stedet for å flytte de ut i den samme løsningen. Informanten mener at de ved å flytte både leverandør master data og artikkel master data over i den samme nye løsningen vil virksomheten skape stadig nye avhengigheter. “Hva oppnår man egentlig ved å flytte avhengighetene fra ett sted til et annet”.

4.4 Bedrift 2

Bedrift 2 anser data som en del av virksomhetens ressurser. Data er viktig for virksomheten og er relatert til en god del av forretningsområdene som kunder, forvaltning og finans. Virksomheten sin datamodell er en definert domenemodell. Når Bedrift 2 refererer til domener gjelder det sju domener for ansatt, produkt og de ulike markedssegmentene. Hvert domene har ulike systemer og applikasjoner de er koblet til, og innenfor hvert domene er det én til flere master data huber som sender og mottar master data fra fagsystemet. Denne modellen er en felles beskrivelse av hvilke data og relasjoner mellom data som benyttes på tvers av virksomheten. “Domenemodellen danner grunnlaget for felles begreper innad i virksomheten, samt et felles datasett som systemene benytter for deling av informasjon”. Bedrift 2 har jobbet bevisst med varige data (master data) fremfor transaksjonsdata. Domenemodellen er logisk og beskriver hvilke data virksomheten har et forhold til og hvordan de er relatert til hverandre. Det er derimot ikke alt som er beskrevet i datamodellen som er realisert som en datakilde i deres tjenesteplattform, og det som er realisert kan ha en noe annen form i praksis. Datamodellen legger til rette for at master data lagres både i kildesystemet og i master data huben. Bedrift 2 er i en prosess med å utfase sitt system, som skal erstattes av flere fagsystemer. Alle disse fagsystemene kommer til å benytte seg av master data fra ett eller flere domener.

Bedrift 2 mener kvaliteten i master data i virksomheten skal være høy. Dette er fordi fagsystemene skal kunne stole på master data de henter eller får via plattformen. Informanten sier at lav kvalitet i master data i virksomheten vil føre til ekstra manuell håndtering og mulige feil i forretningsprosesser, som til syvende og sist kan påvirke sluttkunden. Virksomheten er kjent med hva som er kildene til dårlig datakvalitet, og i noen tilfeller har de iverksatt initiativ for å rydde i master data slik at kvaliteten øker. I Bedrift 2 er det kildesystemet som er ansvarlig for kvalitet i master data, samt at tjenesteplattformen inneholder noen kapabiliteter som duplikatsjekk for å sikre kvaliteten. Da det er kildesystemene som er eier av master data er det der rettighetene til å oppdatere og vedlikeholde data ligger. Forbruker av master data får rettigheter til å lese gjennom godkjenning fra dataeier. Informanten sier:

Dersom forbruker skulle ønske å oppdatere master data, sendes oppdateringen til kilden (der data ligger lagret i systemet) og kilden oppdaterer deretter master data huben med nye data. På denne måten er også hvert fagsystem ansvarlig for å sikre sine data, og virksomheten har for vane å kjøre sikkerhetstester.

Lagringen av master data er avhengig av fagsystemet. Noen system i Bedrift 2 er *on-premise* og noen ligger i skyen, dermed er også lagringen av master data splittet mellom disse to. Fagsystemene eier forretningsregler angående opprettelse, oppdatering og sletting av master data. Master data huben forholder seg til disse oppdateringene fra fagsystemene. De bruker også master data huben som en master data portefølje for data som er relevant for flere systemer.

Da Bedrift 2 skulle velge en leverandør var det viktig at de skulle få en samarbeidspartner som hadde erfaring med digital tjenesteplattform. De gikk gjennom en prosess med flere ulike partnere som fikk muligheten til å presentere seg og sitt produkt. Kriteriene for valg av leverandør var kunnskap rundt skytjenester, digital tjenesteplattform og master data håndtering. I tillegg til at leverandøren hadde mulighet til å stille med flere konsulenter for samarbeid og overføring av kunnskap til virksomhetens ansatte. Virksomheten trekker selv frem at det viktigste kriteriet til leverandør var kunnskap om teknologi og metodikk som kunne støtte realiseringen av domenemodellen.

Årsaken til at Bedrift 2 søkte etter nye løsninger på markedet var utfasing av fagsystemet sitt. Virksomheten hadde hatt et fagsystem som hadde dekket de fleste behov og funksjonaliteter i flere tiår. Da fagsystemet skulle utfases oppstod det et behov for å håndtere master data på en annen måte. Gjennom denne endringsprosessen endret arkitekturen i virksomheten seg, og det ble klart at virksomheten manglet intern kompetanse på en del områder knyttet til håndtering av master data. De søkte derfor etter ekstern kompetanse. I tiden da fagsystemet var kilden til master data opplevde virksomheten duplikater og inkonsistens. Dette ble håndtert direkte i systemet, og opplevdes ikke som noe stort problem. “Nå, når master data distribueres fra master data hub, fører duplikater og feil i datagrunnlaget til utfordringer”.

Bedrift 2 kjenner seg igjen i barrierene knyttet til kvalitet i master data, og synes disse beskriver utfordringene de møter. Spesielt nevnes manglende systemstøtte som tillater duplikater og manglende opplæring av brukere. Nå som virksomheten distribuerer master data fra master data hub til flere fagsystem er håndtering av, og eierskap til, master data en utfordring. “I noen tilfeller er det flere system som eier deler av dataene, og da blir det vanskelig å sikre at master data kun oppdateres fra kilden. Dette er en utfordring vi jobber med å løse på en god måte”.

Bedrift 2 opplever at ting generelt har blitt mer komplisert når data blir fordelt på flere system, og informanten trekker frem at avhengighetene har økt i takt med kompleksiteten. Samtidig har også systemstøtte for brukerne (duplikater) økt, og virksomheten forventer at kvaliteten i master data dermed vil øke over tid. Virksomheten har flere system som er deleier av master data, som øker muligheten til feilkilder. “Feilsøking involverer derfor ikke bare én avdeling lenger, men tre til fire. Dette er tidkrevende og skaper unødvendig arbeid i noen av avdelingene”.

5. Diskusjon

I dette kapitlet skal vi analysere data ved å sammenstille våre teoretiske og empiriske funn, basert på våre forskningsspørsmål. Dataanalyse har ifølge Johannessen et al. (2011) til hensikt å organisere data etter tema og analyse, og tolke data. Når man skal organisere data etter tema vil man redusere, systematisere og ordne datamaterialet slik at det danner et godt grunnlag for analysen, uten å miste viktig informasjon. (Johannessen et al., 2011, s. 185) har basert på (Silverman, 2006) sin forskning funnet ut at i kvalitative studier er det viktig at den som innhenter data også analyser og fortolker disse fordi teorier, hypoteser og forskerens forforståelse er selve utgangspunktet for analysen av dataene. Analyse handler om å dele opp noe i biter, der målet er å finne et budskap, mening og mønstre i datamaterialet. Vi har strukturert våre datafunn etter forskningsspørsmål (se vedlegg 8), og vil i dette kapitlet sammenstille funnene med antakelser fra litteraturstudien vår, med lik struktur. Dette gjør det mer oversiktlig for leseren da Bedrift 1, Bedrift 2 og Leverandør vil nevnes om hverandre, og vi mener derfor en struktur som dette er fornuftig. Litteraturstudien og empirien bruker begrepene informasjonssystem, kjernesystem og kildesystem om hverandre, men har samme funksjon. Når dette diskuteres for begge bedriftene i denne delen, har vi valgt å bruke begrepet kjernesystem.

5.1 Hva er god datakvalitet?

Datakvalitet er et mål på om data er tilstrekkelig for den enkelte prosessen dataene skal brukes i. Det er definert mange dimensjoner som skal måle datakvalitet, men det som går igjen flest ganger er variasjoner av nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet (Christen, 2012; DAMA, 2013; Hikmawati et al., 2021). Ingen av virksomhetene i vår casestudie har noen klare mål på hva god datakvalitet er, men begge er enige om at det skal være høy datakvalitet i master data. Bedrift 2 trekker frem at dette er fordi fagsystemene skal kunne stole på data de henter eller får via sin tjenesteplattform. Virksomhetene er også enige om at dårlig kvalitet i master data vil gå ut over forretningsprosessene, som til syvende og sist går ut over sluttbrukerne.

Bedrift 1 sier at dårlig kvalitet i master data påvirker forretningen og er et hinder for strategisk måloppnåelse. Det kommer frem at virksomheten har mangelfulle data eller datakvalitet i de fleste master data objektene sine. Det er kjernesystemene som vedlikeholder master data, men deleierskap til datakvalitetsdimensjon er en barriere til Bedrift 1. I Bedrift 2 sitt tilfelle er det kjernesystemet som er ansvarlig for kvaliteten i master data, samt at tjenesteplattformen inneholder egenskaper for duplikatsjekk slik at kvaliteten bedres. Begge virksomhetene har behov for bedre kvalitet i master data, og dermed har de også et behov for bedre master data management. Virksomheter forbedrer sin master data management gjennom *data governance*. Med korrekt *data governance* og oversikt vil, data i master data systemet (eller oppbevaringssted/register) kvalifiseres som en enhetlig og konsistent dataressurs, som alle applikasjoner kan stole på for informasjon av høy kvalitet (Loshin, 2009). I master data management hensyntas og ivaretas datakvalitet gjennom dokumenterte roller og ansvar som ligger inn under *data governance* (Buffenoir & Bourdon, 2013; Hikmawati et al., 2021).

Vi mener empirien vår viser et resultat av dårlig *data governance* ved at flere systemer eier data i Bedrift 1. I virksomheten bruker ulike avdelinger master data til ulike formål, noe vi mener gjør det vanskelig å sette standarder for eksempelvis fyllingsgrad. Bedrift 2 har flere fagsystem som er deleier av master data, og master data hub vet ikke hvilket fagsystem som har *golden record*. Dette kan føre til inkonsistens i master data for begge virksomheter. (Hikmawati et al., 2021; Lerche, 2014) mener datakvalitet er et stort problem for virksomheter som har informasjonssystem fordelt på flere forskjellige enheter og avdelinger. Slik arkitektur legger til rette for problemer i datakvalitet som duplikater og inkonsistens, og bruk av master data management skal redusere disse problemene. Våre datafunn viser at master data management har gjort hverdagen til Bedrift 2 mer komplisert ved å involvere flere fagsystemer. Det er dermed ikke gitt at master data management løser arkitektur problemene, arkitekturen blir like dårlig bare med andre systemer. Informasjonssystem må være troverdig og pålitelig for å sikre kvalitet, konsistens og korrekthet i data for å støtte forretningsbeslutninger og som et grunnlag for fremtidig vekst i virksomheten (Hikmawati et al., 2021).

Barrierer knyttet til god kvalitet i master data er ifølge (English, 1999; Haug & Arlbjørn, 2011; Lee et al., 2006; Umar et al., 1999; Xu et al., 2002):

- Mangel på roller og ansvar
- Mangel på eiere knyttet til datakvalitet
- Ineffektive organisasjonsprosesser
- Manglende verktøy og teknologi
- Manglende prosedyrer
- Manglende belønningssystemer
- Manglende forståelse og involvering fra ledelsen
- Forsømmelse av administrative detaljer som opplæring, arbeidsbeskrivelser, ansvarsfordeling og problemer knyttet til kommunikasjon

Bedrift 1 mener roller og ansvar i teorien er godt definert, men at det er krevende når flere deler på ansvaret for master data. Spesielt ser informanten på eierskap som en barriere til datakvalitet i virksomheten, da kvalitet blir en implisitt del av eierskapet. For master data innen ansatte og markedssegmenter er det ingen gode felles løsninger innen teknologi og verktøy, noe som gjør de mer krevende og håndtere. Informanten tror ikke at manglende forståelse og involvering fra ledelsen kan anses som en barriere til kvalitet i master data. Bedrift 2 kjenner seg igjen i barrierer knyttet til god kvalitet i master data, og synes disse beskriver utfordringene de møter godt. Spesielt nevner informanten manglende systemstøtte for duplikatsjekk og manglende opplæring av brukere.

Begge virksomhetene kjenner seg igjen i barrierene som trekkes frem i litteraturen, men trekker frem ulike barrierer som er spesielt utfordrende for dem. For Bedrift 1 er dette manglende eierskap knyttet til master data og for Bedrift 2 er det manglende systemstøtte og opplæring. Vi gjør oss derfor noen tanker om tjenesteplattformen fungerer optimalt for Bedrift 2, da deres ønske var systemstøtte for blant annet duplikater og ekstern kompetanse for å lære opp sine ansatte. Samtidig sitter vi igjen med det inntrykket at Bedrift 2 belager seg på at tjenesteplattformen skal sikre deres datakvalitet, noe som understøtter deres krav til tjenesteplattformen. Under vårt forskningsarbeid kan vi ikke se at tjenesteplattformen har løst Bedrift 2 sine problemer knyttet til datakvalitet.

