

Hvordan brukes evaluering i offentlig sektor?

Et case-studie av Innovasjon Norges evalueringsvirksomhet

Vegard Solhjem Knutsen

Veileder

Gro Kvåle

Masteroppgaven er gjennomført som ledd i utdanningen ved Universitetet i Agder og er godkjent som del av denne utdanningen. Denne godkjenningen innebærer ikke at universitetet inntår for de metoder som er anvendt og de konklusjoner som er trukket.

Universitetet i Agder, 2016

Fakultet for samfunnsvitenskap

Institutt for statsvitenskap og ledelsesfag

Forord

En lang prosess er nå over idet masteroppgaven leveres. Det har både vært spennende og lærerikt, men til tider også strevsomt og frustrerende. Arbeidet har gitt meg kunnskap og innsikt i et faglig tema som interesserer meg, og som jeg hadde lite innblikk i fra før. Ikke minst har det gitt meg verdifull innsikt i forskningsprosessen og det «håndverk» som ligger bak innsamling og analyse av empiri.

Først og fremst ønsker jeg å rette en takk til min veileder, Gro Kvåle som har gitt meg verdifulle kommentarer og tilbakemeldinger gjennom hele prosessen. Hun har vært tilgjengelig og vist interesse for oppgaven. Det motiverer!

Ikke minst ønsker jeg også å rette en stor takk til informantene i Innovasjon Norge som var villig til å sette av tid og stille opp til intervju i en travel arbeidshverdag. På Innovasjon Norge ble jeg møtt på en utmerket måte. Uten deres velvilje ville det ikke blitt noen oppgave.

Takk også til Anne Halvorsen på Institutt for sosiologi ved UiA, som bidro med nyttige faglige avklaringer og tips helt i startfasen av prosjektet.

Til slutt vil jeg rette en takk til min samboer, Margrethe, for korrekturlesing og nyttige kommentarer på den språklige fremstillingen.

De ovennevnte har vært viktige ressurser. Jeg tar selv ansvar for eventuelle feil og mangler i denne oppgaven.

Vegard Solhjem Knutsen

Kristiansand, 28. mai 2016

Sammendrag

Evaluering har i økende grad blitt tatt i bruk av myndigheter som kunnskapsgrunnlag og verktøy for styring og politikkutforming. Disse er ofte utført av eksterne aktører, gjerne private konsultantselskaper eller forskningsinstitutter. Å evaluere er en plikt etter statens økonomireglement. Det produseres nå en mengde evalueringer, og det settes av betydelige midler til formålet. Hvordan evalueringer brukes er derfor et viktig demokratisk anliggende. Økt bruk av evaluering som styringsverktøy har resultert i en rekke initiativ på feltet og det betraktes nå som et eget fagfelt.

Det er imidlertid lite erfaringsmateriale om hvordan evalueringer brukes i norske offentlige virksomheter. Denne oppgaven tar derfor sikte på å bidra med kunnskap på feltet.

Oppgaven er et kvalitativt case-studie av evalueringsbruk i Innovasjon Norge. Formålet med oppgaven har vært å undersøke hvordan evalueringer tas i bruk i en norsk offentlig virksomhet og hvordan evalueringer bidrar til styring og utvikling? Samtidig omfatter oppgaven en kartlegging av Innovasjon Norges samlede evalueringsystem. Den teoretiske tilnærmingen har vært knyttet til en brukstypologi som er sentral i litteraturen. Disse brukstypene har blitt forstått i lys av et rasjonelt og institusjonelt perspektiv på organisasjoner. Gjennom intervjuer av sentrale nøkkelinformanter i Innovasjon Norge belyser oppgaven hvordan evalueringer påvirker styring og utvikling av organisasjonen.

Oppgaven argumenterer for at det er nødvendig med en bred tilnærming til hva evalueringsbruk er, for å kunne danne et bilde av den faktiske betydningen de har på virksomheten som gjennomfører dem. Funnene peker i retning av at flere brukstyper er i spill samtidig. Det vil si at evalueringer blant annet har en direkte innvirkning på at beslutninger fattes. Det har indirekte innvirkning, ved at evaluering inngår som en del av et samlet beslutningsgrunnlag. Funnene peker samtidig i retning av at evaluering inngår som en viktig del av virksomhetens behov for å legitimere innsatsen, spesielt i betydning av legitimering ovenfor oppdragsgivende- og bevilgende departementer.

Innholdsfortegnelse

1.	Innledning.....	1
1.1	Problemstilling	2
1.2	Metode og disposisjon.....	3
1.3	Hva er evaluering?.....	4
1.4	Hvorfor evaluere?.....	7
1.5	Norsk evalueringspraksis - omfang og bruk.....	8
1.6	Konteksten er Innovasjon Norge	10
1.6.1	Næringspolitikk i Norge	11
1.6.2	Innovasjon Norges virksomhet.....	11
1.6.3	Krav til måling.....	13
1.6.4	Hvorfor er Innovasjon Norge et interessant case?	14
2.	Teoretisk rammeverk.....	16
2.1	Utgangspunkt - organisasjonsteori	16
2.1.1	Rasjonelt perspektiv	17
2.1.2	Læringsperspektiv	18
2.1.3	Institusjonelt perspektiv.....	19
2.2	Seks typer evalueringsbruk.....	20
2.2.1	Instrumentell bruk	20
2.2.2	Konseptuell bruk	21
2.2.3	Legitimerende bruk	22
2.2.4	Taktisk bruk.....	22
2.2.5	Rituell eller symbolsk bruk	23
2.2.6	Prosessuell bruk.....	24
2.3	Bruk av funn versus bruk av prosess	25
2.4	Et bredt perspektiv på evaluering	27
3.	Metode.....	30
3.1	Forskningsdesign- om casestudie	30
3.2	Tre prinsipper for datainnsamling	33
3.3	Intervjuet	34
3.4	Utvalg	36
3.5	Dokumentanalyse	38
3.6	Reliabilitet og validitet	39
3.6.1	Reliabilitet	39
3.6.2	Validitet	41

3.7	Etiske refleksjoner	43
4.	Resultater og analyse - Innovasjon Norge og evaluering	45
4.1	Institusjonalisering - hvordan gjennomfører Innovasjon Norge evaluering?	46
4.1.1	Hva er evaluering i Innovasjon Norge?	50
4.1.2	Evalueringsprosessen	51
4.1.3	Kompetanse og evalueringskapasitet.....	57
4.1.4	Hvorfor evaluere?.....	59
4.2	Hvordan brukes evalueringer?.....	62
4.2.1	Prosess eller funn?.....	62
4.2.2	Evaluering som direkte premiss for beslutning	67
4.2.3	Læring og utvikling	71
4.2.4	Evaluering - et verktøy for å legitimere innsatsen?.....	75
4.2.5	Evaluering er en trend?.....	78
5.	Oppsummering og avsluttende betraktninger.....	82
5.1	Et mangfoldig fenomen med empirisk orientering.....	82
5.2	Bidrag, begrensning og forslag til videre forskning	84
6.	Referanseliste	87
7.	Vedlegg	91

Hvordan brukes evaluering i offentlig sektor? Et case-studie av Innovasjon Norges evalueringsvirksomhet

1. INNLEDNING

Myndigheten har i økende grad tatt i bruk evalueringer som kunnskapsgrunnlag og verktøy i styring og politikktutforming. Det produseres en mengde evalueringsrapporter med beskrivelse av funn og anbefalinger til endring og forbedring av offentlige virksomheter, virkemidler og programmer. Å gjennomføre evalueringer er pålagt etter statens økonomiregelverk. Reglementet gjelder for alle statlige forvaltningsorganer og andre foretak som staten eier fullt eller delvis (Finansdepartementet 2005). Samtidig er det utstrakt bruk av evalueringer også på kommunalt styringsnivå, uten at det er pålagt på samme måte som for statlige virksomheter (Halvorsen og Madsen 2013:23).

Kort sagt gjennomføres det en omfattende mengde evalueringer, men ifølge Anne Halvorsen og Einar Lier Madsen er det derimot mangel på kunnskap knyttet til hvorvidt og hvordan informasjonen fra evalueringer brukes i styring og utvikling av offentlig politikk og virksomhet. De hevder at det:

«... i Norge finnes lite erfaringsmateriale om hvordan evaluering inngår i statlig styring, og hvilke prosesser og prosedyrer som faktisk er etablert på området, vi vet lite om hvorvidt de ulike formene for evaluering fungerer etter hensikten og hvordan de brukes» (Halvorsen og Madsen 2013:21)

Selv om evaluering ikke er ett nytt begrep, har bruk av evaluering som styringsverktøy hatt et betydelig oppsving fra årtusenskiftet og som nevnt er det nå pålagt for alle statlige virksomheter å gjennomføre evalueringer (Halvorsen og Madsen 2013:23). I Statens økonomiregelverk § 16 heter det at *«alle virksomheter skal sørge for at det gjennomføres evalueringer for å få informasjon om effektivitet, måloppnåelse og resultater innenfor hele eller deler av virksomhetens ansvarsområde og aktiviteter»* (Finansdepartementet 2003:13).

Økt bruk av evaluering som styringsverktøy har resultert i en rekke initiativ på feltet. Evaluering betraktes nå som et eget fagfelt, med nasjonale og internasjonale fagorganisasjoner og faglige nettverk som arrangerer konferanser, utgir tidsskrifter og faglitteratur (Halvorsen & Madsen 2013: 23). Det global fagnettverket EvalPartners har utropt 2015 til det internasjonale evalueringsåret (Norsk evalueringsforening 2015).

En indikasjon på at evalueringsbruk har blitt en viktig del av styringsverktøyet i staten, kan vi se på Direktoratet for økonomistyrings (DFØ) nettbaserte evalueringsportal, hvor det registreres evalueringer utført på oppdrag fra departementer og andre statlige virksomheter. Her er det i dag registrert nær 2500 evalueringsrapporter, hovedsakelig er det evalueringer som er gjennomført fra 2005 frem til i dag (Evalueringsportalen 2015). Samme direktorat utga i 2011 «*Veileder for strategisk og systematisk bruk av evaluering i styringen*» som tar sikte på å bidra til at statlige virksomheter kan øke kunnskap og bevissthet om hvordan evaluering kan bidra som verktøy for bedre styring (DFØ 2011:3). I tillegg ble det allerede i 1994 opprettet et nettverk for evaluering i staten kalt Eva- forum, som skal bidra til kunnskap og erfaringsutveksling på tvers av virksomheter i statlig sektor. Norsk evalueringsforening ble opprettet så sent som i 2009. De skal være et faglig nettverk som organiserer både bestillere og utførere, og har som overordnet mål å utvikle fagfeltet (Halvorsen og Madsen 2013:23).

Det settes av betydelige midler og ressurser til gjennomføring av evalueringer, og det at det er innført som et sentralt styringsredskap i norsk forvaltning gjør det helt sentralt å forstå om evalueringer virker og gir resultater (Halvorsen og Madsen 2013:26). Viktige spørsmål er om evaluering fungerer som et grunnlag for læring og forbedring? Fører evalueringer til endring? Blir det tatt i bruk i styring, og som grunnlag for beslutninger? Hva må eventuelt til for å oppfylle denne funksjonen (Ibid)?

1.1 PROBLEMSTILLING

En interessant indikasjon på at det er behov for mer kunnskap på feltet, er at DFØ nylig har igangsatt et oppdrag, hvor målet er å beskrive hovedtrekk ved bruk av evaluering i staten med utgangspunkt i rapportene som er publisert i Evalueringsportalen. I følge utlysningsteksten¹ er oppdragets formål å kartlegge hva som blir evaluert, hvordan det blir evaluert, hvilke evalueringer som blir brukt, og hvordan et utvalg av enkelte evalueringer har blitt brukt. Rapporten er forventet ferdigstilt i september 2016 (DFØ 2016, tilbudsforespørsel).