Pipino et al. (2002) mener virksomheter må gjennom tre steg for å forbedre sin datakvalitet:

1. Analysere de subjektive og objektive vurderingene av datakvalitet
2. Sammenligne resultater for å identifisere avvik og årsaker
3. Iverksette tiltak for forbedringer

Det finnes ikke en *one size fits all* løsning, da vurdering av datakvalitet er en kontinuerlig prosess som krever bevissthet om grunnleggende prinsipper for subjektive og objektive målinger av datakvalitet (Pipino et al., 2002). Vi opplever at begge virksomhetene i vår studie har analysert de subjektive vurderingene av datakvalitet. Vi har fått høre fra begge virksomhetene at dårlig datakvalitet fører til feil i forretningsprosesser. Begge virksomhetene i vår studie vet hva avvik i kvaliteten på data er, og hvor årsaken ligger. Bedrift 1 har iverksatt prosjekter for å bedre sine master data for artikkel, ansatte og kunder. Når det gjelder artikkel master data er en master data hub løsningen for virksomheten. Den skal sørge for at data er tilgjengelig, nøyaktig, konsis, tidsriktig, fullstendig og gyldig. Bedrift 2 har i noen tilfeller iverksatt initiativ for å rydde i sine master data.

I følge (Chien & Jain, 2021) har *data quality management* modnet de siste årene. De skriver at et resultat av dette er at man integrerer løsninger for datakvalitet gjennom blant annet data integrasjonsverktøy og master data management produkter. Dette er i samsvar med skiftet innen datakvalitet, der virksomheter har gått fra å redusere operasjonell risiko og kostnader, til at datakvalitet er en nødvendighet når virksomheter skal forsterke analyser for å bedre innsikt, og ta pålitelige, datadrevne beslutninger. Dette skal igjen kunne gi virksomhetene konkurransefortrinn. Leverandør mener at dersom kvaliteten i master data er god kan data migreres rett ut i applikasjoner og tjenester som ligger i tjenestelaget, samt til analyser og rapportering.

Datakvaliteten i tjenesteplattformen skal sikres gjennom *system of entry*, altså i kjernesystemene. Leverandør tilbyr ikke innebygd master data management i sitt datalag, men kunden står fritt til å koble på AI og ML selv. Det å koble på AI og ML krever god innsikt i informasjonsmodellering, noe som henger sammen med at virksomheter er svært modne. Vi mener at virksomheter som er svært modne og som krever avansert master data management, ville kjøpt en ferdig tjeneste, og ikke brukt tid på å modellere dette selv. I

Yuhanna et al., (2021) sin rapport kommer det frem at kunder bør velge leverandører som legger til rette for rask implementering i form av ferdige forretningsmaler, som da inneholder avansert master data management.

Basert på vårt litteratursøk og våre datafunn er det tydelig at virksomheter i praksis ikke har noe entydig svar på hva god datakvalitet er. Det eneste de vet er at de jobber for å ha det og ønsker at den skal være høy. Vi mener Bedrift 1 og Bedrift 2 måler god datakvalitet ut ifra hvor dårlig datakvalitet de har. Det er lettere for virksomhetene å oppdage mangler knyttet til datakvaliteten, enn det er å fortelle oss hva god datakvalitet er for dem. Vi mener det at begge virksomhetene har iverksatt tiltak for å bedre sine master data er en bekreftelse på at de ønsker høy kvalitet i data og at de kontinuerlig søker etter å forbedre denne. På denne måten anerkjenner de hvor viktig den er i forretningsprosesser og analyser.

Det vi mener er alvorlig med dårlig kvalitet i master data er at mangelfull informasjon kan føre til feil i forretningsprosesser og/eller analyser, som fører til at virksomheter tar beslutninger på feil grunnlag. I tillegg er dårlig datakvalitet dyrt, og det hindrer innovasjon og digitalisering. Bedrift 2 forteller de er nødt til å søke etter feil i master data i alle sine fagsystemer når feil først inntreffer, etter de implementerte master data hub. En slik feilsøking tar tid og involverer tre til fire avdelinger. Vi mener det er alvorlig at virksomhetene ikke har noen nøyaktig plan for hvordan de skal oppnå nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet i sine master data. Det fremkommer ikke av vår forskning at virksomhetene har en strategi for master data management, som resulterer at de ikke forbedrer datakvaliteten.

Det fremkommer at leverandøren ikke leverer det som markedet forventer når det gjelder datakvalitet gjennom innebygd master data management, men at de spiller på sine styrker innen integrasjon og legger kontroll og berikelse her. Dermed vil ikke master data huben i dette tilfellet være med på å sikre god datakvalitet, men det er derimot kunden som gjør dette selv gjennom forvaltning og datamodellering. I begge virksomhetene er det kjernesystemet som skal sikre datakvalitet. Bedrift 1 har et ønske om å gå bort fra dette, da optimal arkitektur for dem består av validering, reformatering, forbedringer og berikelse av artikkel data på et så

tidlig tidspunkt som mulig. Altså før artikkel data ender opp i ERP-systemet, klart til bruk og distribusjon.

Master data management kan løse problemer knyttet til datakvalitet forårsaket av data som kommer fra ulike kilder gjennom prosesser som *data profiling* og standardisering (Hikmawati et al., 2021). Det som er vanskelig i IT-verden er at det finnes mange formater og former på data, noe Leverandør refererer til som stammespråket. Vi mener dette gjør det vanskelig å utforme standarder, og kunder kan derfor bli avhengig av omfattende datamodellering for å ta i bruk master data hub.

Kvalitet på data måles som tidligere nevnt i nøyaktighet, fullstendighet, konsistens, aktualitet, tilgjengelighet og validitet. Master data management verktøy skal kunne sikre datakvaliteten. Under har vi illustrert hvordan master data hubene fra vår studie kan måles opp mot disse dimensjonene. Vi har gjort følgende funn blant våre informanter:

Dimensjon	Funn	Informant
Nøyaktighet	Master data hub inneholder kapabiliteter for duplikatsjekk som skal sikre kvaliteten	Alle
Fullstendighet	-	-
Konsistens	I en master data hub vil dataformat fra ulike kilder sammenstilles ved hjelp av datamodellering. Dette skal sørge for konsistente data på tvers av virksomheten.	Alle
Aktualitet	Master data i en master data hub bør være <i>default</i> siste versjon av sannheten	Leverandør
Tilgjengelighet	Master data vil i en master data hub være tilgjengelig når systemer er nede for oppgradering eller hendelser.	Leverandør
Validitet	Master data hub kan ha monitorering og avstamning som viser flyten til data, kilder til feil og hvor det eventuelt stopper i prosesser.	Alle

Vi mener teorien er veldig klar på hva som fører til god og dårlig datakvalitet, og hvilke konsekvenser dette har for virksomheter. Vår oppfatning er at økt kvalitet innen en dimensjon vil føre til økt datakvalitet. Vi ser av tabellen over at master data hub fra vår studie kan bidra til økt datakvalitet i dimensjonene nøyaktighet, konsistens og validitet. I tillegg sier Leverandør at master data hub skal føre til aktualitet og tilgjengelighet i data, men dette har vi ikke fått bekreftet fra virksomhetene våre. Det stemmer at master data management og *data governance* skal sikre datakvalitet. Vi kan ikke utvikle teorien med at master data hub, fra vår studie, skal kunne sikre datakvalitet. Vi ser derimot at de kan bidra til økt datakvalitet.

5.2 Hvilke strategiske, taktiske og operasjonelle valg må en virksomhet ta for å ha god master data management?

Bedrift 1 har delvis klare retningslinjer om hvem som har tilgang til å lese og skrive master data, og hvem som skal bruke master data i de ulike systemene og applikasjonene. Dette er tydelig når man snakker om brukere, da disse er knyttet til autorisasjoner. De er mindre tydelige når virksomheten skal se på hvilke deler av organisasjonen eller hvilke løsninger det er som har eierskap til dataene. Master data brukes i mange systemer, og lagres hovedsakelig i ERP-systemet og i CRM-systemet. For å dra nytte av tekniske innovasjoner er man avhengig av at systemet er oppdatert og vedlikeholdt korrekt, samt at data som er i bruk må holdes ren ved å arkivere eller slette utdaterte data i henhold til livssyklusen (Bowker et al., 2009; Spruit & Pietzka, 2015). Som informanten selv forteller er det ugunstig når systemet som holder master data også er et operativt system, da dette skaper konflikter og at virksomheten låser systemet veldig. I tillegg blir det vanskelig å gjøre endringer i arkitekturen og utvikling i systemene over tid. Virksomheten har satt seg strategiske og operasjonelle gode mål, gjennom både forretningsstrategien og teknologistrategien, men vår forskning viser at de mangler å ta taktiske valg i forhold til datakvalitet og *data governance*. Dette gir ikke et optimalt utgangspunkt for god master data management.

Begge virksomhetene anser data som en viktig ressurs for sin forretning, men informanten for Bedrift 1 sier det kan diskuteres om virksomheten opptrer deretter. Bedrift 1 har tydeliggjort data som ressurs gjennom sin forretningsstrategi og teknologistrategi. I teknologistrategien

fremkommer det flere teknologiske mål som understøtter implementering av en plattform for artikkel master data management. Dette er mål som skal bidra til avkobling og modulisering av systemarkitekturen, redusere ERP-systemet sitt fotavtrykk, data som et produkt, agil utvikling og kontinuerlig implementering. Det som imidlertid er en stor utfordring er at mange selskaper ikke har tilstrekkelige data management strategier (Davenport & Prusak, 1998; Spruit & Pietzka, 2015). Bedrift 1 har en forretningsstrategi og en teknologistrategi, men det kommer ikke frem i vår forskning at de har en egen strategi for master data management.

Bedrift 2 har definert sin datamodell gjennom en domenemodell. Denne modellen er en felles beskrivelse av hvilke data, og relasjoner mellom data, som benyttes på tvers av virksomheten. Domenemodellen danner grunnlaget for felles begreper innad i virksomheten, samt et felles datasett som systemene benytter for deling av informasjon. Det er derimot ikke alt som er beskrevet i datamodellen som er realisert som en datakilde i deres tjenesteplattform, og det som er realisert kan ha en noe annen form i praksis. Her ser vi at Bedrift 2 har skissert hvordan master data skal håndteres, men at det ikke nødvendigvis er slik i praksis. Vi mener en klar og tydelig master data management strategi kan være veien å gå for å tydeliggjøre dette, for begge virksomheter. For å ha en suksessfull master data management prosess må man blant annet ha en strategi for utvikling av master data management (Hikmawati et al., 2021; Vilminko-Heikkinen & Pekkola, 2013). Leverandør sier at de fleste kundene verken har en idé eller en strategi om tjenesteplattform, men at de i noen tilfeller har fokus på digitalisering og digital transformasjon.

Master data strategi skal spesifisere de viktigste prinsippene og retningslinjene som har effekt på beslutninger og aktiviteter når det kommer til *data portfolio*, *data and system design*, *data supply chain* og *data support* (Ofner et al., 2013, s. 480, vår oversettelse). Spesielt ser vi at virksomhetene har mangler innen *data supply chain* som skal lage retningslinjer for opprettelse og vedlikehold av master data, overvåke kvaliteten på master data, samt administrere forespørsler, samle inn datakomponenter, og opprette og distribuere master data.

I praksis har ikke virksomheter noen klare retningslinjer til de som håndterer data når det kommer til dataformat, flyten til registrering av data, hvordan oppdatere data, eller hvem som har ansvaret når det oppstår problemer knyttet til data (Haug & Arlbjørn, 2011; Hikmawati et al., 2021; Silvola et al., 2011; Smith & McKeen, 2008; Vilminko-Heikkinen & Pekkola, 2017). Dette kan føre til at registrering av data blir unøyaktig, overflødig og ikke enhetlig. I Bedrift 2 er det kildesystemene som er eier av master data det er der rettighetene til å oppdatere og vedlikeholde data ligger. Fagsystemene eier forretningsregler om opprettelse, oppdatering og sletting av master data. Master data huben forholder seg til disse oppdateringene fra fagsystemene. Nå som virksomheten distribuerer master data fra en master data hub til flere fagsystem er håndtering av, og eierskap til, master data en utfordring. Siden virksomheten har flere system som er deleier av master data øker også muligheten til feilkilder. Dette er tidkrevende og skaper unødvendig arbeid i noen av avdelingene. Vi mener at både Bedrift 1 og Bedrift 2 setter seg i denne situasjonen selv ved å ha flere informasjonssystem og kildesystem som er deleiere av data. Teorien er tydelig på at dette er en mulig kilde til dårlig datakvalitet som igjen er en konsekvens av dårlig data management og *data governance*. Vi mener en mulig løsning for å bli kvitt deleiere for Bedrift 2, kan være en sentralisert hub. Master data vil i en slik løsning alltid opprettes i det sentrale systemet, noe som sikrer at data opprettes likt og er unik.

Data management er den taktiske utførelsen av virksomhetens *data governance*, slik at IT-ressurser kan garantere og opprettholde datakvalitet, GDPR og korrigerer feil. Effektiv *data governance* kan forbedre kvalitet, tilgjengelighet og integritet til virksomhetens data gjennom å strukturere samarbeid på tvers av virksomheten (Hikmawati et al., 2021). Ingen av virksomhetene i vår studie har klare retningslinjer for hvem som har eierskap til data. Selv om Bedrift 1 har retningslinjer for hvert master data objekt er datakvalitet en tilfeldig konsekvens av eierskapet. Master data ligger lagret i kjernesystemene, som er avhengige av å være oppdaterte og vedlikeholdt korrekt. Dette har vi ikke spurt virksomheten om selv, men som Leverandør viser til vil ikke master data som ligger i kjernesystemene være tilgjengelig under oppdateringer og hendelser. Dersom dette er tilfellet i Bedrift 1 er det et problem. Årsaken til det finnes *policy* per master data objekt, men ikke generelle retningslinjer, mener vi kan ha en sammenheng med lovpålagt krav tilknyttet lagring av data. For eksempel oppgir Bedrift 1 at artikkel master data arkiveres, men ikke hvordan kunde master data håndteres. Her er det knyttet mye mer lovverk til data.

I Bedrift 2 er håndtering og eierskap av master data fordelt mellom flere fagsystemer. Her vil master data huben slite med hvilket fagsystem som distribuerer “*single point of truth*”. Informanten for Leverandør mener dette er komplisert og kan føre til konflikter. Problemer knyttet til hvilke data som er sanne kan løses av master data management verktøy. AI og ML kan si noe om hvilke systemer som sitter med preferanse data, samt logikk for å kunne beregne sannsynligheten for hvilket system som sender *golden record*. Uten mer avansert master data management vil fagsystemene overskrive hverandres data, og master data hub vil ikke kunne skille hva som er korrekt data. Bedrift 2 sier at de har en klar og tydelig datamodell, men vi mener at de mangler god *data governance* som skal sikre datakvalitet, da data overskriver hverandre i master data huben. De poengterer også at systemene deres skal stole på data som kommer fra master data hub, men i slike tilfeller kan ikke dette garanteres.

Bedrift 2 har ifølge vår empiri ingen klar strategi for fremtidig master data management. Vi mener virksomheten har tatt strategisk, taktiske og operasjonelle valg for å endre sin virksomhetsarkitektur, og de har en klar datamodell som i teorien legger til rette for god *data governance* (som igjen bedrer datakvalitet). Leverandør påpeker at de som forvalter master data kanskje kan oppleve at hverdagen blir mer kompleks, sammensatt og krevende når master data kan brukes av flere prosesser og brukere. På den andre siden vil verdien av master data øke fordi den vil være konsis, tidsriktig og korrekt. Dersom man flytter data til en data hub må virksomheter forstå at data forvaltes på vegne av hele organisasjonen, og ikke kun for den enkelte prosessen. Vi mener dette er tilfellet til Bedrift 2 som sliter med at master data er deleid, og at master data hub skaper avhengigheter mellom master data som ligger samlet. Vi opplever ikke at master data pr. i dag er konsis og korrekt, men dette mener Bedrift 2 vil gå seg til over tid ved hjelp av systemstøtte. Master data er imidlertid tidsriktig, og vi vet at oppdaterte master data ligger klar til bruk i master data hub. Dette samsvarer med Leverandør sin filosofi om *default* master data bør være siste versjon av sannheten.

Master data hub er ifølge teorien vår et strategisk valg for virksomheter. Dette mener vi empirien vår også viser, og dermed kan denne teorien beholdes. Bedrift 1 har en strategi om en løsrevet arkitektur, og en master data hub er et steg på vei mot dette. Bedrift 2 sin strategi

er å tilgjengeliggjøre master data i tjenestelaget slik at data er lett tilgjengelig for digitalisering, innovasjon og analyser. Vi mener begge virksomhetene mangler *data governance* som skal føre til god master data management. I tillegg mangler de *data supply chain* som en del av sin master data strategi. En master data hub vil være en del av master data strategi og *data quality management*. For å utnytte de fordelene virksomheten kan oppnå med en master data hub, mener vi det er nødvendig at virksomhetene først og fremst har korrekt *data governance*.

5.3 Kan master data hub være en del av master data lifecycle management?

I vår litteraturstudie gjorde vi oss en antakelse om at master data hub kan være en del av master data lifecycle management. I praksis ser vi også at master data hub kan stille krav til master data gjennom datamodellering, samt design og implementering av datamodell, dataprosess og dataarkitektur. I tillegg skal den også forsyne brukere med master data.

Master data lifecycle management defineres av (Ofner et al., 2013) som:

- Krav til master data
- Master data portefølje
- Design og implementering av en master data modell, -prosess, og -arkitektur
- Hvordan man skal forsyne brukerne med master data
- Støtte brukerne av master data

Leverandør skriver at når det kommer til *master data lifecycle* er det ikke kun snakk om teknologi, det omfatter også informasjonsmodellering, forretningsforståelse, hvordan data anvendes i ulike prosesser osv. Dette er et veldig komplekst fagområde. Når man flytter master data til en master data hub blir data duplisert, noe som naturligvis vil gå ut over lagringskapasiteten til virksomheten. Informanten mener beste praksis er å flytte data som er nødvendig for de scenariene virksomheten mener vil oppstå, ved å gjøre seg noen tanker om hva virksomheten kommer til å trenge i fremtiden. Det er alltid en diskusjon i implementeringen, også med tanke på integrasjoner, om hvilke type data man trenger, hva

data betyr og hvordan skal de flyte. Vedlikehold og opprettelse skjer derfor ikke i datalaget, men datalaget abonnerer på data fra kjernesystemene.

Data and System Design er en viktig del av livssyklusen til master data arkitektur fordi det setter retningslinjer for hvordan master data bør lagres og distribueres, overvåker implementeringen eller forbedringen av applikasjoner og databaser, og vedlikehold av teknisk metadata. Herunder ligger også design av master data modeller, -prosesser, -arkitektur og IT-systemer, for å sikre at IT-systemene er teknisk klare (Ofner et al., 2013). Leverandør sin master data hub er designet for å lagre master data. Denne er bygget på teknologien som ligger i datalaget og infrastrukturen i skyløsningen. Datalaget består av ulike typer teknologier for lagring av data som skal tilgjengeliggjøre data på en enkel måte på tvers av virksomhetens applikasjoner ved hjelp av API-laget. I integrasjonslaget har man et sett med integrasjonsmekanismer som er leverandørens egne produkter som implementeres etter kundens behov. I dette ligger det en logging- og monitoringspakke som gjør det mulig å følge data gjennom arkitekturen (avstamning). Dette gjør at huben inngår som en naturlig del av *master data lifecycle*.

Bedrift 1 mener at domener som kunde, artikkel og leverandør kan gi mening å ha i samme master data hub, fordi de kan være knyttet til hverandre som at en artikkel er knyttet til en spesifikk leverandør. Det er flere master data som ikke har noen forretningsmessig verdi, eller operasjonell verdi, ved å være samlet i en master data hub. Det gir ikke mening å flytte master data objekter inn i samme hub kun for løsningen sin del. Det som gir mening, er å flytte master data med like attributter i samme hub. Vår empiri viser at det skaper flere avhengigheter når man flytter master data inn i samme hub, spesielt om knytningen er transaksjonsdata.

Virksomheten må fremdeles ta hensyn til sin master data portefølje, og ha et forhold til hvilke systemer som skal bruke master data, og dermed hvilke master data det er hensiktsmessig å flytte ut av kjernesystemene. Leverandør bekrefter at en stor del av implementeringen av tjenesteplattformen deres dreier seg om diskusjoner rundt hvilke master data som bør ligge i

master data hub. I de tilfellene der det er gunstig i Bedrift 1 sin optimale arkitektur vil master data hub være en del av *master data lifecycle management*.

I Bedrift 2 lagres master data både i kildesystemet og i master data huben. I og med at oppdatering og vedlikehold av master data skjer i fagsystemet vil ikke master data huben være en del av master data sin livssyklus i dette tilfellet. Master data huben distribuerer data til fagsystemer som har behov for det. I den forstand er master data hub en del av Bedrift 2 sin *master data lifecycle*. Det som mangler her er master data strategi, noe som kan hindre en god livssyklus. De bruker også master data huben som en master data portefølje for master data som er relevant for flere systemer. (Ofner et al., 2013). De mener problemet ligger i at de fleste forretningsdomenene i en virksomhet er interessert i å bruke master data for å dekke sine egne funksjonelle behov, noe som kan resultere i at master data blir tvetydig på tvers av virksomheten

En vanlig utfordring i master data management er manglende kunnskap om livssyklusen til master data på tvers av forretningsenheter, avdelinger, forretningsprosesser, IT-systemer og flere ERP-systemer (Ofner et al., 2013). De mener problemet ligger i at de fleste forretningsdomenene i en virksomhet er interessert i å bruke master data for å dekke sine egne funksjonelle behov, noe som kan resultere i at master data blir tvetydig på tvers av virksomheten. Dette fører ofte til ukoordinerte master data ressurser, og inkonsistent og unøyaktig master data. Forståelse, planlegging og overvåking av master data sin livssyklus sees på som en hjørnestein innen master data management.

Vi mener teorien står seg. Master data hub dekker krav til master data og kan inneholde hele porteføljen til virksomheten etter behov. Her er det viktig at man har et forhold til hvilke data som skal ligge i huben. Master data hub kan også inneholde og bidra til design og implementering av datamodeller, dataprosesser og dataarkitektur. Med utgangspunkt i våre funn kan master data distribueres til alle system og applikasjoner master data hub er koblet til, men uten avansert master data management vil den ikke støtte brukere med økt datakvalitet.

5.4 Hvilke organisatoriske-, forretningsmessige- og IT-systemer berøres av en master data hub?

En master data hub skal håndtere all master data uansett om det skapes i HR-domenet, kundedomenet eller ERP-domenet (Christensen, 2021). Videre skriver han at en master data hub ligger på utsiden av ERP-systemet og blir betraktet som en generisk tjeneste, uavhengig av bransje og prosesser som har behov for master dataene. Bedrift 1 sin arkitektur vil bli påvirket av den nye teknologistrategien, og spesielt med tanke på integrasjoner. Artikkel master data management påvirker flere applikasjoner og applikasjonstjenester, både internt og eksternt i Bedrift 1. Tre eksterne databaser sørger for mesteparten av inputen av artikkel data i virksomheten sin verdikjede. Data fra disse kildene importeres med ulike format og teknikker til master data huben som igjen skal sende data til ERP-systemet. ERP-systemet distribuerer master data videre til virksomhetens systemportefølje. Gjennom master data huben vil artikkel data lett kunne distribueres til andre applikasjoner som trenger nøytralisert artikkeldata, som netthandelsløsningen. En slik arkitektur skal legge til rette for nøyaktighet, tilgjengelighet, aktualitet, konsistens, validitet og fullstendighet i data, og vil samsvare med et *Central Master Data System*:

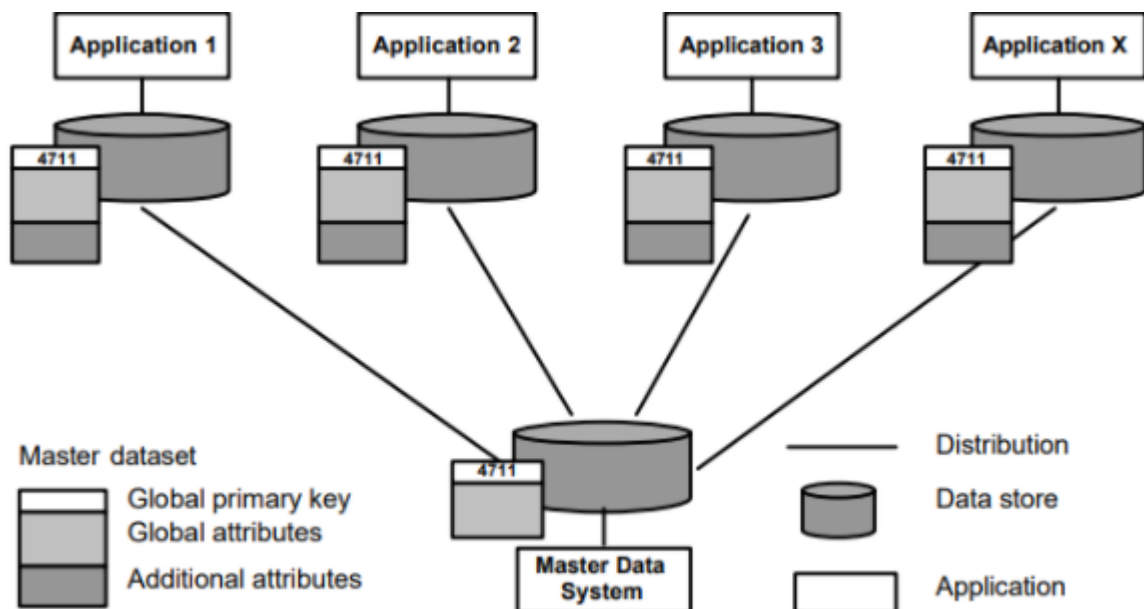
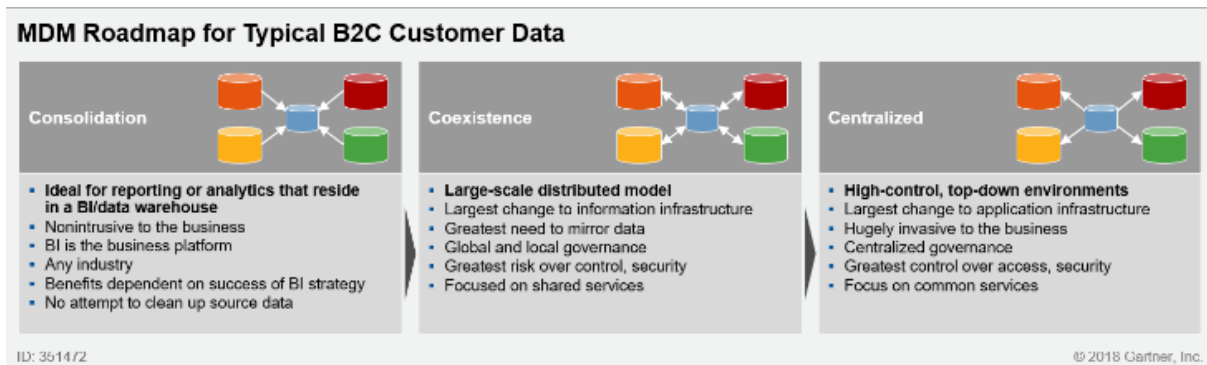