Denne oppgavens tematikk begrenser seg imidlertid til det som kanskje kan sies å være det aller mest sentrale når det utføres en evaluering, nemlig hvordan evalueringen blir tatt i bruk. Problemstillingen i denne masteroppgaven blir dermed:

¹ Tilbudsforespørsel: Bruk av evaluering i statlig styring, ble 16. mars 2016 sendt fra DFØ til Oxford Research og andre aktører. Konkurransgrunnlaget er ikke lagt ut på Doffin.

Hvordan tas evalueringer i bruk og hvordan bidrar evalueringer til styring og utvikling i en norsk statlig virksomhet?

Det kan tenkes at offentlige virksomheter både har ulike motivasjon for å gjennomføre evalueringer og ulike intensjoner om hvordan funnene fra evalueringen skal benyttes. Denne oppgaven tar sikte på å kartlegge evalueringsbruk i Innovasjon Norge (IN), knyttet til en brukstypologi som finnes igjen i de fleste artikler og bøker på feltet (Rich; 1977; Weiss 1977; Levinton & Hughes 1981; Johnsen 1998; Alking & Taut 2003; Henry & Mark 2003; Patton 2008; Vedung 2009; Gildemyn 2014; Herbert 2015). Brukes evalueringsfunns på en *instrumentell* måte, altså til konkrete endringer og nye prosedyrer i det som er gjenstand for evaluering? Brukes evalueringer på en *konseptuell* måte, for eksempel til kunnskapsakkumulering og læring, uten at det fører til konkrete beslutninger og handlinger? Brukes evalueringer som et middel for å *legitimere* egen virksomhet, altså til å bekrefte og rettferdiggjøre egne synspunkter og tidligere avgjørelser? Eller kan evalueringer benyttes på en *taktisk* måte, for å unngå eller unnvike at beslutninger fattes nå? Brukes evalueringer på en *symbolsk* måte, ved å bevise modernitet og fremskritt? (Halvorsen og Madsen 2013:19). Til slutt vil det også undersøkes om evalueringer brukes på en prosessuell måte, det vil si at selve gjennomføringsprosessen blir benyttet som et læringsinstrument (Patton 2008).

For å gi en god beskrivelse av hvordan evalueringer blir brukt i IN, har studie tatt sikte på en kartlegging av hele evalueringsprosessen i virksomheten. Når man undersøker evalueringsbruk, vil det være nødvendig å se det samlede evalueringssystemet og de prosessene som følges i nær sammenheng med typer bruk.

Oppgaven forsøker ikke å trekke kausalslutninger om hva som hindrer og fremmer ulike typer bruk. Formålet er med andre ord å forklare hvordan evalueringer tas i bruk. Det er ikke utformet hypoteser som skal testes. Oppgaven vil likevel forsøke å diskutere hvor overførbart tilfellet IN er til andre deler av statsforvaltningen, uten at det trekkes klare generelle slutninger. For å kunne diskutere overførbarhet, har det også vært viktig å diskutere den konteksten IN befinner seg i.

1.2 METODE OG DISPOSISJON

Denne masteroppgaven er et casestudie hvor Innovasjon Norge behandles som et case på evalueringsbruk i en offentlig virksomhet. Dataene som danner grunnlag for analysen stammer fra intervjuer med nøkkelinformanter i IN. Dataene vil støttes av sentrale dokumenter der det er naturlig, slik som INs siste årsrapport, INs årlig rapport til oppdragsgivere, tildelingsbrev,

brosjyrer og referater fra styringsmøtene mellom IN og NFD. Analysen tar for seg hvordan resultater og funn benyttes til styring og utvikling av INs virksomhet. Det finnes en mengde litteratur som omhandler evalueringsbruk, spesielt fra amerikanske forskere. Det er imidlertid lite empirisk materiale fra norsk forvaltning. Det er også i liten grad utviklet testbare modeller (Halvorsen og Madsen 2013:21), noe som sammen gjør at innsamling av empiri gjennom åpne intervjuer er en egnet metode. Studiet er eksplorerende og søker å kartlegge fenomenet evalueringsbruk i en norsk sammenheng. Funnen vil knyttes opp mot teoretiske perspektiver på hvordan evaluering kan brukes, mer presist vil det knyttes opp til brukstypologien som er nevnt ovenfor.

Kapittel 1 vil gå gjennom definisjon av og formål med evaluering, samt å gi et innblikk i norsk evalueringspraksis. Mot slutten av kapitlet vil jeg gi en fremstilling av caset, Innovasjon Norge (IN), samt en kort diskusjon om betydning av kontekst. I kapittel 2 blir de teoretiske perspektivene presentert. Jeg vil gå nærmere inn på hvordan evalueringsbruk må forstås, hvor utgangspunktet er en bred tilnærming til hva evalueringsbruk er eller kan være. Det vil bli gitt en detaljert redegjørelse for de ulike brukstypologiene. I kapittel 3 vil jeg gjennomgå metode og forskningsdesign. Her vil det nærmere beskrives hvordan jeg har gått frem ved innsamling av empirien. Metodekapitlet inneholder også en redegjørelse av reliabilitet og validitet, tett knyttet til oppgavens innhold. I kapittel 4 vil funn fra datainnsamlingen presenteres og analyseres i tematiske kategorier, inndelt etter brukstypene. I femte og siste kapittel vil oppgaven oppsummeres og konklusjon presenteres.

1.3 HVA ER EVALUERING?

Et sentralt kjennetegn ved evalueringsbegrepet er at det har vært og fremdeles er vagt og inkonsistent (Herbert 2014). Det samme trekkes frem i Halvorsen og Madsen (2013) som hevder at evaluering er et vidt og mangfoldig begrep som viser til en rekke ulike former for undersøkelser. Begrepet brukes også gjerne i dagligtalen på samme måte som vurdering.

Denne oppgaven skal undersøke evalueringsbruk i en norsk offentlig virksomhet. Det kan derfor være nyttig å se hvordan det offentlige selv bruker begrepet. I Finansdepartementets «*Veileder til gjennomføring av evaluering i staten*» (2005) defineres evaluering på denne måten:

«En systematisk datainnsamling, analyse og vurdering av en planlagt, pågående eller avsluttet aktivitet, en virksomhet, et virkemiddel eller en sektor. Evalueringer kan gjennomføres før et tiltak iverksettes (*ex ante*), underveis i gjennomføringen, eller etter at tiltaket er avsluttet (*ex post*). Evalueringen kan utføres av interne eller eksterne fagmiljøer» (Finansdepartementet 2005: 8)

Denne definisjonen er bred og vil dermed omfatte en mengde ulike typer evalueringer. Dette inkluderer *ex-ante* evalueringer, slik som konsekvensutredninger, eller samfunnsøkonomiske analyser som tar sikte på å undersøke nytte/kostnad av et aktuelt tiltak før det iverksettes. Det kan omfatte underveisevalueringer som gjennomføres for at det skal være mulig å endre kurs eller justere et tiltak underveis, og *ex-post* evalueringer som typisk tar sikte på å undersøke om et tiltak når de mål og resultater som er satt på forhånd (Finansdepartement 2005). Evaluering kan også omfatte det som vi normalt ikke betegner som evalueringer slik som forvaltningsrevisjon, selskapskontroller og granskninger (Halvorsen og Madsen 2013:16).

Evert Vedung benytter seg av en minimal definisjon av evalueringsbegrepet, hvor evaluering er en «*systematisk vurdering i etterkant av et utfall*» (Vedung 2009:22). Han gjør det klart at denne definisjonen ikke inkluderer vurdering i forkant, evaluering er en retrospektiv øvelse: evalueringer ser tilbake og vurderer det som alt er besluttet, pågår eller er avsluttet. Denne begrensingen gjøres for at begrepet skal være mer håndterbart for den offentlige sektoren og for at det ikke skal være en samlebetegnelse for alt hva offentlig sektor foretar seg av analyser (Vedung 2009, 2010). Han mener videre at definisjoner som inkluderer for mye er et resultat av den semantiske magneten som ligger i ordet evaluering - det har blitt et moteord og det forventes at evaluering er «godt og riktig».

Robert Patton, en av de viktigste bidragsyterne i utviklingen av evalueringsfaget, er opptatt av at en forståelse av begrepet må være pragmatisk og praksisnær. Evalueringer vurderer et tiltaks «*merit, worth, value, or significance*», det handler dermed om å trekke slutninger om kvalitet, effektivitet og måloppnåelse (Patton 2008:5). Hovedargumentet er at evalueringer skal være bruksfokuserte, evaluering handler om «*intended use for intended user*» (Patton 2008:37).

Evalueringer er forskningslignende, men det finnes viktige forskjeller mellom forskning og evaluering. Begge tilfeller dreier seg om kunnskapsproduksjon. Evalueringer og forskning er også produsert på lignende måte, men med ulike mål. Målet med evalueringer er å produsere kontekstspesifikk kunnskap - kunnskap som kun er gjeldene innenfor en spesifikk setting og

innenfor et spesifikt tidspunkt, og tenkt brukt for en spesifikk gruppe mennesker. Evalueringer er beslutningsorienterte, det innebærer at de skal brukes til noe (Alkin og Taut 2003:3).

Det å *bruke* evalueringer handler om hvordan informasjon og innsikt som produseres av evalueringer har konsekvenser for tiltaket eller organisasjonen som har blitt evaluert (Alkin og Taut 2003:1) Førte for eksempel en evaluering til ny forståelse? Hvilke endringer har blitt gjennomført på bakgrunn av en evaluering? Tilegnet de ansatte seg ny kunnskap og forståelse som en konsekvens av evaluering? (Alkin og Taut 2003:1). Dette er blant spørsmålene som denne oppgaven vil forsøke å adressere.

Det er ikke hensiktsmessig her å trekke en streng grense for hva som er evalueringer og hva som faller utenfor begrepet. Oppgaven har ikke gjort et utvalg av evalueringer, for så å følge disse spesifikt. Det er heller det samlede evalueringssystemet som er forsøkt kartlagt, noe som reduserer behovet for en streng avgrensning. I en eksplorerende oppgave slik som denne, kan det snarere tvert imot være viktig å få frem det store mangfoldet i evalueringspraksiser. Enkelte vil kanskje mene at det er nødvendig med en presisering og avgrensning, men den saken er ifølge Halvorsen og Madsen allerede tapt (2013:16). Egen- eller internevalueringer er imidlertid holdt utenfor rammene av denne oppgaven, både fordi det synes som det er viet liten oppmerksomhet i litteraturen, men også fordi at ressursene som settes av til eksterne oppdragstakere (ofte private selskaper) vil ha et større krav til legitimitet, og derfor etter egen mening er av større interesse.

Det finnes stor variasjon i evalueringspraksis, og ulike organisasjoner vil kanskje også variere i hvorvidt de omtaler noe som en evaluering eller som en undersøkelse eller analyse. Det er derfor nødvendig å tilnærme seg evalueringbegrepet ut i fra hvilke fenomener man selv skal studere. Her er det ikke satt noen strengere grense enn at det er bruk av eksternt utførte evalueringer, altså at det offentlige utformer et konkurransegrunnlag som inneholder en beskrivelse av oppdraget, hva man ønsker å finne ut av og hva som er overordnet formål med evalueringen. Slike evalueringer er som regel utført av eksterne aktører, enten private konsultantselskaper, forskningsinstitutter eller ekspertpaneler (Ervik og Horrigmo 2015:3). Det evalueres enten før, underveis, eller i etterkant av et program. Det kan også dreie seg om evalueringen av hele organisasjonens virkeområde og organisering. Videre rapporteres funn, analyser og ofte anbefalinger til bestilleren, disse skal i tur føre til utvikling og forbedring (Ervik og Horrigmo 2015). Grunnen til at det er denne typen evaluering som skal undersøkes her, er at det gjennomføres en betydelig mengde slike evalueringer, og dermed brukes også betydelige ressurser. Evalueringer skal ha en funksjon, de skal brukes til noe. Spørsmålet er derfor hvordan de brukes?