Figure 2: Central Master Data System

Gjenbruk av Figur 3: Central Master Data System (Loser et al., 2004, s. 2)

Og (O’Kane et al., 2018) sin *Centralized* implementeringsstil av master data hub:



Gjenbruk av Figur 7: Best-Practice MDM Implementation Roadmap for Typical B2C Customer Data Scenario (O’Kane et al., 2018, s. 8)

Bedrift 2 har en desentralisert master data hub. Det vil si at all master data som skal benyttes av andre systemer og applikasjoner enn kjernesystemet, ligger i ulike master data huber. Dette er system for HR, Microsoft Dynamics, samt finansielle og operasjonelle systemer. Denne arkitekturen samsvarer med Loser et al. (2004) sitt *Repository*:

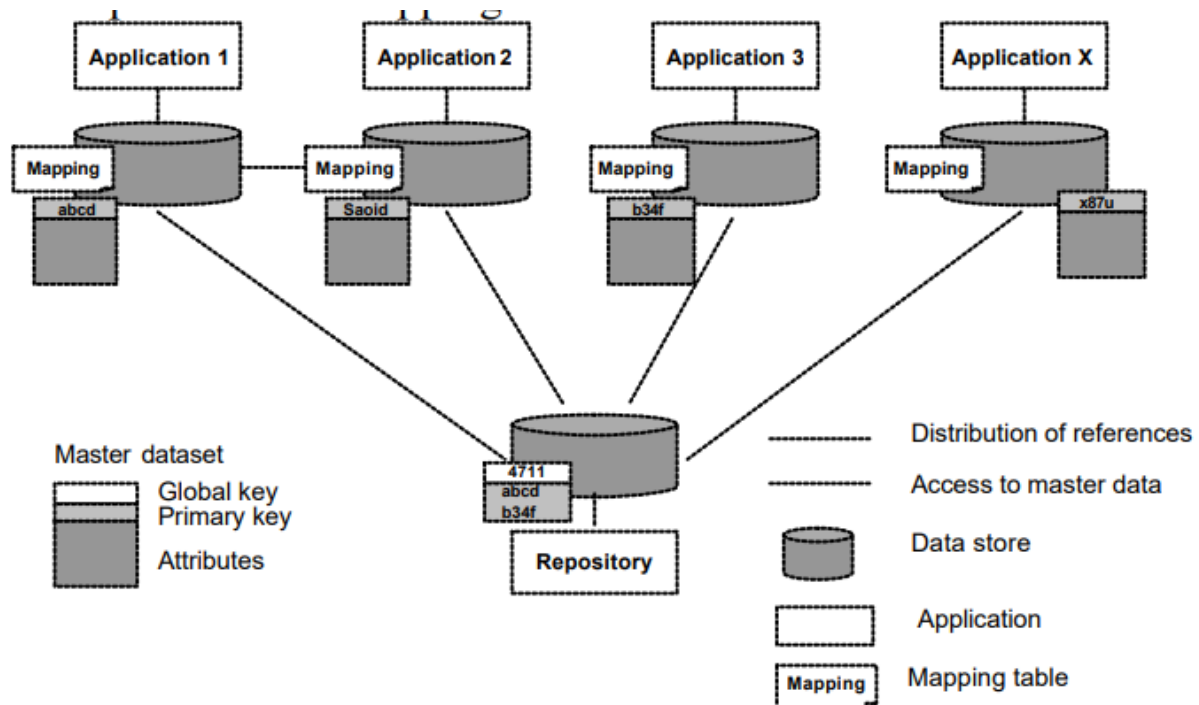


Figure 5: Repository

Figur 4: Repository (Loser et al., 2004, s. 4)

Som Bedrift 2 sier selv vil alle fagsystemene bruke master data, i tillegg til fagsystemene som eier master data, som blir berørt av en master data hub. Virksomheten står også fritt til å koble på de digitale tjenestene de måtte ønske. I Leverandør sin arkitektur for en master data hub berøres alle applikasjoner i tjenestelaget, alle analyseverktøyene i *Intelligence*-laget, og kundens kjernesystemer inkludert *Software as a Service* (SaaS) og *Platform as a Service* (PaaS).

Hvordan virksomheten skal utveksle data må avgjøres på et organisatorisk nivå der prosesser knyttet til harmonisering, *cleansing*, opprettelse, vedlikehold og arkivering av data må defineres (Loser et al., 2004). Det finnes ikke begrensninger for distribusjon av master data. Virksomhetene må ha kunnskap innen modellering og koding (API). I tillegg blir data lettere å få tak i, og bruke til utvikling og innovasjon. Eksempelvis har ikke Bedrift 2 tilgjengeliggjort master data for analyser og rapportering enda, men dette er noe de ser for seg på sikt. Altså blir ikke all master data tilgjengelig før man setter opp integrasjoner.

Litteraturstudien vår viste at forretningsenhetene til virksomheter skal bruke master data som ligger i huben til å utføre sine forretningsprosesser, rapporter og analyser. Denne teorien kan beholdes. Så lenge virksomheten har mulighet til å koble til applikasjoner og systemer gjennom bruk av API og integrasjon, vil master data hub kunne distribuere data. Master data vil også være lagret i en master data hub, enten det ligger som duplikat i en desentral løsning eller en unik verdi i et sentralt system. Vår oppfatning er at dersom kvaliteten på data som er i en master data hub er dårlig, vil data som distribueres fra huben også være dårlig: “*Shit in, shit out*”. Dårlig kvalitet på master data er som tidligere nevnt dyrt, og vil hindre innovasjon og digitalisering.

5.5 Hvilke faktorer påvirker valg av master data hub?

Fra litteraturstudien vår har vi en antakelse om at modenhet innen master data management kan føre til implementering av master data hub. Spruit & Pietzka (2015) laget MD3M-modellen for å analysere organisasjonens modenhet når det kommer til master data

management. Modellen ser på en virksomhets *data model*, *data quality*, *usage & ownership*, *data protection* og *maintenance* for å måle grad av modenhet. *Data model* ser på hvilke data virksomheten definerer som master data, hvordan disse er strukturert, hvilke systemer som bruker master data og hvor master data ligger lagret. Begge virksomhetene har klart å definere hva som er master data for dem, og hvor denne brukes i organisasjonen. Bedrift 1 definerer artikkeldata, kundedata, leverandørdato, ansattdata og de ulike markedssegmentene sine som master data. Bedrift 2 har splittet sine master data inn i domenene produkt, kunder og de ulike markedssegmentene.

Bedrift 1 har behov for å gjøre artikkel master data tilgjengelig for blant annet netthandel. Når data ligger lagret i ERP-systemet er ikke dette mulig, uten omveier, for å migrere denne informasjonen ut til netthandelsløsningen. Dette er ikke i tråd med en løsrevet arkitektur som skal hjelpe med virksomhetens strategiske måloppnåelse. En optimal løsning, ifølge dem selv og som Christensen (2021) oppfordrer til, er en master data hub som ligger på utsiden kjernesystemet som kan koble seg opp mot nødvendige applikasjoner og tjenester. Det fører til at Bedrift 1 kan utføre de aktivitetene som er i tråd med deres strategiske måloppnåelse. Virksomheten ser større nytte av å flytte artikkel master data ut i en hub, enn utfordringene de kommer til å møte i integrasjonen for å sammenstille artikkel og leverandør i de prosessene som krever det. Bedrift 2 identifiserte at de måtte tenke helt nytt angående master data håndtering, da deres kjernesystem skulle fases ut. Dette var i tråd med virksomhetens ønske om å ha en digital tjenesteplattform. Begge virksomhetene har klart å definere hva som er master data for dem, og hvor denne brukes i organisasjonen.

Data quality går ut på om virksomheten har et forhold til hvilken kvalitet master data har, hvordan kvaliteten kan påvirke forretningen, hvor kilden til dårlig master data kvalitet ligger og hvordan virksomheten kan forbedre master data kvaliteten. Dette er viktig da det er datakvaliteten som til syvende og sist gjør det mulig for virksomheter å oppnå konkurransefordeler og redusere feil i forretningsprosesser. Begge virksomhetene har klart å definere hva som er kilden til dårlig datakvalitet, hvor denne ligger og hva de kan gjøre for å forbedre den. Med det mener vi at de vet hvordan utfordringene kan løses, men har ikke noen klar handlingsplan for å oppnå dette. Bedrift 1 har mangelfulle master data i all data bortsett

fra leverandørdata. Bedrift 2 sliter med duplikater i sin master data hub, men tror at systemstøtte for akkurat dette skal kunne løse problemet på sikt.

Usage & ownership gir virksomheten muligheten til å definere hvem som bruker data i hvilke system, hvem har lese/skrive tilganger og om man har klare retningslinjer knyttet til hvem som har disse autorisasjonene. Bedrift 1 har uklare retningslinjer når det kommer til eierskap av master data, og som en følge av dette kan også prosesser for vedlikehold, standarder og fremtidige forbedringer være uklare. Samtidig ser vi at de søker forbedringer innen master data når det kommer til artikler, markedssegment og ansatte (*MDMS, sammenstilling av datagrunnlag for markedssegment og ansatte*). Bedrift 2 har som nevnt en klar og tydelig datamodell som dekker dette området. Dette er i tråd med kravene til datamodellering som ligger i tjenesteplattformen de benytter. Virksomheter må huske at data rives løs fra autorisasjonsmodeller når de flyttes ut av kjernesystemene. Leverandør påpeker at virksomheter må tenke på hva som skal være allment tilgjengelig og hva som skal ligge i kjernesystemet. Spesielt viktig er dette i forhold til kundedata som skal være beskyttet gjennom GDPR. Informanten trekker frem at et interessant aspekt med dette er at man kan unngå for eksempel lisenser til CRM-systemet for å vise data til andre grupper med behov i virksomheten ved å tilgjengeliggjøre data i tjenesteplattformen. Relevant kundedata kan derfor bli tilgjengelig for flere ansatte.

Data protection handler om hvor og hvordan data er sikret mot mulige forhold som svikt i komponentene, software bugs eller spesifikke angrep. Bedrift 1 sikrer data gjennom tilgangsstyring på brukernivå, redundante løsninger og sikring av virksomhetens tekniske infrastruktur. Redundante løsninger skaper høy pålitelighet ved at to eller flere systemer jobber parallelt. På denne måten sikrer virksomheten at systemene “alltid” vil være operative. I Bedrift 2 er hvert fagsystem ansvarlig for å sikre sine data, og virksomheten har for vane å kjøre sikkerhetstester.

Maintenance handler om fysisk lagring og livssyklusen til master data. For å kunne dra nytte av de tekniske innovasjonene er man avhengig av at systemet er oppdatert og vedlikeholdt korrekt, samt at data som er i bruk må holdes ren ved å arkivere eller slette utdaterte data i

henhold til livssyklusen. Bedrift 1 lagrer master data i sine kjernesystemer for ERP og CRM, i tillegg er det her master data opprettes, vedlikeholdes og distribueres. I deres optimale løsning vil master data opprettes og vedlikeholdes i databaser, før de importeres til MDMS. I MDMS skal data være validert, reformatert og klar til distribusjon. Bedrift 2 lagrer sine master data i kildesystemet og i master data hub, noe som vil gå utover deres lagringskapasitet. Opprettelse og vedlikehold av master data skjer i fagsystemene, som sender informasjonen til master data hub. Leverandøren sin arkitektur legger også opp til at det er kjernesystemet som eier og vedlikeholder master data, og at det er kundene som skal modellere og forvalte data de bruker.

Bedrift 1 jobber mot en løs kobling i arkitekturen, som skal gjøre at systemene mindre avhengige av hverandre og mener et eget lagringssted for master data kan være en god løsning. Derimot kan det å samle all master data gi en motsatt effekt ved å skape avhengigheter. Dette er i samsvar med Destein (2007) som mener at virksomheter vil støte på utfordringer ved å sammenstille data på tvers av ulike datamodeller, system og eksterne datakilder. Bedrift 1 jobber med å tilgjengeliggjøre artikkel master data i en master data hub. Dette vil gi virksomheten en autoritativ kilde til master data som ikke er tett knyttet til et spesifikt fagsystem, som logistikk eller økonomi. Bedrift 1 ser større verdi av å løfte artikkel master data ut i en master data hub, enn det koster av ressurser for å få på plass integrasjonen som skal sammenstille data. Faktoren for å tilgjengeliggjøre master data for Bedrift 1 er i dette tilfellet en løsrevet arkitektur.

Årsaken til at Bedrift 1 søker en løsning for master data håndtering av artikkel master data er utdatert i den forstand at den ikke lenger støttes av support, samt at den ikke fungerer optimalt. Teknisk sett har de en løsning i dag som fungerer som et *repository* for artikkel, men i realiteten er det ERP-systemet som fungerer som den autoritative kilden til artikler. I tillegg har ikke dagens *repository* kapabilitetene virksomheten ønsker fra en master data management løsning. Den er ikke fleksibel nok i distribusjon av master data, og den mangler verktøy for å orkestre litt mer avanserte valideringer av mange datastrukturer (datamodeller) samtidig. Da det kun er artikkel master data som skal flyttes ut i en nøytralisert løsning vil det føre til utfordringer og belastninger for integrasjonen til virksomheten, fordi master data objektene artikkel og leverandør ligger lagret på ulike steder.

Bedrift 1 har valgt å fokusere på å sammenstille objektene ved behov, i stedet for å flytte de inn i den samme løsningen. Informanten mener at de ved å flytte både leverandør master data og artikkel master data over i den samme nye løsningen vil virksomheten skape nye avhengigheter.

Da fagsystemet til Bedrift 2 ble faset ut oppstod det et behov for å håndtere master data på en annen måte. I tiden da fagsystemet var kilden til master data opplevde virksomheten duplikater og inkonsistens. Dette ble håndtert direkte i systemet, og opplevdes ikke som noe stort problem. Dette var ikke en årsak til implementering av master data management, og virksomheten opplever flere duplikater og feil i datagrunnlaget nå, når master data distribueres fra master data hub. Vår observasjon er at dårlig datakvalitet i en virksomhet ikke direkte fører til master data management tiltak.

Selv om Bedrift 1 og Bedrift 2 mener de vet hva kilden til dårlig datakvalitet er, opplever vi at begge virksomhetene velger å jobbe rundt dårlig datakvalitet, eller gå for *quick fixes*, i stedet for å løse kilden til problemet. Bedrift 2 opplever at ting har blitt mer komplisert når data er fordelt på flere system, og informanten trekker frem at avhengighetene har økt i takt med kompleksiteten. Samtidig har systemstøtte for brukerne (duplikatsjekk) økt, og Bedrift 2 forventer at kvaliteten i master data vil øke over tid. Leverandør mener tjenesteplattformen vil føre til mer kompleks forvaltning da master data skal brukes av alle i virksomheten.

Vår forskning viser at dårlig datakvalitet ikke er en direkte faktor for valg av master data hub. Vi har fra litteraturstudien gjort en antakelse om at virksomheter som er modne innen master data management vil implementere en master data hub. I praksis ser vi at det er utfasing av gamle systemer og gammel integrasjon som tvinger virksomheter over på nye løsninger, og at dette ikke nødvendigvis har noe med master data management å gjøre. I tillegg er det utfasing av systemer som har mistet support som er en faktor for begge virksomheter. Bedrift 2 har fått en master data hub via overgangen fra et gammelt fagsystem til tjenesteplattform. Vi mener ønsket om en master data håndtering fra deres side handlet mer om innovasjonstakten enn om datakvalitet.

Leverandør mener dersom kvaliteten i master data er god kan den migreres rett ut i applikasjoner og tjenester som ligger i tjenestelaget, samt til analyser og rapportering. Feil i master data kan resultere i signifikante kostnader eller eventuelt inntektsfall for virksomheten. Årsaker til feil i eller inkonsistent master data kan være systemarkitektur, utilstrekkelig koordinering med forretningsprosesser, utilstrekkelig implementering av programvare eller uoppmerksom brukeradferd (Haug & Arlbjørn, 2011). Loshin (2009) mener det er flere fordeler knyttet til master data management, blant annet konsistent rapportering, forbedret konkurranseevne, forbedret beslutninger, og bedre kostnadsanalyser og kostnadsplanlegging.

Virksomheter må være nøyaktig i valg av leverandør av master data hub, og de må velge en leverandør som kjenner bransjen de jobber i. Mange master data management løsninger er spesialiserte på de horisontale kravene virksomheten har, men ikke nødvendigvis de vertikale (Beasty, 2008). Her er de horisontale kravene virksomhetens strategi og de vertikale kravene leverandørens produkt. Dersom vi skal se dette opp mot valget Bedrift 2 har gjort av leverandør, ser vi at de ikke satte krav til avansert master data management, men til en enkel master data hub som er en del av en tjenesteplattform. De sier også selv at det var den digitale tjenesteplattformen de var ute etter. Leverandør har valgt å levere en enkel master data management løsninger fordi det er ikke alle kunder som krever eller trenger noe mer enn distribusjon av master data. Noen kunder har god kompetanse innen informasjonsmodellering og sterkt datavarehus, mens andre har lite kompetanse og trenger å bli forklart hvordan de skal forholde seg til datamodellering. De mindre kundene er ifølge informanten ikke veldig opptatt av hvordan plattformen fungerer i detalj, men hva man kan få ut av tjenestene på toppen som rapporter, mobilapper og de funksjonelle behovene.

Vi mener derfor at det ikke er teknisk kompetanse som ligger til grunn for Bedrift 2 sitt valg av master data hub, men heller mulighetene som kommer i tjenestelaget. Det er her konkurransefortrinnet ligger da master data er tilgjengelig for tjenestelaget til enhver tid, og innovasjon blir enklere når man har muligheten til å prøve og feile med løsninger. Leverandøren mener det er kunder som har flere *systems of entries* som krever mer avansert

master data management. Typisk kan dette være ved oppkjøp der man ikke ønsker å slå sammen kjernesystemene. Vilminko-Heikkinen & Pekkola (2013) mener det er spesielt problematisk når virksomheter lagrer samme data i flere ulike informasjonssystem, som en konsekvens av at data har blitt utviklet og lagret i siloer over de siste tiårene. Master data management, og bruk av master data hub, er en løsning til dette problemet og skal hjelpe med å avdekke data som ligger inne i siloene. De som har kraft til å investere i arkitektur og som har et visst aktivitetsnivå til å utvikle digitale tjenester er typiske kunder til Leverandør.

Gjennom vår litteraturstudie kom vi frem til at det er virksomheter som er modne innen master data management som ender opp med å ta i bruk master data hub. Dårlig datakvalitet fører ikke til direkte tiltak for bedring av eller implementering av master data management løsninger. Av vår forskning fremkommer det at faktorer for valg av master data hub er tilgjengeliggjøring av data til netthandel, utgående support og utfasing av system. Ut ifra vår forskning må teorien modifiseres.

5.6 Hvem leverer denne type teknologi og hvilke tjenester/kvalitet er bygget inn i disse, der man går ett steg lenger?

Styrker i en leverandørs strategi vil være at de har *roadmap* for innovasjon, partnerskap, visjon for produktet, *roadmap* for utførelse, teknisk support og profesjonelle tjenester. Styrker til en leverandørs produkt vurderes av *matching*, kobling og oppløsningen til *entities*; datakvalitet, data kontekst, sammenstilling, *data governance*, dataintegrasjon, maler til forretning, distribusjon, avstamning, sikkerhet, HA/DR, skalerbarhet, domener og støtte for flere domener (Yuhanna et al., 2021). Rapporten trekker fram datakvalitet, dataintegrasjon og datasikkerhet som de viktigste aspektene ved leverandørens produkt.

Bedrift 1 sin optimale arkitektur for en master data hub legger til grunn at huben skal tilpasse seg eksisterende datamodeller i databasene, uten tap av struktur og bredde. Dette krever sofistikert datamodellering fra leverandøren som skal sammenstille alt til *one single truth*.

Løsningen krever nye mønster for integrasjon som legger til rette for eksisterende og fremtidige egenskaper i databasene, og leverandører av master data hub må derfor være tilpasningsdyktig når det kommer til integrasjonsmønstre. Dette trekker også informantene vår for Leverandør frem som et viktig evalueringskriterium. Bedrift 1 skal unngå transformasjon av data i integrasjonsnivået så langt det er praktisk mulig.

Da Bedrift 2 skulle velge en leverandør var det viktig at de skulle få en samarbeidspartner som hadde erfaring med digital tjenesteplattform. Kriteriene for valg av leverandør var kunnskap rundt skytjenester, digital tjenesteplattform, master data håndtering, og kunne stille med flere konsulenter som jobber sammen med og overfører kunnskap til virksomhetens ansatte. Selv trekker virksomheten frem at det viktigste kriteriet til leverandør var kunnskap om teknologi og metodikk som kunne støtte realiseringen av domenemodellen.

Teknologien og tjenestene Leverandør leverer er tilgjengeliggjøring av informasjon til det digitale tjenestetilbudet. De har et datalag med teknologier, lagring og distribusjon av data, som deres master data hub er bygget på. Integrasjonslaget deres ekstraherer og modellerer data, men det er kunden selv som må kode denne modelleringen. For å sikre datakvalitet integrerer leverandører innsikt for å optimalisere og automatisere master data management og forretningsens *data governance* (Parker et al., 2021). Kunden selv står for datamodelleringen tjenesteplattformen krever. Dette gjør plattformen tilpasningsdyktig til kundens behov, men gjør også at den ikke inneholder funksjoner som AI og ML, med mindre kunden modellerer dette selv. Informanten mener virksomheter bør ha evalueringskriterier som pris- og supportmodell, integrasjonsvennlighet og integrasjonsmønstre, og brukervennlighet. Brukervennlighet og brukergrensesnitt er viktige aspekter ved master data hub og analyser. Ofte er det brukere med kompetanse innen forretning og data som skal sitte med selvbetjeningsverktøyene master data huben legger til rette for. Dermed blir dette løserevet fra den tradisjonelle IT-flyten, der det er IT-avdelingen som eier hele verdikjeden fra ETL til kjernesystemet, og til rapportene ligger klare på dashbordet.

Formålet med dette forskningsspørsmålet var å få innsikt i teknologien en master data hub er bygget på. Vi oppfatter at manglende kunnskap om master data hub gjør det vanskelig for

virksomheter å ta stilling til hva de skal kreve av en hub, og hvilke muligheter en hub skaper. Behovene til virksomheter er ikke generelle, men individuelle, noe som gjør det vanskelig å gi et klart svar på hvilke behov master data hub skal dekke. En master data hub fører til at virksomheter kan løsrive rapportering og analyser fra IT-avdelingen, men modellering og integrasjoner fører til at man trenger intern eller ekstern teknisk kompetanse.

6. Konklusjon

Formålet med denne studien er å utrede om bruk av master data hub bidrar til økt datakvalitet. Vår tilnærming til denne studien har vært kvalitativ. Forskningsdesignet er en casestudie, der vi har samlet inn empiri gjennom semistrukturerte intervju og elektronisk innsamling. Utvalgsstørrelsen vår er hensiktsmessig med flere informanter, og har gitt oss mye informasjon angående caseobjektet vårt, master data hub. Vår utvalgsstrategi var å finne informanter innen målgruppe “kunde” og leverandør” av master data hub.

For å kunne svare på denne problemstillingen utarbeidet vi seks forskningsspørsmål som i denne studien besvares gjennom teori og empiri. Teorien baserer seg på en litteraturstudie som består av 38 fagartikler, fagbøker og uavhengige rapporter forsknings- og rådgivningsselskap. Fra denne litteraturstudien gjorde vi oss følgende antakelser:

- Master data management kan bidra til økt datakvalitet
- Datakvalitet måles både subjektivt og objektivt
- Jo høyere kvalitet man har i master data, jo bedre er det for virksomheten
- Virksomheter må innarbeide master data management i sin strategi
- Master data management er en taktisk beslutning som skal garantere og opprettholde datakvalitet
- Master data hub kan være en del av *master data lifecycle management*
- Alle applikasjoner og systemer som bruker master data berøres av master data hub
- Modenhet innen master data management påvirker valg av master data hub
- Avansert master data management som AI og ML sørger for at leverandører er “leaders” i markedet

Antakelsen vår om at master data management kan bidra til økt datakvalitet støttes av vår empiri. I vår forskning er master data hub en enkel komponent i en helhetlig tjenesteplattform, som ikke har avansert master data management innebygd. Vår empiri viser at en master data hub kan bidra til økt datakvalitet. Samtidig gjør den hverdagen til brukere mer kompleks. Videre skapes det avhengigheter i master data objekter om virksomheten ikke er nøye med hvilke master data de samler i samme hub.

Det at datakvalitet måles både subjektivt og objektivt fører ifølge vår forskning til at det er vanskelig for virksomhetene å si hva som er god datakvalitet for dem. Teori og empiri samsvarer når det gjelder antakelsen om at høyere kvalitet i master data er bedre for virksomhetene. Dette legger til rette for digitalisering, innovasjon og god beslutningsstøtte.

Vår empiri viser at antakelsen om at virksomheter må innarbeide master data management i sin strategi er viktig for at de skal kunne håndtere master data på en god måte. Begge virksomhetene i vår studie ønsket seg bedre kvalitet innen master data, men ingen av dem hadde en klar strategi for hvordan oppnå dette. Empirien vår viser at fravær av klare og tydelige retningslinjer tilknyttet eierrollen for master data gjør det utfordrende for virksomheter å garantere og opprettholde god datakvalitet.

Gjennom litteraturstudien vår gjorde vi oss en antakelse om at master data hub kunne være en del av *master data lifecycle management*, plassert under systemdesign innen master data strategi. Vår forskning viser at det master data hub brukes til å forsyne brukere med master data, samt designe og implementere datamodell, -arkitektur og -prosesser. Det er opp til brukeren i hvilken grad de benytter seg av funksjonene.