1.4 HVORFOR EVALUERE?

Evalueringsbruk har lenge vært fokus for teoriutvikling og forskning. Hvordan evalueringer benyttes kan sies å være «*the core construct in the field of evaluation*» (Henry og Mark 2003:293). Målet med å gjennomføre evalueringer beskrives gjerne med overordnede perspektiver, slik som Henry og Mark, som mener at evalueringer skal føre til «*social betterment*» (2003:295) eller Herbert som sier at målet med å gjennomføre evalueringer er «*to affect positive social change*» (2014:389). Dette kan imidlertid oppleves som litt underlig da det er vanskelig å tenke seg at for eksempel symbolsk eller taktisk bruk av evaluering fører til «*social betterment*» (Højlund 2014: 29). En beskrivelse som kanskje passer bedre til praksisfeltet finner vi hos Weiss, som sier at «*the overall aim of evaluations is to assist people and organizations to improve their plans, policies and practices*» (Weiss 1999:469). Selv om det også her må sies å ligge et entydig rasjonelt utgangspunkt til grunn.

Ideelt sett kan vi si vi si at evaluering skal føre til en tilstand som er bedre enn den tilstanden som eksisterte før evalueringen, og at evalueringer dermed representerer et bidrag i endringsprosesser (Henry og Mark 2003) Dette perspektivet er et viktig utgangspunkt for denne oppgaven. Når man undersøker evalueringsbruk i en organisasjon, må det spores gjennom hvilke endringer evaluering har ført med seg. Endring bør da ses på en så bred måte som overhodet mulig, og ikke begrenses til kun å gjelde de direkte og «synlige» endringene. Det legges heller et syn på evalueringer som et mangfoldig fenomen med en rekke mulige formål, virkninger og egenskaper (Henry og Mark 2003, Kirkhart 2000).

I DFØs veileder *Strategisk og systematisk bruk av evaluering i styringen* (DFØ 2011) heter det at evalueringer «*bør være en sentral del av en virksomhets styring- og resultatinformasjon*». Samtidig skal evalueringer også bidra med læring og utvikling, utover mål- og resultatstyringen (DFØ 2011:5). Evaluering beskrives som et verktøy som skal sikre effektiv ressursbruk og gode beslutninger (DFØ 2011).

Som det fremgår ovenfor eksisterer det en rekke ulike former for evaluering. Det kan for eksempel dreie seg om en *utredning*, *formativ* eller *summativ* evaluering (Askim, Døving & Johnsen 2013:74-76). I tillegg vil det være stor variasjon i hva som evalueres, enten det er et fast tiltak, et tidsbegrenset tiltak eller evaluering av organisasjonsform (Ibid). De ulike evalueringsformene og hva som evalueres vil åpenbart innebære ulike aktiviteter, også når det gjelder hvordan evalueringen skal brukes underveis i prosessen eller etter at den foreligger. Denne oppgaven legger til grunn at evalueringer ikke kun har klare instrumentelle mål som

passer inn i rammen av et rasjonelt perspektiv på organisasjoner. Det må også ses i sammenheng med et institusjonelt perspektiv, hvor evalueringer gjennomføres som en rituell handling, og for å vise modernitet (Dahler-Larsen 2012). Det innebærer at evaluering kan ses som et mål i seg selv (Furre og Horrigmo 2013:181).

Hvorfor man evaluerer må naturligvis ses i nær sammenheng med hvordan evalueringer brukes. I teorikapittelet, hvor jeg gir en nærmere redegjørelse for brukstypene, vil formål og bruk ses i sammenheng.

1.5 NORSK EVALUERINGSPRAKSIS - OMFANG OG BRUK

Hvordan evalueringer brukes er et viktig demokratisk anliggende i ethvert moderne samfunn. Utviklingen av evaluering som et særskilt og selvstendig felt har kommet relativt sent til Norge, men har til gjengjeld skutt rask fart. Moderniseringsbølgen generelt, og NPM spesielt, har bidratt til å bane vei for etterspørselen av evalueringer i Norge (Dahler-Larsen 2013:46) Økt etterspørsel kan med andre ord ses i lys av den vertikale spesialiseringen som har foregått de siste tiårene, hvor det oftere enn før styres på «en armlengdes avstand» (Askim et. al. 2013:72).

Jeg har tidligere nevnt at det finnes lite systematisk erfaringsmateriale om evalueringsbruk i Norge. Det er imidlertid i skrivende stund igangsatt et oppdrag fra DFØ, hvor formålet blant annet er å kartlegge hovedtrekk ved rapportene som er publisert på evalueringsportalen, herunder å kartlegge hva som blir evaluert, hvordan det blir evaluert og hvordan et utvalg rapporter har blitt brukt (DFØ 2016, konkurransegrunnlag). Arbeidet med kartlegginger vanskeliggjøres av at det er utfordrende å avgrense evalueringer fra andre typer utredninger og analyser (Foss Hansen 2003).

I 1997 ble det gjennomført en lignende undersøkelse i Norge, hvor departementenes evalueringspraksis ble kartlagt på oppdrag fra Statkonsult. Undersøkelsen var basert på departementenes egen innrapportering gjennom et registreringskjema. Spørsmålene som ble tatt opp var blant annet knyttet til hvem som var initiativtaker, type og ressursbruk. Det inngikk også en nærmere undersøkelse av 100 evalueringer, blant annet om hvordan evalueringene ble fulgt opp, og hvordan og av hvem de ble brukt (Foss Hansen 2003: 66-68)

Konklusjonene var at evalueringer ble gjennomført i et betydelig omfang i sentraladministrasjonen, men med store variasjoner mellom departementene både med hensyn til antall og ressursbruk. Det var tydelig at det ble satt av betydelige summer til formålet. Over

90 prosent av respondentene ga uttrykk for at evalueringene var viktige som beslutningsgrunnlag, og at konklusjonene og anbefalingene i evalueringsrapportene ble brukt, først og fremst til å forbedre eksisterende tiltak. Det ble særlig fremhevet at evalueringene ga mer systematisk kunnskap enn det departementene var i stand til å tilegne seg selv. Kvaliteten på evalueringen ble vurdert som høy, og det ble rapportert om at de i stor grad svarte på de problemstillingene som var reist (Statkonsult 1997:11-30).

Undersøkelsen pekte videre på at de fleste av de gjennomførte evalueringene var effekt- og virkemiddelanalyser og at de ble brukt mest i forbindelse med oppfølging av omorganisering og reformvirksomhet. Evalueringer viste seg også å være utbredt på områder som i stor grad hadde offentlighetens oppmerksomhet rettet mot seg (Ibid).

Undersøkelsen pekte også på at evalueringen i liten grad ble fulgt opp i gjennomføringsprosessen, men med en tettere oppfølging i start- og slutfasen av prosjektene. Evalueringene ble vurdert som godt organisatorisk forankret, ved at det var fagavdelingene som sto ansvarlig for både initieringen av- og som brukere av evalueringsresultatene (Ibid)

Askim. et. al. (2013) trekker på informasjon fra evalueringsportalen i tidsrommet 2005-2011, når de systematiserer evalueringspraksis inn i evalueringens betingelser, objekt, type og metode. En slik klassifisering gjør det mulig å studere offentlig evalueringspraksis, for eksempel ved å se hvilke politikkfelt som blir grundig evaluert, og hvilke som har et mindre erfaringsbasert kunnskapsgrunnlag (Askim et. al. 2013:85). Når det gjelder betingelser, finner de at det oftest er tiltakseier selv, altså den underliggende etaten, som bestiller evalueringer og ikke det ansvarlige departementet. Det er departementsområdene arbeid- kunnskap- og miljø som bestiller flest. Forskningsinstitutter, universiteter og høyskoler var tidligere de mest brukte evalueringsmiljøene, men det ser ut til at private konsultentselskaper nå har gått forbi. Dette reflekterer at evalueringsarbeid ofte befinner seg i gråsonen mellom oppdragsforskning og avansert konsulentarbeid (Askim et. al. 2013:78-86) De finner videre støtte for at evalueringer oftere gjennomføres innenfor saksfelt som er preget av tidsavgrensede eller midlertidige tiltak, for eksempel innenfor bistand og næringsutvikling, snarere enn innenfor områder som er mer preget av faste tiltak. Summativ evaluering (etterevaluering) er den vanligste typen, dernest formativ (underveisevaluering), mens utredninger og samfunnsøkonomiske analyser gjerne forekommer innenfor sektorene transport, olje og skatt. En enkelt evaluering omfatter gjerne flere typer (Ibid).

Foss Hansen (2003) gjennomgikk og analyserte 452 evalueringsrapporter fra dansk statsforvaltning. Likheter mellom nordiske politiske systemer gjør at analysene fra dansk evalueringspraksis også kan være relevant i en norsk sammenheng. Funnene støtter et bilde av evalueringer som et forvaltningspolitisk ledelsesredskap og som en dialogbasert læringsprosess. Det bidrar med dokumentasjon og kontroll over aktiviteter, resultater og måloppfyllelse. Det resulterer også i anbefalinger, som sammen med andre former for input, kan bidra i beslutningsprosesser (Foss Hansen 2003: 102-105). Det understrekes at Hansens analyse har undersøkt innholdet i evalueringsrapporter, og ikke hvordan forvaltningen selv rapporterer om bruk (Foss Hansen 2003: 102-105).

1.6 KONTEKSTEN ER INNOVASJON NORGE

Konteksten en evaluering gjennomføres i, har betydning for om og hvordan den blir brukt. Man har i litteraturen blitt mer oppmerksom på at evaluering er avhengig av den sosiale- og organisatoriske konteksten den foregår innenfor (Dhaler-Larsen 2013:34).

Alkin og Taut (2003) deler inn trekk ved selve evalueringen, trekk ved brukeren og trekk ved konteksten som variabler som kan påvirke bruk. I Cousins og Leitwood (1986) trekkes det frem trekk ved implementering og trekk ved «policy settingen» som uavhengige variabler. Når det gjelder trekk ved implementeringen kan det for eksempel dreie seg om kvaliteten på evalueringen, om den oppleves som relevant for virksomheten og om den blir kommunisert på en god måte. Trekk ved policy-settingen kan handle om hvorvidt det er behov for informasjonen som gis, hvordan det politiske klimaet er, om det finnes konkurrerende informasjon eller hva slags kompetanse og holdninger som eksisterer hos de ansatte.

Et annet sentralt begrep som kan knyttes til trekk ved brukeren, er evalueringskapasitet, som viser til den evnen en organisasjon har til å gjennomføre og bruke evaluering. Det kan for eksempel dreie seg om størrelse på virksomheten, strategi, organisatorisk forankring, kompetanse etc. (Halvorsen og Madsen 2013:20).

Jeg understreker at disse bakgrunnsvariablene ikke skal undersøkes direkte, men at det likevel er viktig å trekke frem en slik forståelse fordi det har betydning for å vurdere ekstern validitet, altså i hvilken grad bruk av evaluering i dette tilfellet kan vurderes å være gjeldene i andre offentlige virksomheter. I tillegg tar oppgaven sikte på en bred kartlegging av INs evalueringsvirksomhet og funn som knytter seg til slike variabler vil da være interessante i seg

selv. For å danne en grunnleggende forståelse for caset som undersøkes vil jeg i det følgende beskrive trekk ved næringspolitiske virkemidler og selve virksomheten Innovasjon Norge.

1.6.1 Næringspolitikk i Norge

I Norge føres en aktiv næringspolitikk. Formålet med det er å legge til rette for samfunnsøkonomisk lønnsom produksjon, effektiv ressursbruk og bærekraftig vekst. I en situasjon hvor markedet ikke fungerer perfekt, vil det være nødvendig med reguleringer og inngrep for å justere for markedets imperfeksjoner. Staten har derfor utviklet en rekke virkemidler som skal bidra til å utløse samfunnsøkonomiske prosjekter som ellers ikke ville blitt iverksatt (Meld. St. 22 (2011-2012):20-21).