Antakelsen vår om at alle applikasjoner og systemer som bruker master data berøres av en master data hub stemmer. Virksomheter løfter master data de skal bruke nå og i overskuelig fremtid opp i huben, og distribuerer denne til systemer og applikasjoner som trenger den.

Ifølge vår studie stemmer det ikke at modenhet innen master data management påvirker valg av master data hub. Vår forskning viser at utdaterte systemer og integrasjoner gjør at virksomheter søker etter nye plattformer eller verktøy til å håndtere sine master data. Vi har ikke funnet empiri som understøtter denne antakelsen.

Litteraturstudien vår legger til grunn en korrelasjon mellom AI og ML og god datakvalitet. Dette er tjenester “*leaders*” i markedet har innebygd i sin master data management, og tilbyr

kundene sine. Vår forskning kan ikke bekrefte eller avkrefte denne antakelsen, da AI og ML ikke er innebygd i tjenesteplattformen vi har studert, og virksomheten som bruker master data hub ikke har implementert slike avanserte master data management verktøy.

Vilminko-Heikkinen & Pekkola (2013) mener det er et stort problem at data ligger lagret i ulike informasjonssystem. Datakvalitet er et stort problem for virksomheter som har informasjonssystem fordelt på flere forskjellige enheter og avdelinger. Slik arkitektur legger til rette for problemer i datakvalitet som duplikater og inkonsistens. Master data management kan løse problemer knyttet til dårlig datakvalitet forårsaket av data som kommer fra ulike informasjonssystem, gjennom prosesser som *data profiling* og standardisering (Hikmawati et al., 2021). Christensen (2021) mener en desentralisert master data hub kan være med å løse problemer knyttet til dette. AI og ML er avansert master data management som kan bidra til økt datakvalitet. Da vi startet dette forskningsprosjektet visste vi midlertidig ikke om implementering av master data management, i form av en master data hub, kunne bidra til økt datakvalitet. Vi ønsket derfor studere dette nærmere.

Denne forskningen omhandler større og komplekse virksomheter, da vi i utgangspunktet mente det ville være hensiktsmessig å studere disse fordi slike virksomheter mest sannsynlig hadde et forhold til sine master data og master data management. Vi mener dette ikke er tilfellet, og at utvalget vårt også kunne inneholdt mindre virksomheter med lavere krav til master data management. Flere informanter innen samme virksomhet kunne styrket vår empiri gjennom å bekrefte eller avkrefte påstander som kommer fra personer innen samme virksomhet. Utvalget vårt inneholdt ikke brukere eller leverandører av master data management som inkluderer AI og ML.

Vår forskning viser at master data management må tilpasses den enkelte virksomheten. Uavhengig av om en virksomhet er moden eller ikke innen master data management, kan vi ikke konkludere med at master data hub i seg selv vil kunne sikre datakvalitet. Derimot vil master data hub kunne bidra til økt datakvalitet i dimensjonene nøyaktighet, konsistens og validitet. *Data governance* skal garantere for og opprettholde datakvalitet. Det fremkommer av vår forskning at uavhengig av lagringssted for master data vil ikke kvaliteten øke uten

korrekt *data governance*. Master data management i seg selv reduserer ikke problemer som duplikater og inkonsistens i data. Med utgangspunkt i våre funn vil det være naturlig å forske videre på om avansert master data management, som AI og ML, kan bidra til økt datakvalitet i flere dimensjoner. Det vil også være naturlig å forske videre på hvorfor hverdagen til brukere blir mer kompleks ved implementering av master data hub.

Referanser

- Beasty, C. (2008). The master piece. *CRM magazine*, 12(1), 39–42.
- Borghoff, U., & Pareschi, R. (1998). *Information Technology for Knowledge Management*.
<https://doi.org/10.1007/978-3-662-03723-2>
- Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward information infrastructure studies: Ways of knowing in a networked environment. I *International handbook of internet research* (s. 97–117). Springer.
- Brocke, J. vom, Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature search process*.
- Buffenoir, E., & Bourdon, I. (2013). Managing extended organizations and data governance. I *Digital Enterprise Design and Management 2013* (s. 135–145). Springer.
- Chien, M., & Jain, A. (2021). *Magic Quadrant for Data Quality Solutions*.
<https://www.gartner.com/doc/reprints?id=1-27L65HH8&ct=211004&st=sb>
- Christen, P. (2012). *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer.
- Christensen, B. H. (2021). *Forretningsutvikling og digitalisering*.
- DAMA, U. (2013). The six primary dimensions for data quality assessment. *DAMA UK*, October.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.
- Destein, M. (2007). The Master Data Hub Solution. *Securities industry news*.
- English, L. P. (1999). *Improving data warehouse and business information quality: Methods for reducing costs and increasing profits*. John Wiley & Sons, Inc.

- Haug, A., & Arlbjørn, J. S. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, 24(3), 288–303.
<https://doi.org/10.1108/17410391111122862>
- Hikmawati, S., Santosa, P., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 5, 90.
<https://doi.org/10.22146/ijitee.66307>
- Johannessen, A., Christoffersen, L., & Tufte, P. A. (2011). *Forskningsmetode for økonomisk-administrative fag* (3. utg.). Abstrakt forl.
- Joshi, A. (2007). MDM governance: A unified team approach. *Cutter IT Journal*, 20(9), 30.
- Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM*, 45(4), 184–192.
- Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006). *Journey to data quality*. The MIT Press.
- Lee, Y. W., Strong, D. M., Wang, R. Y., & Pipino, L. L. (1998). Manage your information as a product. *Sloan Management Review*, 95–105.
- Lerche, S. (2014). *Achieving customer data integration through master data management in enterprise information management*. University of Johannesburg (South Africa).
- Loser, C., Legner, C., & Gizanis, D. (2004). *Master data management for collaborative service processes*.
- Loshin, D. (2009). *Master data management*. Elsevier/Morgan Kaufmann.
- Morbey, G. (2013). *Data Quality for Decision Makers: A dialog between a board member and a DQ expert*. Springer Science & Business Media.
- Ofner, M., Straub, K., Otto, B., & Oesterle, H. (2013). Management of the Master Data Lifecycle: A Framework for Analysis. *Journal of Enterprise Information*

Management, 26, 472–491. <https://doi.org/10.1108/JEIM-05-2013-0026>

- O’Kane, B., Moran, M., & Walker, S. (2018). Accelerate Business Value by Using Gartner’s MasterData Management Implementation Styles. *Gartner*, 1–14.
- Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, 29(1), 3.
- Parker, S., Hawker, M., & Walker, S. (2021, desember 6). *Magic Quadrant for Master Data Management Solutions* [Report]. Gartner.
<https://www.gartner.com/doc/reprints?id=1-28X0POPI&ct=220127&st=sb>
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211–218.
- Ryen, A. (2002). *Det kvalitative intervjuet: Fra vitenskapsteori til feltarbeid*. Fagbokforlaget.
- Schäffer, T., & Leyh, C. (2017). *Master Data Quality in the Era of Digitization—Toward Inter-organizational Master Data Quality in Value Networks: A Problem Identification* (s. 113). https://doi.org/10.1007/978-3-319-58801-8_9
- Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011). Managing one master data—challenges and preconditions. *Industrial Management & Data Systems*.
- Smith, H. A., & McKeen, J. D. (2008). Developments in practice XXX: master data management: Salvation or snake oil? *Communications of the Association for Information Systems*, 23(1), 4.
- Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, 51, 1068–1076.
<https://doi.org/10.1016/j.chb.2014.09.030>

- Umar, A., Karabatis, G., Ness, L., Horowitz, B., & Elmagardmid, A. (1999). Enterprise data quality: A pragmatic approach. *Information Systems Frontiers*, 1(3), 279–301.
- Vilminko-Heikkinen, R., & Pekkola, S. (2012). *Organizational Issues in Establishing Master Data Management Function*. 1–13.
- Vilminko-Heikkinen, R., & Pekkola, S. (2013). *Establishing an organization's master data management function: A stepwise approach*. 4719–4728.
- Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management and its organizational implementation: An ethnographical study within the public sector. *Journal of Enterprise Information Management*.
- Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, 47, 76–87.
- Xu, H., Nord, J. H., Brown, N., & Nord, G. D. (2002). Data quality issues in implementing an ERP. *Industrial management & data systems*.
- Yin, R. K. (2007). *Fallstudier: Design och genomförande*. Liber.
- Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, 19(3), 321–332.
- Yuhanna, N., Leganza, G., & Monteverde, K. (2021, november 21). *The Forrester Wave™: Master Data Management, Q4 2021* [Report]. Forrester.
<https://reprints2.forrester.com/#!/assets/2/87/RES176577/report>

Vedlegg

Vedlegg 1 - Oversikt over vårt litteratursøk

	Master data	Master data hub/Master data platform	Master data management	Data Quality	Data Quality Management
Antall treff	4 980 000	295 000	4 400 000	7 420 000	6 070 000
Antall treff på review artikkel	4 100 000	192 000	3 290 000	3 700 000	3 690 000
Beasty (2008). The master piece. CRM magazine		x	x		
Borghoff & Pareschi (1998). Information Technology for Knowledge Management			x		
Bowker et al. (2009). Toward information infrastructure studies			x		
Buffenoir & Bourdon (2013). Managing extended organizations			x		
Chien & Jain (2021). Magic Quadrant for Data Quality Software			x	x	x
Christen (2012). Data matching: Concepts and techniques for				x	
Christensen (2021). Forretningutvikling og digitalisering.		x	x		
DAMA, U. (2013). The six primary dimensions for data quality					x
Davenport & Prusak. (1998). Working knowledge: How organizations manage what they know			x		
Destin (2007). The Master Data Hub Solution	x				
English (1999). Improving data warehouse and business information systems				x	
Haug & Arlbjørn (2011). Barriers to master data quality	x			x	
Hikmawati et al. (2021). Improving Data Quality and Data Governance			x	x	x
Joshi (2007). MDM governance: A unified team approach			x		
Kahn et al. (2002). Information quality benchmarks: Product data				x	
Lee et al. (2006). Journey to data quality				x	
Lee et al. (1998). Manage your information as a product			x		
Lerche (2014). Achieving customer data integration through data integration			x	x	
Loser et al. (2004). Master data management for collaborative business	x		x		
Loshin (2009). Master data management	x		x		x
Morbey (2013). Data Quality for Decision Makers: A dialog	x			x	
Ofner et al. (2013). Management of the Master Data Lifecycle	x		x		
O'Kane et al. (2018). Accelerate Business Value by Using Cloud			x		
Otto (2011). Organizing data governance: Findings from the					x
Parker et al. (2021, desember 6). Magic Quadrant for Master Data Management			x		
Pipino et al. (2002). Data quality assessment				x	
Schäffer & Leyh (2017). Master Data Quality in the Era of Information				x	
Silvola et al. (2011). Managing one master data—challenges and opportunities					x
Smith & McKeen (2008). Developments in practice: Master data management					x
Spruit & Pietzka (2015). MD3M: The master data management	x		x		
Umar et al. (1999). Enterprise data quality				x	
Vilminko-Heikkinen, R., & Pekkola, S. (2012). Organizational data quality					x
Vilminko-Heikkinen, R., & Pekkola, S. (2013). Establishing data quality	x		x		
Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management					x
Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in data quality					x
Xu et al. (2002). Data quality issues in implementing an ERP system					x
Yuhanna et al. (2021). The Forrester Wave™ Master Data Management			x		

Vedlegg 2 - Data Quality Dimensions, vurdering av datakvalitet

Table 1. Data quality dimensions.

Dimensions	Definitions
Accessibility	the extent to which data is available, or easily and quickly retrievable
Appropriate Amount of Data	the extent to which the volume of data is appropriate for the task at hand
Believability	the extent to which data is regarded as true and credible
Completeness	the extent to which data is not missing and is of sufficient breadth and depth for the task at hand
Concise Representation	the extent to which data is compactly represented
Consistent Representation	the extent to which data is presented in the same format
Ease of Manipulation	the extent to which data is easy to manipulate and apply to different tasks
Free-of-Error	the extent to which data is correct and reliable
Interpretability	the extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	the extent to which data is unbiased, unprejudiced, and impartial
Relevancy	the extent to which data is applicable and helpful for the task at hand
Reputation	the extent to which data is highly regarded in terms of its source or content
Security	the extent to which access to data is restricted appropriately to maintain its security
Timeliness	the extent to which the data is sufficiently up-to-date for the task at hand
Understandability	the extent to which data is easily comprehended
Value-Added	the extent to which data is beneficial and provides advantages from its use

Table 1. Data quality dimensions (Pipino et al., 2002, s. 212)

Vedlegg 3 - Informasjonsskriv til informanter, meldt inn til NSD

Vil du delta i forskningsprosjektet

”Kan bruk av master data hub bidra til økt datakvalitet”?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å finne ut om master data hub kan bidra til bedre datakvalitet, og dermed også analyser. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Dette er en masteroppgave i Digital ledelse og business analytics, der formen på oppgaven er master of science. Omfanget er 30 studiepoeng.

Problemstilling:

- **Kan bruk av master data hub bidra til økt datakvalitet, og dermed også bedre analyser**

Forskningsspørsmål:

- Hvilke strategiske, taktiske og operasjonelle valg må en bedrift ta for å ha god master data management?
- Kan master data hub være en del av master data lifecycle management?
- Hva er (god) datakvalitet?
- Hvilke faktorer påvirker valg av master data hub?
- Hvilke organisatoriske-, forretningsmessige- og IT-systemer berøres av en master data hub?
- Hvem leverer denne type teknologi og hvilke tjenester/kvalitet er bygget inn i disse, der man går ett steg lenger?