Det er imidlertid utfordrende å gi konkrete svar på hvor stor den samfunnsøkonomiske lønnsomheten er ved å iverksette næringsrettede virkemidler. Jo større denne usikkerheten er, jo mer utfordrende er det å ha kunnskap om hvordan ressursene bør fordeles gjennom det samlede virkemiddelapparatet. Usikkerheten bidrar også til at det er vanskelig å evaluere hvor berettiget et virkemiddel er (Ibid:21). Et viktig spørsmål er derfor om «*virkemiddelet faktisk gir en samfunnsøkonomisk meravkastning, eller om den bare kommer til erstatning for privat investeringsinnsats*»? Dette fører igjen til at «*virkemiddelapparatet daglig bør gjøre denne typen vurderinger*» (Meld. St. 22 (2011-2012):21).

1.6.2 Innovasjon Norges virksomhet

Når man snakker om virkemiddelapparatet på næringspolitikken, snakker man om alle offentlige institusjoner som har fått i oppdrag å forvalte ulike offentlige programmer og tiltak som skal bidra til forskning, innovasjon og næringsutvikling (Meld. St. 22 (2011-2012)). Dette systemet har i Norge vokst frem over tid og det finnes flere aktører på feltet. En av disse institusjonene, og kanskje den viktigste, er Innovasjon Norge. IN skal være pådriver og partner for bedriftene som mottar støtte, men skal også fungere som rådgiver for myndighetene i næringspolitiske spørsmål. IN står selv for kontakten mot bedriftene og skal selv foreta vurdering av deres behov (Meld. St. 22 (2011-2012)).

Innovasjon Norge ble opprettet i 2004 gjennom en sammenslåing av flere mindre aktører og har nå ansvar for forvaltning av virkemidlene som tidligere var tillagt Statens nærings- og distriktsutviklingsfond (SND), Statens veiledningskontor for oppfinnere, Norges Turistråd og Norges Eksportråd. Formålet med å slå sammen de ulike aktørene var å skape en mer helhetlig

og brukervennlig organisasjon, slik at næringslivet skulle få vurdert behovene på ett sted, og slippe å bli sendt mellom de ulike aktørene (Ibid).

I Styrets beretning fra 2015 heter det at:

Innovasjon Norges formål er å være statens og fylkeskommunenes virkemiddel for å realisere verdiskapende næringsutvikling i hele landet.

Hovedmålet med Innovasjon Norge er at selskapet skal utløse bedrifts- og samfunnsøkonomisk lønnsom næringsutvikling og ulike regioners næringsmessige muligheter. Dette skal nås gjennom delmålene flere gode gründere, flere vekstkraftige bedrifter og flere innovative næringsmiljøer.

Innovasjon Norge mottar årlige oppdragsbrev fra departementer, fylkeskommuner og fylkesmenn. Med utgangspunkt i statsbudsjettet gir oppdragsbrevene en samlet oversikt over oppdragsgivernes forventninger og krav knyttet til Innovasjon Norges virksomhet. Brevene gir videre oversikt over Innovasjon Norges bevilgningsrammer og føringer knyttet til satsinger, programmer og tjenester.

IN tilbyr finansieringstjenester, kompetansetjenester, rådgivningstjenester, nettverkstjenester og profileringstjenester, og bidro i fjor med til sammen 6,1 milliarder kroner til næringslivet. I 2015 ble selskapet omstrukturert og består nå av fem divisjoner fordelt på virksomhetsområdene Grundere og oppstartbedrifter, Vekstbedrifter og klynger, Regioner og finansiering, Merkevarer Norge og Bærekraft Norge 2030 (Innovasjon Norge 2015).

IN har ansvaret for å forvalte midler fra en rekke aktører på tvers av mange sektorer. Virksomheten mottar oppdrag fra Nærings- og handelsdepartementet, Kommunal og regionaldepartementet, Fiskeri- og kystdepartementet, Landbruks- og matdepartementet, Utenriksdepartementet, og samtlige fylkeskommuner og fylkesmenn (Meld. St. 22 (2011-2012):23).

IN er organisert som et særlovselskap i henhold til lov om Innovasjon Norge. Det innebærer at organisasjonene er et eget rettssubjekt med et selvstendig og faglig ansvar knyttet til enkeltsaker. Virksomheten har omlag 700 ansatte, er organisert på 22 distrikts- og lokalkontorer i Norge og har i tillegg det som kalles utekontorer i omlag 30 land (Årsrapport 2015, Meld. St. (2011-2012):23).

1.6.3 Krav til måling

I Styrets beretning for 2015 trekkes det videre frem at virksomheten som forvalter av offentlig midler har fokus på å dokumentere måloppnåelse og effekt. På et slikt virksomhetsområde er det imidlertid krevende å dokumentere både kortsiktige og langsiktige effekter, fordi det også er mange andre forhold enn INs bidrag som spiller inn på suksessen (Innovasjon Norge 2015).

Ved bruk av offentlige midler er det naturligvis krav til at ressursene brukes på en måte som er til nytte for samfunnet. Når det gjelder næringsrettede virkemidler snakker man gjerne om markedssvikt, som tar utgangspunkt i at det finnes svakheter i markedet som bør korrigeres. Man snakker også gjerne om addisjonalitet, som dreier seg om at ressursene som gjøres tilgjengelig til tilskuddsmottaker også skal utløse ønskede tilleggsaktiviteter- eller tilleggsproduksjon (Madsen, Alsos, Clausen & Ljungren 2013:59). Dette er komplekse prosesser, og dermed utfordrende å måle, fordi det innebærer å skulle si noe om hvordan situasjonen er nå, sammenlignet med hvordan det hadde vært hvis virkemiddelet ikke hadde eksistert (Ibid: 70).

Det at det er vanskelig å måle effekter bidrar kanskje til et økt behov for dokumentasjon fra omgivelsene om hva som virker og ikke virker. Samtidig som det også bidrar til at virksomheten selv er avhengig av systematiske undersøkelser for å kunne få informasjon om måloppnåelse og effektivitet. Det finnes imidlertid andre politikfelt hvor det også er utfordrende å dokumentere effekter, slik at IN ikke nødvendigvis er i en særstilling i denne sammenhengen. Tidligere har jeg vist både til Norad og Utdanningsdirektoratet som eksempler på at evalueringsvirksomheten er institusjonalisert, i betydningen av at det er opprettet egen avdeling for evaluering. Disse tre politikfeltene har kanskje det til felles at det er utfordrende å dokumentere effekter og i tillegg er et relativt politisert felt.

I INs tilfelle må det i hvert fall sies å være rettet betydelig oppmerksomhet fra media, slik NRK-debatten 28.04.16 om gründerstøtte illustrerer. I tillegg har det vært rettet kritikk fra Civita (Nordbakken 2015) om hvordan effekten av INs virksomhet blir målt. Samtidig som det fremkommer i Dagsavisen (04. 01.2016) at IN har blitt kritisert for svak rapportering på måloppnåelse fra Nærings- og fiskeridepartementet. Det er også en rekke andre artikler og kommentarer som gjelder IN og deres virksomhet, noe et enkelt Google søk vil vitne om.

1.6.4 Hvorfor er Innovasjon Norge et interessant case?

Det som er nevnt ovenfor gjør utvilsomt at Innovasjon Norge er et interessant case. Men det er også interessant fordi de gjennomfører et stort antall evalueringer og har utviklet prosedyrer og systemer på hvordan kunnskap fra evalueringer skal benyttes i virksomhetens styring- og utvikling. Ansvar for gjennomføring av evalueringer er skilt ut i en egen avdeling og må derfor sies å være institusjonalisert (Foss Hansen 2003).

På evalueringsportalen er det registrert 72 unike evalueringer hvor IN med forløpere var registrert som oppdragsgiver. Fra 2004-2014, når IN eksisterer i sin nåværende form, er det registrert 59 evalueringsrapporter. I tillegg finnes det minst tre evalueringer hvor daværende Nærings- og handelsdepartementet var oppdragsgiver, deriblant «Evaluering av Innovasjon Norge» fra 2010, som er en evaluering av hele INs virksomhet. Jeg understreker at de siste som er nevnt, kun er rapporter hvor Innovasjon Norge var nevnt i tittelen på prosjektet, slik at det også kan finnes andre som er bestilt fra andre departementer. I én evalueringsrapport var Kommunal- og moderniseringsdepartementet (KMD) og Nærings- og handelsdepartementet (NHD) sammen oppdragsgivere. I tillegg kommer Riksrevisjonens undersøkelse av «Innovasjon Norge som statlig næringsutviklingsaktør» fra 2008.

11 av rapportene som er registrert har «kundeeffektundersøkelse» (KEU) som en del av tittelen. Jeg var innledningsvis inne på at evaluering kan omfatte en mengde ulike undersøkelser, og det kan eksistere ulike oppfatninger om dette kan betegnes som en evaluering. Siden dette studie ikke tar utgangspunkt i spesifikke evalueringsrapporter, men heller søker å danne et bredt bilde av evalueringsbruk, er det ikke av avgjørende betydning å begrense hvilke typer undersøkelser som faller utenfor eller innenfor evalueringsbegrepet. Det er likevel verdt å nevne at det i en av KEU-rapportene opplyses om at *«kundeeffektundersøkelsen er en av flere undersøkelser som inngår i et totalt evalueringssystem for Innovasjon Norge»* (Meltevik et. al. 2010:15). Samtidig som det i INs eget evalueringsstrategidokument (2011) nevnes at evaluering regnes først og fremst som «øvrige evalueringer», som innebærer at man ser nærmere på årsaks-virkningsforhold i den enkelte tjeneste, eller en *gruppe av tjenester*. Dette skal for øvrig diskuteres nærmere i resultat og analysekapittelet.

For å oppsummere kan vi si at IN er et interessant case fordi evalueringsvirksomheten er institusjonalisert, med egne ansatte med lang erfaring på evalueringsfeltet. IN er som tiltakseier selv bestiller av evalueringer, heller enn overliggende departement. Det eksisterer et komplekst nettverk av eiere og oppdragsgivere. Noe som innebærer at det er mange å rapportere til på,

tvers av sektorer. Det er også et felt hvor det er utfordrende å måle effekter, noe som kan bidra til økt krav om at effekter faktisk måles. I tillegg kan det sies å være et relativt politisert felt.

2. TEORETISK RAMMEVERK

Evalueringsbruk har lenge vært fokus for teoriutvikling, forskning og debatt. Litteratur om evaluering og evalueringsbruk har imidlertid gjerne vært konseptuell og teoretisk, og har sjelden involvert systematisk forskning (Henry og Mark 2003:294). Når en evaluering blir gjennomført er det mye som kan skje. Enkelte konsekvenser er direkte, mens andre er mer indirekte og «sniker» seg inn i beslutningsprosesser (Henry og Mark 2003:295; Weiss 1980). I andre tilfeller dreier det seg snarere om evalueringens funksjon som legitimerende middel og uttrykk for modernitet (Røvik 2007).

I denne oppgaven undersøkes evalueringsbruk på organisasjonsnivå. Det er ikke undersøkt innenfor rammene av individer og dermed kognitive og atferdsmessige modeller, slik som enkelte forskere trekker frem (Kirkhart 2000; Henry og Mark 2003; Mark og Henry 2004;).

Nedenfor skal jeg først sette bruk av evaluering inn i et større organisasjonsteoretisk perspektiv. Deretter vil jeg gi en relativt detaljert redegjørelse for de ulike brukstypene som oppgaven tar sikte på å undersøke og som Henry og Marks har karakterisert som «*the fundamental taxonomy of use*» (Henry og Mark 2003:295). Mot slutten skal jeg kort diskutere hvordan bruk av evaluering bør forstås med en bred tilnærming.