Hvem er ansvarlig for forskningsprosjektet?

Høgskolen i Innlandet er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Du får spørsmål om å delta som informant i denne oppgaven fordi du sitter på verdifull informasjon om master data og arkitektur.

Du får spørsmål om å delta som informant i denne oppgaven fordi du sitter på verdifull informasjon om tjenesteplattformer (inkl. Master data hub) og arkitekturen og teknologien bak en slik plattform.

Hva innebærer det for deg å delta?

Semistrukturerte intervju; vi forbereder noen spørsmål og samtalepunkter til møtet, og vi diskuterer fritt rundt disse og andre tema i samtalene. Intervjuene blir tatt opp ved hjelp av skjermopptak på teams med godkjenning i forkant. Dersom det lar seg gjøre kan enkelte spørsmål følges opp på mail.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Det er kun vi som skriver oppgaven som har tilgang til opptakene som blir tatt
- Både du og din bedrift er anonymisert i våre notater og i oppgaven.

Opplysninger om deg som person eller din bedrift skal ikke være gjenkjennelig i vår oppgave. Du vil få tilgang til å lese gjennom oppgaven før den sendes til sensur og eventuelt publiseres.

Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?

Prosjektet vil etter planen avsluttes 15.06.2022.

Etter prosjektslutt vil alt materiale knyttet til deg og din bedrift slettes.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Høgskolen i Innlandet har Personverntjenester vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Høgskolen i Innlandet ved Christoffer Aakervik Tangen; 238338@stud.inn.no / Therese Thorvaldsen; 238336@stud.inn.no / Muhammad Yahya; muhammad.yahya@inn.no (veileder for prosjektet).
- Vårt personvernombud: anne.lofthus@inn.no

Hvis du har spørsmål knyttet til Personverntjenester sin vurdering av prosjektet, kan du ta kontakt med:

- Personverntjenester på epost (personverntjenester@sikt.no) eller på telefon: 53 21 15 00.

Med vennlig hilsen

Muhammad Yahya

Prosjektansvarlig

Christoffer Aakervik Tangen

Masterstudent

Therese Thorvaldsen

Masterstudent

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet [*sett inn tittel*], og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i [semistrukturerte intervju]
- å delta i [spørsmål på mail]

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Signert av prosjektdeltaker, dato)

Vedlegg 4 - Første intervjuguide til virksomhetene i casestudien

Intervjuguide til Bedrift 1 og Bedrift 2: MD3M Modellen

Dette gjelder fem nøkkeltema: *data model, data quality, usage & ownership, data protection og maintenance.*

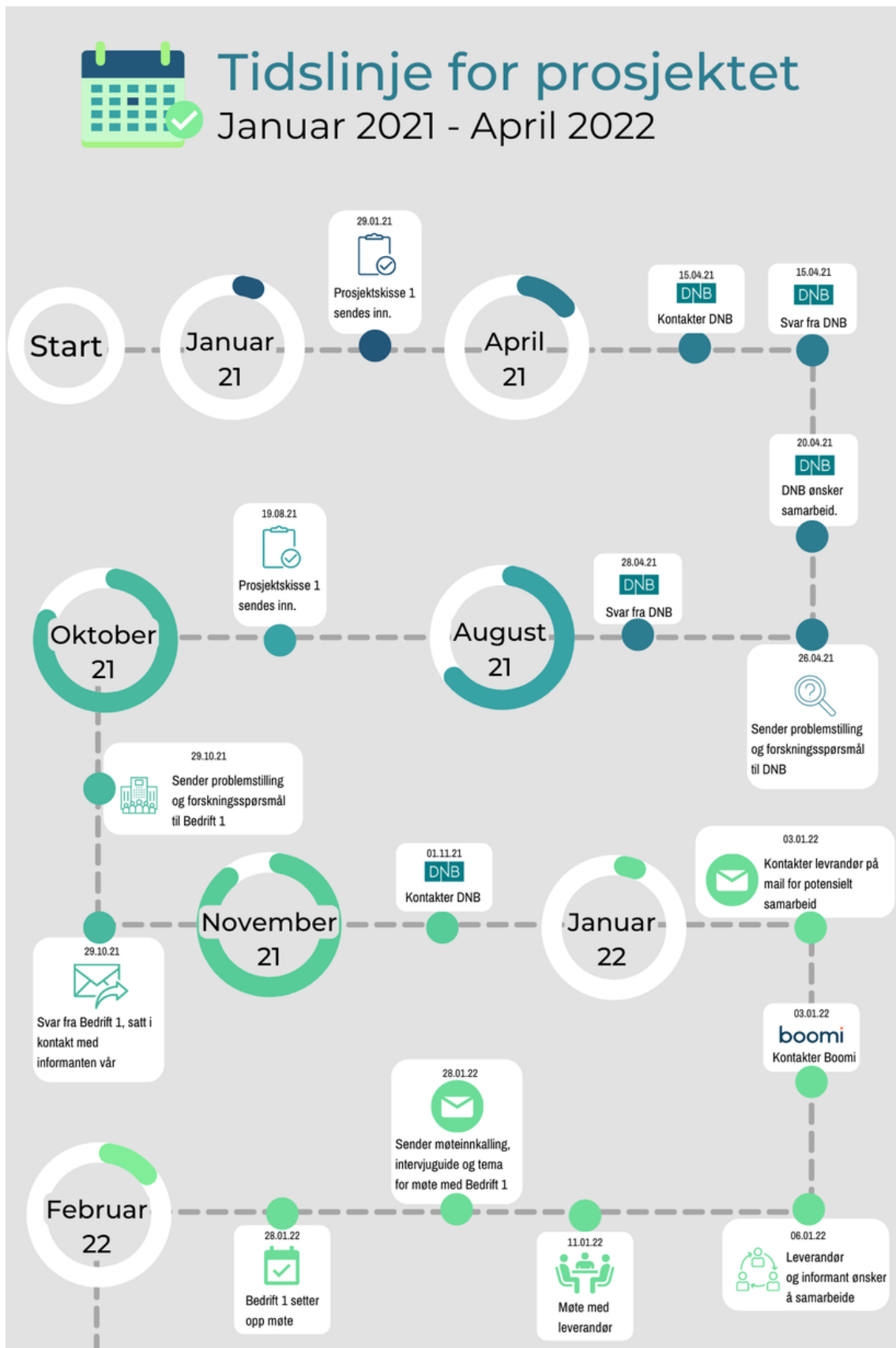
- Data model: hvilke data defineres som master data i virksomheten, hvordan er master data strukturert, hvilke systemer bruker master data og hvor ligger master data lagret.
- Data quality:
 - Har virksomheten et forhold til hvilken kvalitet data skal ha?
 - Har virksomheten et forhold til hvordan kvaliteten påvirker forretningen?
 - Vet virksomheten hvor kilden til dårlig datakvalitet ligger og hvordan man kan forbedre den?
 - Anser virksomheten data som en del av organisasjonens ressurser?
 - Kjernen her er at «data er en del av organisasjonens ressurser, og derfor må kvaliteten på data være god slik at man har mulighet til å oppnå konkurransefordeler og slik at organisasjonen unngår å støte på problemer knyttet til datakvalitet».
- Usage & ownership: Er det klare retningslinjer for hvem som har read/write tilganger til data, og hvem som kan bruke data i de ulike systemene?
- Data protection: hvor og hvordan er data sikret mot mulige forhold som svikt i komponenter, software bugs eller spesifikke angrep?
- Maintenance: hvilket forhold/retningslinjer har virksomheten til fysisk lagring og livssyklusen til data?

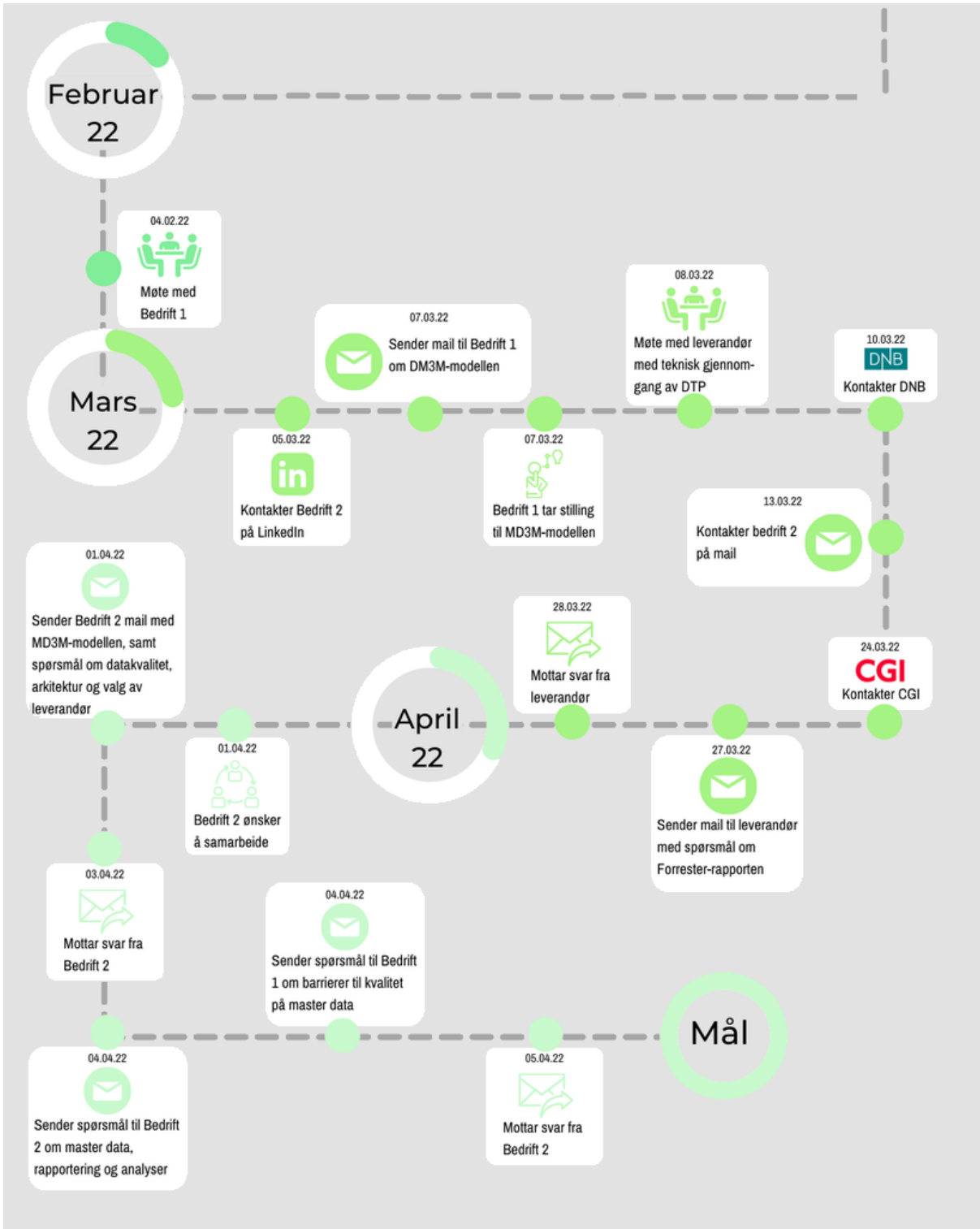
Vedlegg 5 - Første intervjuguide til leverandøren i casestudien

Intervjuguide til Leverandør: Master data hub, teknologi og kunder

- Hva innebærer en master data hub for dere?
- Hvilke faktorer påvirker valg av hub/ hvorfor velger virksomheter hub?
- Hvilke tjenester/kvaliteter har dere spesifikt bygget inn i deres master data hub som gjør at kundene velger dere?

Vedlegg 6 - Tidslinje for forskningsarbeidet





Vedlegg 7 - Andre intervjuguide til virksomhetene

Intervjuguide til Bedrift 1 og Bedrift 2: Barrierer knyttet til kvalitet på master data

Haug & Arlbjørn (2011) lister opp følgende barrierer:

- Mangel på roller og ansvar
- Mangel på eiere knyttet til datakvalitet
- Ineffektive organisasjonsprosesser
- Manglende verktøy og teknologi
- Manglende prosedyrer
- Manglende belønningssystemer
- Manglende forståelse og involvering fra ledelsen
- Forsømmelse av administrative detaljer som opplæring, arbeidsbeskrivelser, ansvarsfordeling og problemer knyttet til kommunikasjon

Er det noe her du kjenner deg/virksomheten igjen i, eller tenker er spesielt viktig?