2.1 UTGANGSPUNKT - ORGANISASJONSTEORI

Et første utgangspunkt for å redegjøre om evalueringsbruk er å sette det inn i en organisasjonsteoretisk ramme. Grunnen til at det gjøres, er at evaluering er et fenomen som utspiller seg innenfor organisasjoner, hvor de mest innflytelsesrike ideene på feltet er organisasjonsideer. Evaluering har blitt en institusjonalisert praksis og organisasjonsteori kan da bidra med å forstå hvorfor og hvordan evaluering utspiller seg (Dhaler-Larsens 2012:33). Dhaler-Larsen beskriver evaluering innenfor tre organisasjonsteoretiske perspektiver, det rasjonelle, det lærende og det institusjonelle.

Røvik (1998) gjør et lignende skille mellom verktøyperspektivet som er utledet fra den rasjonelle-instrumentelle tradisjonen, hvor oppmerksomheten er rettet mot en sterkt tro på forandring og forbedring gjennom planmessig strukturering og læring, og symbolperspektivet som er utledet fra ny-institusjonell teori, hvor oppskrifter ikke kun blir sett på som verktøy for effektiv problemløsning, men også til å skape legitimitet og identitet (Røvik 1998).

2.1.1 Rasjonelt perspektiv

Innenfor det *rasjonelle organisasjonsperspektivet* ser man organisasjoner som instrumenter som er designet for å nå spesifikke mål. Organisasjoner er funksjonelt utformet for å koordinere handlinger, slik at man når forutbestemte mål på en så effektiv måte som mulig (Scott og Davis 2007:35). Formulering av spesifikke målsettinger gir entydige kriterier for valg mellom ulike alternativer. Disse alternative vurderes innenfor en klar hierarkisk preferanseorden (Ibid:36). Organisasjoner sikter også på å maksimere forutsigbarhet. Dette gjøres ved å utforme standardiserte prosedyrer som best fører til måloppnåelse. Organisasjoner definerer problemer og løsninger på disse problemene ved at det implementeres en *formell rasjonalitet*, gjennom regler og prosedyrer som gir forutsigbarhet og effektivitet. Videre foregår det en spesialisering ved at arbeidet fordeles i ulike avdelinger med spesialiserte oppgaver og ansvar (Dhaler-Larsen 2002:38-41). En slik formalisering skal innenfor dette perspektivet bidra med stabilitet og forutsigbarhet i adferdsmønstre (Scott og Davis 2007:37).

I dette perspektivet er organisasjonsstrukturer og virkemidler som blir implementert, gjenstand for endring og modifisering dersom det er behov for å forbedre resultater (Scott og Davis 2007:38). Evalueringer kan dermed bidra til å sikre nødvendig informasjon, slik at man kan justere og produsere bedre strategier og prosedyrer. Altså forventes det at evaluering er et instrument (Dhaler-Larsen 2002:38-41).

Siden 80-tallet har mål- og resultatstyring vært innført som et sentralt styringsprinsipp i staten. Dette innebærer at offentlige virksomheter setter mål for hva som skal oppnås, måler resultater og sammenligner resultater med målene. Ideelt sett skal informasjonen som innhentes om effektivitet og måloppnåelse brukes til styring, kontroll og læring for å utvikle virksomhetene (DFØ 2016). Det må sies å ligge et klart rasjonelt perspektiv til grunn i denne logikken. Man forbinder det også gjerne med organisasjonsideer som har fulgt med NPM-bølgen. Selv om NPM ikke er en enkelt-teori, men heller en samling av ideer, spores det gjerne tilbake til public choice-teori og prinsipal-agent teori, som begge er retninger som legger et rasjonelt perspektiv til grunn (Eliassen og Sitter 2008, Howlett, Ramesh & Perl 2009:22-34, Røvik, 2007:36).

Det finnes også eksempler på at det rasjonelle perspektivet er nedtonet. Ved at man ser individer og organisasjoner som *bundet rasjonelle*. Det vil si at man er rasjonell i forstand av at man søker etter å maksimere nytte, basert på et sett av stabile mål og preferanser, men at man ikke alltid vil være i stand til å forstå alle problemer og mulige løsninger, grunnet mangel på informasjon (Scott 2014:80). I et slikt perspektiv på organisasjoner kan evaluering dermed

bidra til å redusere informasjonsmangel. Dette må også sies å være nært beslektet med det neste perspektivet som Dhaler-Larsen trekker frem.

2.1.2 Læringsperspektiv

Det *lærende* organisasjonsperspektivet blir av Dhaler-Larsen fremstilt som et alternativ til en rasjonell modell, men har likevel viktige likhetstrekk. Det kan derfor nevnes at dette perspektivet vanligvis ikke fremstilles som en selvstendig tilnærming, på samme måte som skillet mellom et rasjonelt og institusjonelt perspektiv (Røvik 2007).

I det lærende perspektivet fremholdes det at en organisasjons verdier og mål ikke er konsistente, og at ulike aktører har ulike problemoppfatninger og løsningsforståelse. (Dhaler-Larsen 2012). Dette kan også ses i sammenheng med et klassisk perspektiv på at organisasjoner er organiserte anarkier. Det innebærer at en organisasjon står ovenfor en strøm av ulike problemoppfatninger, mulige løsninger, preferanser og tolkninger. Dette fører i tur til beslutningsmuligheter, som kan eksistere på et gitt problem til en gitt tid, men ikke nødvendigvis på et senere tidspunkt (March og Olsen 1972). Dette skiller seg klart fra et rasjonelt perspektiv, fordi beslutninger i en organisasjon ikke bare er et resultat av en klar nytte-kostnad tenkning, men også basert på tilfeldigheter. Faktisk adferd er med andre ord ikke nødvendigvis så sammenkoblet med formelle prosedyrer og organisasjonsstrukturer som antatt i et rasjonelt perspektiv (Røvik 2008).

Aktørene i en organisasjon har begrenset informasjonskapasitet og forståelse for konsekvensen av tidligere avgjørelse, og har derfor også begrenset mulighet til å planlegge og velge fremtidige alternativer. For å kunne løse utfordringer er organisasjoner derfor avhengig av «feedback» (Dhaler-Larsen 2003). Evaluering er et sentralt verktøy i en slik «feedbacksprosess». Dette perspektivet knytter sammen evaluering med lærings- og endringsprosesser, hvor evaluering bidrar med innsikt som kan benyttes i det praktiske arbeidet (Ibid).

Cohen og Levinthal (1990) argumenterer for at en organisasjons evne til å lære, eller det de kaller *absorberende kapasitet* er et sentralt trekk ved enkelte organisasjoner. Det dreier seg om hvorvidt organisasjoner er i stand til å anerkjenne verdien av ny ekstern informasjon, og hvilken evne de har til å absorbere og benytte seg av slik kunnskap. Videre argumenteres det for at dette i stor grad hviler på tidligere kunnskap, altså at organisasjonslæring er kumulativt. Derfor er det vesentlig for organisasjoners suksess at det avsettes tilstrekkelig med ressurser til forskning og utvikling (Cohen og Levinthal 1990). I sammenheng med evaluering vil man på samme måte

kunne snakke om evalueringskapasitet eller organisatorisk beredskap for evaluering (Halvorsen og Madsen 2013). I et slik perspektiv vil gjennomføring og bruk av evaluering potensielt være et viktig bidrag i læringsprosessen.

2.1.3 Institusjonelt perspektiv

I et *institusjonelt* perspektiv på organisasjoner, tar man utgangspunkt i at individer også handler etter andre prinsipper enn rasjonelle beregninger og refleksjoner (Dahler-Larsen 2012). Det er ikke tilstrekkelig å forklare organisasjoner som rasjonelt sammenhengende og autonome, fordi adferd også styres av helt andre faktorer. Adferd vil for eksempel være betinget og begrenset av kulturelle og sosiale normer. Selv om egeninteresse også vil være styrende, er adferd ofte mer basert på det som er normativt passende, heller enn beregning av nytten til alternative valg (March og Olsen 1984:744).

Viktig innenfor dette perspektivet er at kulturelle konstruksjoner blir forstått av mennesker som objektive sannheter. Selv om organisasjoner er sammensatt av menneskelige konstruksjoner opplever man det ikke på den måten. I en organisasjon skapes det dermed en felles forståelse av verdier, organisasjonsoppskrifter og rutiner som blir reproduisert og tatt for gitt (Dahler-Larsen 2012:57). Organisasjoner må også søke legitimitet fra omgivelsene ved å demonstrere at man lever opp til fundamentale modernitetsnormer om kontinuerlig fremskritt, fornyelse og rasjonalitet. Slike ideer kommer gjerne til uttrykk både gjennom generelle idéer og tydelige oppskrifter (Røvik 2008: 26). Det som kjennetegner slike populære oppskrifter er blant annet at de er *allestedsnærværende* og sprer seg på tvers av organisasjoner og land. De er *gjennomtrengende* og preger i stor grad strukturer, aktører og aktiviteter. Ofte har de også *tvetydige virkninger*. Det vil si at det ofte er betydelig usikkerhet knyttet til hvordan de virker, altså hvor effektive de er som verktøy (Røvik 2008: 16-19).

I et institusjonelt perspektiv gjennomføres og brukes evalueringer på en rituell og symbolsk måte som en vane, og for at det forventes av en moderne organisasjon fra de eksterne omgivelsene (Dahler-Larsen 2012:65). Evaluering blir da et; «*taken for granted aspect of public administration and organizational procedures for management and development*» (Dahler-Larsen 2012:2).

Jeg vil hevde at alle brukstypene som presenteres nedenfor, kan forstås i lys av disse tre organisasjonsteoretiske retningene. Det er imidlertid tydelig at et perspektiv på egenhånd ikke dekker en teori om organisasjoner generelt, eller om evalueringsbruk spesielt. Evaluering kan

ikke alene tilbyr entydige løsninger, slik at en organisasjon løser sine problemer og når sine mål. Det brukes heller ikke bare som «feedback» inn i et læringssystem, og heller ikke kun som en symbolsk og rituell handling. Det er derfor viktig å ta hensyn til at mange typer bruk kan være i spill samtidig, og at det disse vil kunne spores tilbake til både et rasjonelt og institusjonelt perspektiv på organisasjoner.

2.2 SEKS TYPER EVALUERINGSBRUK

Litteraturen på evalueringsbruk kan knyttes tett til en større litteraturstrøm om bruk av forskningsresultater og kunnskap i politikktutforming (Højlund 2014), hvor tidlig litteratur på feltet ikke spesifikt behandlet bruk av evaluering men bruk av samfunnsvitenskap mer generelt (Levinton & Hughes 1981:526). Bruk av evaluering utviklet seg imidlertid tidlig til å omfatte minst fem hovedtyper som har vært vanlig å benytte i litteraturen; instrumentell, konseptuell, legitimerende, taktisk og rituell/symbolsk. Denne typologien og lignende kategoriseringer trekkes frem i de fleste artiklene om bruk av evaluering, og synes derfor fremdeles å være relevant (Rich; 1977; Weiss 1977; Levinton & Hughes 1981; Johnsen 1998; Alking & Taut 2003; Henry & Mark 2003; Vedung 2009; Gildemyn 2014; Herbert 2015). Enkelte har også lagt til en sjettede type, prosessuell bruk av evaluering, som vil si at individer og organisasjoner lærer å tenke på en «evaluerende» måte gjennom å delta i evalueringsprosesser (Alking & Taut 2003; Patton 2008; Vedung 2010) I det følgende skal vi se nærmere på hver enkelt type.

2.2.1 Instrumentell bruk

Instrumentell bruk av evaluering er kanskje det en intuitivt først vil knytte til bruksbegrepet. I utgangspunktet er det en relativ enkel ide om at beslutningstakere bruker evalueringsfunnene til å endre det som er gjenstand for evaluering (Halvorsen & Madsen 2013). I en tidlig definisjon av instrumentell bruk forklarer Robert F. Rich at instrumentell bruk er «*tilfeller hvor respondenten viser til og kan dokumentere spesifikke tilfeller hvor forskning har blitt brukt til å ta beslutninger*» (Rich 1977:200, egen oversettelse). På samme måte forklarer Henry og Marks (2003) at instrumentell bruk generelt referer til direkte handlinger som har blitt gjennomført som et resultat av en evaluering. Videre hevder de at dette er et kjernebegrep som stort sett alle de som gjennomfører evalueringer vil streve for å oppnå (Henry & Mark 2003). I dette klassiske perspektivet på evalueringsbruk er altså målet for å gjennomføre en evaluering

at informasjonen som den produserer benyttes på en direkte måte, i beslutningstaking og tiltaksforbedring (Henry & Mark 2003).

Selv om bruksbegrepet som vi skal se, også er tillagt flere dimensjoner og bredere forståelse, må instrumentell bruk sies å fremdeles være et klart mål med å gjennomføre evalueringer. En evalueringsrapport inneholder for eksempel ofte konkrete punkter med forslag til endring av systemer og prosedyrer i en organisasjon eller et program. Hvorvidt slike forslag manifesterer seg i faktisk endring og implementering av nye beslutninger gjenstår imidlertid å se.

2.2.2 Konseptuell bruk

Konseptuell bruk beskrives av Rich (1977) som en «*påvirkning av beslutningstakeres tenkning om et tema, uten at informasjonen blir omsatt til spesifikk og dokumenterbar bruk*» (Rich 1977:200, egen oversettelse). Denne bruken av evalueringsfunn bidrar til at organisasjoner og beslutningstakere forstår tiltaket på en ny måte og dermed påvirker kunnskap, tenking og begrepsbruk uten at det fører til konkrete beslutninger som kan spores direkte tilbake til en evalueringsrapport. Det kan med andre ord bidra til å forme beslutningstakeres, medarbeideres og allmenhetens måte å tenke rundt et program eller tiltak på (Weiss 1979:429). Weiss har kalt dette «*Enlightenment Model*» hvor forskning og evalueringer bidrar til å skape orden i en kompleks verden (Weiss 1979:429).

Cousins og Leithwood peker også på det de kaller bruk som utdanning. Det innebærer at evaluering ikke brukes som grunnlag for direkte beslutninger, men heller til å «opplyse» beslutningstakere, ved at evaluering for eksempel påvirker oppfatninger om nåværende tilstand og ideelle måter å strukturere et tiltak på (Cousins og Leithwood 1986:345). En evaluering bidrar gjerne med å peke ut styrker og svakheter ved et virkemiddel. Bruk som utdanning har derfor blant annet blitt operasjonalisert ved at respondenter har rapportert om i hvilken grad en evalueringsrapport har påvirket de ansattes synspunkter om et program. Enkelte har også inkludert saker som begrepsutvikling, bekrefte tidligere antagelser og styrke moral som utdanningsbruk (Ibid).

I dette perspektivet er evalueringer kilde til ny kunnskap og informasjon på feltet, som setter beslutningstakere og ansatte i stand til å ta kunnskapsbaserte avgjørelser (Weiss 1979). Med andre ord kan vi si at det legges et tydelig læringsperspektiv til grunn.

2.2.3 Legitimerende bruk

Legitimerende bruk er når innsikt og informasjon fra evalueringer anvendes for å «rettferdiggjøre, forsvare og bekrefte allerede eksisterende overbevisninger» (Vedung 2009: 264). I dette perspektivet endres ikke eksisterende standpunkt, men nye argumenter kommer i tillegg. Det kan tenkes to situasjoner hvor slik bruk kommer til uttrykk. I det ene tilfellet er brukeren overbevist om riktigheten av sin egen posisjon eller beslutning, men vil ha faktabasert støtte for sitt synspunkt, og bruker evalueringsresultater for å overbevise meningsmotstandere eller tvilere. I det andre tilfellet kan brukere være noenlunde overbevist om at sin egen posisjon er rimelig, men det eksisterer en viss tvil hvor man leter etter argumenter som kan støtte posisjonen (Vedung 2009).

På samme måte beskriver Weiss, i det hun har kalt den politiske modellen, at forskningsresultater kan brukes som «ammunisjon» til støtte for den posisjonen man allerede anser som korrekt. Evalueringsresultater kan dermed brukes som evidens for å nøytralisere motstandere, overbevise tvilerne eller styrke støttespillerne (Weiss 1977:14; 1979). I utgangspunktet kan kanskje slik bruk oppfattes som uredelig. Evalueringer som er utført på oppdrag fra offentlige virksomheter har ofte en iboende politisk dimensjon over seg og dersom resultater benyttes på en selektiv måte vil det kunne ha problematiske sider. Weiss gjør imidlertid et poeng av at bruk av evalueringsfunn som underbygger en eksisterende posisjon må anses som bruk på lik linje som andre kategorier, og er ofte hverken uviktig eller uriktig (Weiss 1979).

2.2.4 Taktisk bruk

Taktisk bruk går ut på at evalueringer benyttes for å unngå beslutninger på nåværende tidspunkt. Denne brukstypen innebærer at intensjonen bak gjennomføring av en evaluering ikke nødvendigvis er begrunnet i et mål om å frembringe kunnskap om et tiltaks resultater, men heller for å lede negativ oppmerksomhet bort fra tiltaket (Vedung 2009). Det kan for eksempel dreie seg om et tilfelle med dårlige resultater hvor man ønsker å forhindre kritisk debatt eller trenere beslutninger. På denne måten kan tiltakshaver vise at det er satt i gang en aktivitet som man kan henvise til for å «desarmere» konflikten og hindre kritiske røster som krever umiddelbare endringer (Vedung 2009). Ved å initiere og vise til at en evalueringsprosess er i gang kan myndighetene signalisere at noe blir gjort med et problem, samtidig som beslutninger og mulige tiltak blir utsatt eller unngått. Informasjon fra evaluering blir da ignorert i selve

beslutningsprosessen (Knorr 1977:171). Det er altså ikke innholdet i evalueringen man er ute etter å utnytte, men det faktum at en evaluering i seg selv blir gjennomført (Weiss 1979). Slik som Weiss formulerer det er det ikke uvanlig å høre uttalelser fra offentlige virksomheter som «vi er klar over utfordringene derfor har vi satt i gang en evaluering» eller «vi venter på resultatene fra evalueringen» - evalueringer blir dermed et bevis på evne til å reagere, uten at man nødvendigvis reagerer (Weiss 1979:429). I dette tilfellet kan det være naturlig å stille seg spørsmålet om dette er bruk i det hele tatt?

Skillet mellom legitimerende bruk og taktisk bruk kan synes noe uklar og overlappende. Vedung presiserer imidlertid at forskjellen mellom disse to kategoriene er at legitimerende bruk innebærer at evalueringens resultater blir tatt i bruk for å rettferdiggjøre virksomheter, intervensjoner og argumenter. Ved taktisk bruk benyttes selve evalueringsprosessen. Det kan derimot til å begynne med være at evalueringen gjennomføres av taktiske årsaker, men at andre brukskategorier kommer til uttrykk på et senere tidspunkt (Vedung 2009). Kategoriseringen er med andre ord ikke gjensidig utelukkende, og kan anvendes på flere ulike måter (Halvorsen og Madsen 2013).

Det er viktig å gjennomgå denne brukstypen fordi det er en del av den fundamentale typologien som trekkes frem i litteraturen. Men det er nødvendig å gjøre det klart at denne bruksformen ikke er forsøkt operasjonalisert og undersøkt i dette studie. Det bygger på en forventning om at det sannsynligvis vil være utfordrende å dokumentere de mer «spektakulære» bruksformene, slik som taktisk bruk (Foss Hansen 2003: 89).

2.2.5 Rituell eller symbolsk bruk

Rituell eller symbolsk bruk av evaluering viser til at evalueringer gjennomføres fordi det er moderne. Organisasjoner slutter seg til populære organisasjonsideer og adferd uten at det nødvendigvis har enn reell og synlig betydning. Dagens moderne samfunnsordning bygger på myten om fremskritt, og denne fremskrittmyten manifesterer seg i rituelle aktiviteter som planlegging og evaluering (Vedung 2009). I følge Vedung har evaluering i dag blitt en «megatrend» med utgangspunkt i at offentlig sektor ikke lenger er legitimert kun av regel- og prosedyrestyring, slik som i det tradisjonelle Weberske byråkratiet. Det er derfor behov for andre instrumenter som kan bidra til å legitimere virksomheten (Vedung 2010). Evalueringer kan være en aktivitet som annonseres, startes, gjennomføres og rapporteres på en rutinisert

måte. Dette er imidlertid ikke til hinder for at funn kan brukes på de andre måtene som er nevnt til nå (Vedung 2009).

Hos Vedung gjøres et tydelig skille mellom legitim, taktisk og symbolsk bruk. Hos andre er dette skillet mer uskarpt. For eksempel omtaler Burke Johnson (1998) symbolsk bruk som bruk av evalueringsinformasjon for å tjene politisk egeninteresse, noe som i denne oppgaven heller er plassert under taktisk bruk. I tillegg blir legitimerende og «persuasive» bruk ofte omtalt som nært beslektede kategorier (Johnson 1998:94) Patton plasserer også både «persuasive», legitimerende og politisk bruk under fanen symbolsk bruk (Patton 2008:104). Alkin og Taut (2003) hevder også at det ikke nødvendigvis gjøres et skarpt skille mellom symbolsk og legitimerende bruk, men argumenterer samtidig for at begrepet har vært forbeholdt de tilfeller hvor evaluering er gjennomført bare som et symbol eller en statushandling (Alkin og Taut 2003).

I organisasjonsteori har det innenfor den amerikanske neoinstitusjonelle tradisjonen vært vanlig å legge vekt på at organisasjoner må vise at man lever opp til fundamentale vestlige modernitetsteorier om kontinuerlig fremskritt, fornyelse og rasjonalitet. Disse normene kan uttrykkes gjennom ideer og oppskrifter på hvordan moderne organisasjoner bør fremstå og hva slags rutiner og prosedyrer som bør gjennomføres (Røvik 2007). Evaluering er en populær styringsoppskrift, og kan passe fint inn i et slikt perspektiv på organisasjoner.

2.2.6 Prosessuell bruk

Prosessuell bruk av evaluering er også en kategori som har fått feste i evalueringslitteraturen. Den viktigste bidragsyteren innenfor dette perspektivet må sies å være Michael Quinn Patton, som ifølge Herbert (2014) uttrykte denne formen for bruk første gang i 1986, i boken «Utilization Focused Evaluation». Denne tilnærmingen konsentrerer seg ikke bare om bruk av informasjon fra selve evalueringen, men heller at måten evalueringen gjennomføres på har en viktig innvirkning på individer og organisasjoner (Alkin og Taut 2003). Ifølge Patton viser prosessuell bruk til:

«Cognitive, behavioral, program, and organizational changes resulting either directly or indirectly, from engagement in the evaluation process, and learning to think evaluatively» (Patton 2008:108).

Prosessuell bruk er altså når de som er involvert i selve evalueringsprosessen lærer fra deltagelse eller gjør endringer i adferd eller et tiltak basert på evalueringsprosessen, snarere enn at selve funnene fra evalueringen benyttes. En implikasjon ved denne brukstypen er at det i et utfører-bestiller forhold, bør være en tett interaksjon mellom aktørene i gjennomføringen av evaluering. Bestiller skal ikke være en passiv mottaker av evalueringsfunn. Her er altså prosessen et mål i seg selv, så hvordan brukes prosessen?