Vedlegg 8 - Sentrale datafunn

Forskningsspørsmål	Bedrift 1	Bedrift 2	Leverandør
Hva er god datakvalitet?	<ul style="list-style-type: none"> • Har ikke noe mål på hva som er god kvalitet på master data, men det er generelt fokus på god kvalitet • Virksomheten har et forhold til hvordan datakvalitetene på master data påvirker forretningen, og informanten trekker spesielt frem at mangelfull artikkel-, kunde-, og brukerdata oppleves som et hinder i virksomhetens strategiske måloppnåelse • Det kommer frem at det er mangelfulle data eller datakvalitet i de fleste master data objektene, bortsett fra leverandordata. • Iverksatt prosjekter for å bedre artikkel master data, kunde master data og sikkerheten rundt data. 	<ul style="list-style-type: none"> • Bedrift 3 mener at kvaliteten på master data i virksomheten skal være høy. Dette er for at fagsystemene skal kunne stole på master dataene de henter eller får via plattformen. Informanten sier at lav kvalitet på master data i virksomheten vil føre til ekstra manuell håndtering, mulige feil i forretningsprosesser, som til syvende og sist kan påvirke sluttikunden. • Virksomheten er kjent med hva som er kildene til dårlig datakvalitet, og i noen tilfeller har de iverksatt initiativ for å rydde i master data slik at kvaliteten øker. • I Bedrift 3 sitt tilfelle er det kildesystemet som er ansvarlig for kvaliteten i master data, samt at tjenesteplattformen inneholder noen kapabiliteter som duplikatsjekk for å kunne understøtte kvaliteten • Bedrift 3 kjenner seg igjen i barrierene knyttet til kvalitet på master data, og synes disse beskriver utfordringene de møter godt. Spesielt nevnes manglende systemstøtte som tillater duplikater og manglende opplæring av brukere. 	<ul style="list-style-type: none"> • God datakvalitet er data som kan dyttes rett ut i applikasjoner og tjenester som ligger i tjenestetilaget, samt analyser og rapportering • Sikres gjennom <i>system of entry</i>. • Det er ikke innebygd master data management i <i>hub</i>.

Forsknings spørsmål	Bedrift 1	Bedrift 2	Leverandør
Hvilke strategiske, taktiske og forretningsmessige valg må en bedrift ta for å ha god master data management?	<ul style="list-style-type: none"> Anser data som en viktig ressurs for forretningen. Dette blir tydeliggjort gjennom forretningsstrategien og teknologistrategien Det er delvis klare retningslinjer om hvem som har tilgang til å lese og skrive masterdata, og hvem som skal bruke masterdata i de ulike systemene og applikasjonene. Det er tydelig når man snakker om brukere da disse er knyttet til autorisasjoner. Så er det mindre tydelig når virksomheten skal se på hvilke deler av organisasjonen eller hvilke løsninger det er som har eierskap til dataene. Data sikres gjennom tilgangsstyring på brukernivå, redundante løsninger og generell sikring av den tekniske infrastrukturen Det er ingen generelle retningslinjer når det kommer til fysisk lagring og livssyklusen til masterdata. Hver masterdata objekt har hver sin policy knyttet til livssyklus håndtering. prosjektet er en direkte konsekvens av teknologistrategien virksomheten har utarbeidet. Virksomheten ønsker å gjøre dataprodukt (masterdata og transaksjonsdata) tilgjengelig gjennom åpne, og definerte grensesnitt. I denne teknologistrategien fremkommer det flere teknologiske mål som understøtter implementering av en artikkel master data management plattform; avkobling og modularisering av systemarkitekturen, redusere ERP-produkt, agil utvikling og kontinuerlig implementering. 	<ul style="list-style-type: none"> Bedrift 2 anser data som en del av virksomhetens ressurser. Dataene er viktige for virksomheten og er relatert til en god del av forretningsområdene som kunder, forvaltning og bank. Virksomheten sin datamodell er en definert domenenmodell. Denne datamodellen er en felles beskrivelse av hvilke data og relasjoner mellom data som benyttes på tvers av virksomheten. Domenemodellen danner grunnlaget for felles begreper innad i virksomheten, samt et felles datasett som systemene benytter for deling av informasjon. Da det er kilde systemene som er eier av master data er det der rettighetene til å oppdatere og vedlikeholde data ligger. Forbruker av master data får rettigheter til å lese gjennom godkjenning fra dataeier. Dersom forbruker skulle ønske å oppdatere master data så sendes oppdateringen til kilden (der data ligger lagret i systemet) og kilden oppdaterer så master data huben med nye data. På denne måten er også hvert fagsystem ansvarlig for å sikre sine data, og virksomheten har for vane å kjøre sikkerhetstester. Fagsystemene eier forretningsregler om opprettelse, oppdatering og sletting av master data. Master data huben forholder seg til disse oppdateringene fra fagsystemene. Noen systemer i Bedrift 3 er on-premise og noen ligger i skyen, dermed er også lagringen av master data splittet mellom disse to. Nå som virksomheten distribuerer master data fra en master data hub til flere fagsystemer er håndtering av og eierskap til master data en utfordring. I noen tilfeller er det flere systemer som eier deler av dataene, og da blir det vanskelig å sikre at master data kun oppdateres fra kilden. Dette er en utfordring de jobber med å løse på en god måte. Siden virksomheten nå har flere systemer som er "deleier" av master data øker det også muligheten til feilkilder. Feilsøking involverer derfor ikke bare en avdeling lenger, men 3-4. Dette er tidkrevende og skaper unødvendig arbeid i noen av avdelingene. 	<ul style="list-style-type: none"> den skal gi muligheten til å effektivisere og skape nye arbeidsprosesser, og den skal tilgjengeliggjøre data for analyser og beslutningsstøtte i real time. De fleste kundene har ingen ide eller strategi om en tjenesteplattform Noen har fokus på digitalisering og digital transformasjon

Forsknings spørsmål	Bedrift 1	Bedrift 2	Leverandør
Kan master data hub være en del av master data lifecycle management?	<ul style="list-style-type: none"> Informanten mener at domener som kunde, artikkel og leverandør kan gi mening å ha i samme master data hub fordi de kan være knyttet til hverandre, som at en artikkel er knyttet til en spesifikk leverandør. Men det er også mange masterdata som ikke har noen forretningsmessig verdi eller operasjonell verdi ved å være samlet i en masterdata hub. Det gir ikke mening å flytte masterdata objekter inn i samme hub kun for løsningen sin del; for eksempel vil det ikke gi mening å flytte produkt masterdata og kunde masterdata i samme hub da linken mellom disse er transaksjonsdata. De er to separate domener. Hvilke master data er det gunstig å flytte 	<ul style="list-style-type: none"> Master data lagres både i kilde systemet og i master data huben. I og med at oppdater og vedlikehold av master data skjer i fagsystemet vil ikke master data huben være en del av master data sin livssyklus i dette tilfellet. Master data huben distribuerer data til systemer og applikasjoner og systemer som trenger det. 	<ul style="list-style-type: none"> Kjernesystemet eier og vedlikeholder master data

Forsknings spørsmål	Bedrift 1	Bedrift 2	Leverandør
Hvilke faktorer påvirker valg av master data hub?	<ul style="list-style-type: none"> • Bedrift 1 jobber veldig rettet mot en løs kobling av arkitekturen sin, som vil si hvordan de skal få systemer til å være minst mulig avhengige av hverandre. Man kan si at et lagringssted for master data kan være et virkemiddel til det, men alle master data på ett sted gir den motsatte effekten gjennom å skape nye avhengigheter som ikke er nødvendige. • Bedrift 1 jobber med å flytte ut i en masterdata hub er produkt master data. Gjennom dette vil virksomheten kunne få en autoritativ kilde til masterdata som ikke er tett knyttet til et spesifikt fagsystemet, som logistikk eller økonomi, men som et nøytralt masterdata system. • Årsaken til at Bedrift 1 søker en løsning for master data håndtering av artikkel master data er fordi dagens løsning har gått ut av support, samt at den ikke fungerer optimalt. Teknisk sett har de en løsning i dag som fungerer som et <i>repository</i> for artikkel, men i realiteten er det ERP-systemet deres som fungerer som den autoritative kilden til artikler. I tillegg har ikke dagens <i>repository</i> kapabilitetene virksomheten ønsker fra en master data management løsning; det er ikke fleksibel nok i distribusjon av master data og den mangler verktøy for å orkestrere litt mer avanserte valideringer av mange datastrukturer (datamodeller) samtidig. • De har også utfordringer knyttet til kunde master data som ligger lagret i et <i>customer relationship management</i> (heretter kalt CRM) system. Her er ikke kundedatabasen nøytralisert, som vil si at den er systemavhengig, og skaper direkte integrasjoner mellom CRM som holder masterdataene og alle andre systemer og applikasjoner som bruker kunde master data. Når systemet som holder masterdata også er et operativt system opplever informanten at det skaper konflikter og at virksomheten låser systemet veldig, og at dette er noe som gjør det vanskelig å gjøre endringer i arkitekturen og utvikling i systemene over tid. • Må gå omveier for å bruke artikkel master data i f.eks netthandel. Dette er ikke i tråd med en løsrevet arkitektur som skal hjelpe med virksomhetens • strategiske måloppnåelse. • Da det kun er artikkel master data som skal flyttes ut i en nøytralisert løsning så vil det føre til utfordringer og belastninger for integrasjonen til virksomheten, da master data objektene ligger lagret på ulike steder. Bedrift 1 har valgt å fokusere på å sammenstille objektene ved behov, i stedet for å flytte de inn i den samme løsningen. Informanten mener at de ved å flytte både leverandør master data og artikkel master data over i den samme nye løsningen vil virksomheten skape stadig nye avhengigheter, og hva oppnår man egentlig ved å flytte avhengighetene fra ett sted til et annet. 	<ul style="list-style-type: none"> • Bedrift 3 er også i en prosess med å utfase det ene systemet sitt som skal erstattes av flere fagsystemer. Alle disse fagsystemene kommer til å benytte seg av master data fra ett eller flere domener. • Grunnen til at Bedrift 3 søkte etter nye løsninger på markedet var på grunn av utfasing av fagsystemet sitt • Da fagsystemet ble utfaset oppstod det et behov for å håndtere master data på en annen måte. Gjennom denne endringsprosessen endret også arkitekturen i virksomheten seg • I tiden da fagsystemet var kilden til master data opplever virksomheten duplikater og inkonsistens. Dette ble håndtert direkte i systemet, og opplevdes ikke som noe stort problem (Var overraskende nok ikke en årsak til implementering av master data management, og virksomheten opplever mer duplikater og feil i datagrunnlaget nå, når master data distribueres fra en master data hub • Bedrift 3 opplever at ting generelt har blitt mer komplisert da dataene blir fordelt på flere systemer, og informanten trekker frem at avhengighetene har økt i takt med kompleksiteten. • Samtidig har også systemstøtte for brukerne (duplikater osv) økt og virksomheten forventer at kvaliteten i master data vil øke over tid. 	<ul style="list-style-type: none"> • Utdaterte integrasjoner med gammel teknologi • Ønske om en skybasert tjeneste • kunder som tenker på selvbetjening • Heltetlig tilnærming kontra lift and shift • Modenhet?

Forskningsspørsmål	Bedrift 1	Bedrift 2	Leverandør
Hvilke organisatoriske, forretningsmessige- og IT systemer berøres av en master data hub?	<ul style="list-style-type: none"> • Bedrift 1 sin arkitektur vil også bli påvirket av den nye teknologi strategien, og da spesielt med tanke på integrasjoner. Artikkel master data management påvirker mange applikasjoner og applikasjonstjenester både internt og eksternt i bedrift 1. • Tre eksterne databaser sørger for mesteparten av inputen av artikkel data i bedrift 1 sin verdikjede. Data fra disse kildene importeres i ulike formater og teknikker til SAP Master Data Management. • Systemporteføljen, system i verdikjeden, POS-systemer, e-commerce systemer, planleggingsystemer, PIM & DAM systemer 	<ul style="list-style-type: none"> • Alle fagsystemer – kunde, forvaltning og bank 	<ul style="list-style-type: none"> • Alle applikasjoner i tjenestelaget, alle analyseverktøyene i Intelligencelaget og alle kjernesystemene inkl SaaS og PaaS (Microsoft 365).

Forskningsspørsmål	Bedrift 1	Bedrift 2	Leverandør
Hvem leverer denne type teknologi og hvilke tjenester/kvalitet er bygget inn i disse, der man går ett steg lenger ?	<ul style="list-style-type: none"> • skal tilpasse seg eksisterende datamodeller i databasene, uten tap av struktur og bredde. Krever sofistikert datamodellering fra leverandøren som skal sammenstille alt til "one single truth" • nye mønstre for integrasjon som legger til rette for eksisterende og fremtidige egenskaper i databasene • skal unngå transformasjon i integrasjonslaget for en hver pris 	<ul style="list-style-type: none"> • Da Bedrift 3 skulle velge en leverandør var det viktig at de skulle få en samarbeidspartner som hadde erfaringer med digitale tjenesteplattformer. • Kriteriene for valg av leverandør var kunnskap rundt skytjenester, digital tjenesteplattform, master data håndtering, kunne stille med flere konsulenter som jobber sammen med og overfører kunnskap virksomhetens ansatte. • Selv trekker virksomheten frem at det viktigste kriteriet til leverandør var kunnskap om teknologi og metodikk som kunne støtte realiseringen av domenemodellen. 	<ul style="list-style-type: none"> • API-laget tilgjengeliggjør informasjon til tjenestelaget • Datalaget har teknologier for lagring og distribusjon av data • Integrasjonslaget ekstraherer og modellerer data • Kjernesystemene vedlikeholder og oppretter data • Kunden står selv for datamodellering (inkl AI og ML) • Pris- og supportmodell, integrasjonsvennlighet og integrasjonsmønster, og brukervennlighet • Mener rapporten er tilpasset relativt store kunder med avansert behov for master data management