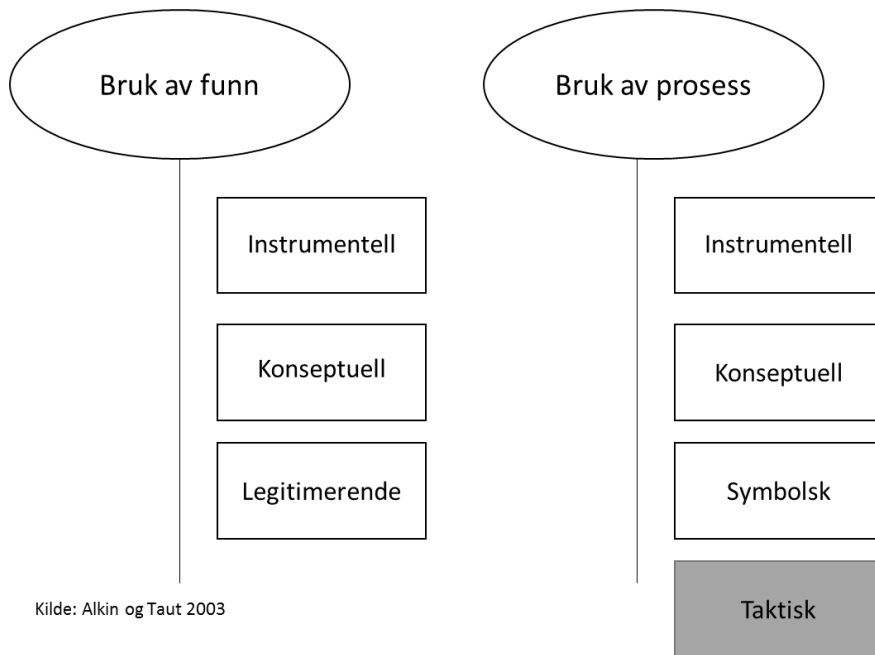
Et viktig argument hos Patton er at evaluering utgjør en kultur, en måte å tenke på. Prosessuell bruk er forskjellig fra bruk av substansielle funn på samme måte som det er forskjell mellom å lære, og å lære hvordan man lærer. Gjennom prosessuell bruk av evaluering kan man skape lærende organisasjoner (Patton 1988). Med andre ord dreier det seg ikke om bruk av funn fra evalueringen på samme måte som i instrumentell, konseptuell, taktisk og legitimerende bruk. Man benytter heller prosessen til å lære læring. Organisasjoners læringskultur har for øvrig vært en viktig tematikk innenfor organisasjonsteorien (se for eksempel Cohen og Levinthal 1990). Både i den prosessuelle og konseptuelle kategorien må læringsperspektivet sies å være fremtredende.

Pattons standpunkt om at utførere av evalueringer aktivt bør engasjere seg i å levere informasjon som møter oppdragsgivers forventninger, og at man gjennom aktiv deltakelse i større grad kan hente ut gevinster fra en evaluering (Patton 1988), har fått stor betydning for praksisfeltet. Patton utga også i 1986 første versjon av boken «Utilization focused Evaluation» - Pattons standpunkt må sies å være tett knyttet til det instrumentelle perspektivet, samtidig som han argumenterer for at dette muliggjøres gjennom prosessuell bruk, altså aktiv deltakelse og tett interaksjon mellom oppdragstaker- og giver.

2.3 BRUK AV FUNN VERSUS BRUK AV PROSESS

Typologiene av bruk som er presentert her, er fortsatt gjenstand for teoretisering og debatt, mest sannsynlig fordi det fortsatt viser seg å være utfordrende å vurdere hvilken påvirkning og effekter evaluering har (Højlund 2014:27). Alkin og Taut gjør en finere inndeling på hvordan disse typene brukes. Det gjøres ved å skille mellom bruk av funn og bruk av prosess som to separate måter å bruke en evaluering på (Alkin og Taut 2003). Bruk av funn kan være av typen instrumentell, konseptuell og legitimerende, mens bruk av prosess kan være av typen instrumentell, konseptuell og symbolsk, slik det fremkommer i figuren nedenfor. *Taktisk bruk* vil naturlig også kunne plasseres under fanen «bruk av prosess». Slik vi har sett ovenfor (Vedung 2009) er det nettopp skillet mellom bruk av funn og bruk av prosess som også utgjør

skillet mellom legitimerende og taktisk bruk. Højlund har for øvrig valgt samlebetegnelsen «justificatory uses» for å beskrive brukstyper slik som legitimerende, taktisk og symbolsk, fordi de alle har det til felles at det dreier seg om ekstern rettferdiggjøring av organisasjonen (Højlund 2014:27).



Denne enkle modellen er et godt bidrag til et mer presist rammeverk om evalueringsbruk. Når et evalueringsoppdrag gjennomføres av en ekstern oppdragsgiver presenteres arbeidet som regel i form av en evalueringsrapport. I disse rapportene fremkommer resultater og funn som analyseres og gjerne ender opp i noen helt konkrete anbefalinger. Spørsmålet er da hvordan slike funn brukes? Fører resultater og anbefalinger til at konkrete beslutninger fattes, slik som at et program omstruktureres eller legges ned? Bidrar funnene med ny og nyttig kunnskap som virksomheten lærer noe av? Kan man bruke resultatene til å vise til eierdepartementer og andre i omgivelsene at man er suksessfull og virker, at man når de resultatene og målene som er satt for virksomheten?

Vi har sett en antagelse om at prosessuell bruk og tett involvering gjør at organisasjonen lærer å tenke på en evaluerende måte. Hva vil det si? Er virksomheten i det hele tatt en aktiv deltaker i evalueringsprosessen, eller er det viktigere å holde avstand? Hvilke direkte handlinger eller beslutninger er et resultat av å delta i evalueringsprosessen? Finnes det støtte for at evaluering gjennomføres som et rituale, noe man bare gjør?

2.4 ET BREDT PERSPEKTIV PÅ EVALUERING

På et tidlig tidspunkt ble det uttrykt bekymring blant forskere om at evalueringer ofte ikke ble benyttet (Levinthorn og Hughes 1981). Utgangspunktet i Patton et. al. (1977) var at det i evalueringsslitteraturen eksisterte stor grad av konsensus rundt oppfatningen om at evalueringer sjelden hadde en særlig innvirkning på beslutningsprosesser i det offentlige. Kort sagt ble det ansett at evalueringer ikke ble benyttet eller at de ble benyttet i liten grad. Argumentet i Patton et. al. var at det skyldes et for snevert syn på hvordan evalueringer påvirker, det førte dermed til at bruk ble undervurdert (Patton et. al. 1977). Her blir det argumentert for at evaluering ofte blir brukt på en diffus og indirekte måte. Spesielt mente de at idealtypisk bruk, hvor avgjørelser og aktiviteter kunne spores direkte tilbake til en gjennomført evaluering (instrumentell), er sjelden. Evalueringer ble ikke bare brukt til: *“grand decisions with immediate concrete and visible impact, but in a more subtle, clarifying, reinforcing, and reorienting way»* (Patton et. al. 1977:161). I vurdering av evalueringsbruk burde man derfor fokusere mindre på synlige virkninger og konkrete avgjørelser i styringen, og heller fokusere på hvordan evalueringer bidrar med å fastsette retning og utvikling (Patton et. al. 1977). I dette arbeidet ble bestillere av flere evalueringer på helsefeltet intervjuet og de fant at offentlige beslutningsmyndigheter brukte evalueringene til å fylle kunnskapshull, forsterke egen oppfatning av situasjonen, redusere usikkerhet og fremskynde/sette opp farten på beslutninger.

Et lignende argument ble fremsatt i Cousins og Leithwoods (1986:332) som utvider evalueringsbegrepet - det er simpelthen bruk av informasjon som på enhver måte kan bli assosiert med resultatene fra en evaluering, for eksempel data, tolkninger og anbefalinger, som blir kommunisert under eller etter en evaluering. Ved å gå igjennom empiriske studier, identifiserte de; bruk som støtte for spesifikke beslutninger, bruk som utdanning av beslutningstakere og bruk som prosessering av- og tenking rundt evalueringen (Cousins og Leithwood 1986:345).

Weiss understreker også at systematiske undersøkelser og analyser ofte ikke benyttes på en direkte og instrumentell måte. Hun hevder at er en sjeldenhet at evalueringer leverer tydelige svar som aktører i det offentlige kan anvende for å løse problemer. Forskning påvirker snarere utvikling og modifisering av tiltak, gjerne på diffuse måter. En evaluering gir kunnskap som «sniker» seg inn i diskusjoner; *«its influence is exercised in more subtle ways than the word «utilization»* (Weiss 1980:381). I tillegg påpeker hun at beslutninger ikke fattes på en entydig måte. Spesielt i store organisasjoner kommer avgjørelser til uttrykk uten systematisk behandling, de formes heller gradvis uten at en formell styringsprosess med agenda-setting, deliberasjon,

avgjørelse og implementering, følges. Ansatte i en organisasjon baserer handlingene sine på summen av kunnskap og dømmekraft som de bærer med seg (Weiss 1980). Da vil evalueringen være et av mange verktøy for beslutningstakerne.

Man må anerkjenne at organisasjonsatferd ikke alltid følger en rasjonell styringsmodell, men heller består av byråkratiske prosedyrer, politiske drakamper og «garbage- can» prosesser (Weiss 1988), hvor organisasjoner oppfattes som anarkiske i betydningene av at avgjørelser kan komme i stand basert delvis på tilfeldigheter, ved at en strøm av deltakere, problemer, løsninger og valgmuligheter eksisterer i en gitt situasjon, og ikke i neste (Cohen og Olsen 1972).

Hvor mye en evaluering blir brukt må derfor ikke bare vurderes ut fra et instrumentelt perspektiv, uten at det betyr at man ikke oppfordrer til instrumentell bruk. Weiss påpeker også at evaluering utvilsomt har reelle konsekvenser; de kan utfordre gamle ideer; tilby nye perspektiver og gjøre om på agendaen. I tillegg viser det modernitet, kunnskap og kan rettferdiggjøre beslutninger (Weiss 1999:468). På bakgrunn av dette vektlegges det som ovenfor er kalt «enlightenment» eller konseptuell bruk.

Vi har sett ovenfor at Patton anerkjenner at evalueringsbruk ofte kommer til uttrykk på andre enn instrumentelle måter, men mener imidlertid at Weiss tegner et for dystert bilde av hvordan evalueringer benyttes (Patton 1988). Patton mener at Weiss gir inntrykk av at evalueringer som spesifikt bidrar til tiltaksforbedring er sjeldent og at det beste vi kan håpe på er indirekte påvirkning, med en viss avstand fra det som i utgangspunktet var argumentert for i evalueringen (Patton 1988). Patton mener tvert om at det finnes utallige eksempler på målrettet utnyttelse av evalueringer, og at dette heller er regelen enn unntaket. Perspektivet som Weiss tar, mener Patton er mer hensiktsmessig innenfor visse kontekster, slik som store nasjonale politiske diskusjoner og drakamper og ikke i programevalueringer, hvor et offentlig organ bestiller tjenester fra ulike tilbydere. Utførere har utviklet kompetanse på hvordan evalueringer best skal kunne tjene oppdragsgiver og et tett forhold mellom bestiller og utfører bidrar til at evalueringer benyttes (Patton 1988).

Som vi har sett ovenfor setter også Patton fokus på opplevelsen som deltakerne har av å delta i selve evalueringsprosessen, at endring vil inntreffe gjennom prosessen (Patton 1988). Altså at evalueringer utnyttes best gjennom aktiv deltakelse og tett interaksjon mellom partene. Pattons posisjon er nært knyttet til praksisfeltet og hans bakgrunn som evaluator, evaluatorer inntar en sentral rolle og tankegangen er; «intended use for intended users» (Patton 2008:37). Dette

perspektivet har drevet moderne tilnærminger til evaluering både på praksisfeltet og i litteraturen, som en samarbeidsprosess (Herbert 2015).

Kirkhart (2000) argumenterer for at det er nødvendig med en mer helhetlig teoretisk tilnærming til hvordan evalueringer kan påvirke. For å være i stand til å fange opp alle mulige virkninger av evaluering er det derfor nødvendig å tilnærme seg evalueringsbruk i vid forstand. Når man undersøker evalueringsbruk er målet også å avdekke alle endringer som en evaluering har ført til, både inkrementelle, episodiske, ikke-instrumentelle konsekvenser, sammen med de mer direkte og instrumentelle virkningene (Kirkhart 2000:7). Kirkhart har forøvrig fått bred støtte for sitt synspunkt, og det er i forlengelsen utviklet mer omfattende rammeverk for å undersøke bruk, som ikke bare tar for seg endring på organisasjonsnivå, men som også inkluderer kognitive og atferdsmessige endringer på individ- og interpersonlig nivå (Herbert 2014; Mark og Henry 2003).

Poenget med denne diskusjonen er å formidle at bruksbegrepet har endret seg over tid og at det nå i evalueringslitteraturen eksisterer en bred tilnærming til hvordan man skal forstå og analysere evalueringsbruk, samtidig som man anerkjenner at tidlige perspektiver fremdeles er relevant og viktige. Den smale forståelsen er ikke forkastet, men den har blitt supplert. Denne oppgaven er preget av å være en eksplorerende undersøkelse. Det er da viktig å ta utgangspunkt i et bredt perspektiv på hvilke virkninger evaluering kan ha i en organisasjon, slik at så mange virkninger som mulig kan avdekkes.

3. METODE

Metodelæren gir råd om hvordan man skal gå frem for å samle inn, behandle og analysere data. Det bygger på sentrale vitenskapelige normer som vil gjelde uavhengig av hvilke metodevalg som er gjort. Vektlegging av vitenskapelige prinsipper er det som skiller forskning fra andre fremgangsmåter for å besvare spørsmål om faktiske forhold (Hellevik 2009). En bevisst og kritisk refleksjon omkring undersøkelsesopplegg og metodebruk er derfor helt vesentlig i ethvert vitenskapelig arbeid. En forutsetning for en god diskusjon om metode, er en klar forståelse og forklaring på styrker og svakheter ved metodevalg som er gjort (Georg og Bennet 2005).

King Keohane og Verba (1994) definerer forskning i samfunnsvitenskap ut i fra fire karakteristiske trekk. For det første er målet å kunne trekke slutninger. Det handler altså om et forsøk på å konkludere utover dataen man har, til noe bredere som ikke er direkte observert. For det andre skal prosedyrene være offentlige slik at resultater kan testes og ettergås av andre, det handler altså om reliabilitet. For det tredje må man anerkjenne at konklusjonene er usikre. Et vitenskapelig arbeid krever derfor at man gjør refleksjoner rundt usikkerheten (King et. al. 1994) Det vil også gjelde i kvalitative studier hvor man ikke kan rapportere om usikkerhet i form av tall og prosentverdier. For det fjerde; innholdet er metoden. Det vil si at validiteten til et vitenskapelig arbeid er priggitt den metodiske kvaliteten (Ibid).

I det følgende kapittelet vil jeg gjennomgå valg av metode, forskningsdesign og datainnsamling. Det vil også mot slutten av kapittelet bli redegjort for validitet og reliabilitet. Oppgaven er et casestudie av Innovasjon Norge, hvor det empiriske materialet er samlet inn gjennom intervjuer med nøkkelinformanter i organisasjonen. Det er samtidig benyttet dokumentstudier som støttende analyse. Datainnsamlingen baserer seg dermed på metodetriangulering med vekt på informantintervjuet. Triangulering innebærer at det er tatt i bruk mer enn én metode eller datakilder (Bryman 2001:274). Studiet er eksplorerende i betydningen av at det ikke er utledet hypoteser som skal testes, men at temaet kartlegges deskriptivt ut fra teoretiske forventninger.

3.1 FORSKNINGSDESIGN- OM CASESTUDIE

Mye av det vi vet om den empiriske verden er utarbeidet gjennom case-studier, og metoden utgjør fremdeles en stor andel av arbeidet som gjøres innen statsvitenskap (Gerring 2004). Case-

studie er et undersøkelsesopplegg som går ut på intensiv undersøkelse av mye informasjon om få enheter (Thagaard 2013:56).

Å vise til at et arbeid er en case-studie kan innebære flere ulike metodiske tilnærminger. En vanlig tilnærming er at det innebærer forskning på et enkelt fenomen, et tilfelle eller et eksempel (Gerring 2004). I denne sammenhengen behandles IN som et case på fenomenet evalueringbruk. Et mål med denne studien er å koble beslutningstakere, og andre ansatte i INs egne beskrivelser av hvordan de bruker evalueringer med abstrakte teoretiske referanser fra litteraturen (Andersen 1997:11).

Interessen og behovet for case-studiet springer ut fra muligheten det gir for å forstå komplekse sosiale fenomener. Case muliggjør et virkelighetsnært perspektiv, slik som å forstå adferd i styringsprosesser og endrings- og organisasjonsprosesser (Yin 2014). Case er nær det virkelige liv, og tester dermed et fenomen direkte i sammenheng med hvordan det utfolder seg i praksis (Flyvbjerg 2006:235). Det finnes ikke en objektiv oppskrift på når case-studie bør benyttes, men det avhenger hovedsakelig av hva det er du ønsker å besvare. Jo mer spørsmålet søker å forklare hvordan og hvorfor enkelte sosiale fenomener fungerer, desto mer relevant vil et case-studie være (Yin 2014).

Det finnes ingen klar fellesforståelse av hva case- metode referer til, og det er omstridt hva case-studier er eller bør være innen samfunnsvitenskapen. Enkelte mener at case-studier er en støttedisiplin for «virkelig» vitenskap, andre mener at det er en alternativ tilnærming til konvensjonell vitenskap, med en kritisk eller anti-positivistisk tilnærming som gjør det mulig å gripe det «særegne». Disse to perspektivene har til felles et syn på at case-studier kun er begrenset egnet til å generalisere (Andersen 1997:10). Single case-studier har vært kritisert for å ha en begrenset nytte for teoriutvikling, og for at kausalkjeder ikke kan avdekkes. Samtidig som det ikke muliggjør ekskludering av alternative teorier (Georg og Bennet 2005:221). Atter andre ser case-studie metoden som en selvstendig vitenskapelig tilnærming på lik linje med andre metoder, med sikte på å generalisere (Anderen 1997:10). Blant disse er Yin som fremhever at det ikke er statistisk generalisering som er målet, men heller analytisk generalisering man tar sikte på (Yin 2014).

Et case vil alltid inngå i et univers av mulige caser, hvor et enkelt case er et eksempel trukket ut fra en større klasse av tilfeller hvor fenomenet som skal studeres utspiller seg (Georg og Bennet 2005:69). Universet i dette tilfellet kan ses som alle statlige virksomheter som er underlagt § 16 i økonomiregelverket eller alle statlige virksomheter som man vet gjennomfører

evalueringer. Mulig utvelgelse av caser er i prinsippet derfor stort. Med den bakgrunnen kan IN antas å representere en større klasse av mulige tilfeller.

En annen grunn for at IN er valgt ut, er den teoretiske forventningen om at størrelse på organisasjonen og evalueringskapasitet kan ha betydning for om evalueringer blir brukt (Halvorsen og Madsen 2013). Jeg har ikke utarbeidet noen oversikt over slike bakgrunnsvariabler i statlige virksomheter, men IN er en stor organisasjon, hvor det blir gjennomført relativt mange evalueringer og hvor hovedansvaret er organisert i en spesifikk avdeling. Det samme er tilfellet i organisasjoner som Norad (Norad 2016) og Utdanningsdirektoratet (Udir. 2016). Det er politikkfelt som man har kjennskap til at evaluering har vært mye brukt som styringsinstrument (Askim et. al. 2013:80). Dette er organisasjoner som også kunne blitt undersøkt. Pragmatiske hensyn spiller imidlertid også en rolle når man velger et forskningsopplegg (Andersen 1997:8). Denne oppgaven består av et single-case studie. Noen ganger velger man case-studie fordi det kun finnes et eller noen få tilfeller, andre ganger fordi det kun er et eller noen få caser som er tilgjengelige for forskeren. Det siste er antageligvis en vanligere begrunnelse (Andersen 1997:8). Tid, ressurser og tilgjengelighet har dermed også spilt en rolle for utvelgelse av kun et case.

Man kan hevde at caset som er valgt ut kan være representativt for en større klasse av caser, men man bør være forsiktig med å generalisere funn. Et argument for valg av single-case studie kan være at det er et eksempel på et typisk tilfelle (Yin 2014). Basert på dataene kan man ikke trekke slutning om hvorvidt IN er et typisk tilfelle, eller om det er en virksomhet i særstilling på evalueringsbruk. For å kunne trekke en slik slutning er det nødvendig med mer systematisk kunnskap fra andre deler av statsforvaltningen.

Et annet seleksjonskriterie kan være at det er et kritisk case - i betydningen av at det er kritisk for å teste teoretiske antagelser. I en slik sammenheng kan caset bidra til teoribygging, enten ved å bekrefte, utfordre eller utvide teorien (Yin 2014:51). Jeg mener at dette studiet kan bidra med ny kunnskap på feltet, og dermed teste teoretiske antagelser. Som nevnt finnes det lite empirisk materiale om evalueringsbruk i Norge. I en slik sammenheng kan det være nyttig å trekke frem virkelige eksempler hvor fenomenet utspiller seg. Da er et single-case studie en velegnet metode, fordi det kan bidra med viktig læring og kunnskapsakkumulering, uavhengig av i hvor stor grad man er i stand til å gjøre *formelle* generaliseringer (Flyvbjerg 2006).

Dette studie kan videre beskrives som teoretisk fortolkende, noe som innebærer en interesse for det spesifikke caset som undersøkes, samtidig som det ses på som et tilfelle av en større klasse

av fenomener. Studiet tar sikte på å bruke begreper og teoretiske sammenhenger til å oppsummere og strukturere empirisk materiale (Andersen 1997:68-69). Et annet begrep som kan brukes er instrumentelt case-studie, hvor et fenomen utforskes ved hjelp av en enkelt enhet, hvor den primære interessen er å studere et fenomen, og enheten representerer det empiriske grunnlaget for å utvikle forståelse av fenomenet (Thagaard 2013:56). Det teoretiske rammeverket som er presentert ovenfor har gitt forventninger til hvilke brukstyper som kan finnes i det empiriske materiale. Tanken er at dette skal bidra med kunnskap og innsikt om hvilke prosesser og prosedyrer som er etablert på området og hvordan evalueringer brukes og bidrar til styring og utvikling.

Som nevnt tidligere er det lite empirisk materiale i litteraturen og det er i liten grad utviklet modeller for å teste evalueringsbruk (Halvorsen og Madsen 2013). I tilfeller hvor det eksisterer lite systematisk kunnskap og lite testbare modeller på et tematisk felt passer det best med et eksplorerende undersøkelsesopplegg hvor formålet er å komme frem til mer presis kunnskap og presise problemstillinger (Hellevik 2009). I et eksplorerende undersøkelsesopplegg vil aktuelle informanter ofte være personer med spesiell kjennskap til fenomenet som undersøkes, altså eksperter på området (Hellevik 2009). IN var tidlig i oppgaveprosessen vurdert som case, men det er begrenset hvor inngående kunnskap man har om en organisasjon før man har vært der. Forskningsopplegget har derfor vært fleksibelt, tenkt som en prosess hvor aktuelle informanter ble identifisert etter hvert som oppgaven skred frem.

Yin (2014) poengterer at case-studier bør rettes mot hendelser som er relativt nære i tid. Dette må sies å være spesielt sentralt når man studerer evalueringsbruk. Dersom det er lenge siden en evaluering er gjennomført og brukt, kan man ikke forvente at respondenter har den samme detaljkunnskapen lenger. Det kan også antas at det er utskiftninger i organisasjonen som gjør at kunnskap om hvordan og hvorvidt «gamle» evalueringer har blitt tatt i bruk, er redusert. Dette er en av grunnene til at jeg ikke undersøkte hvordan spesifikke rapporter var blitt tatt i bruk, men det er rimelig å anta at informantene har mer kunnskap om, og derfor heller formidler, det som er nært i tid snarere enn det som ligger lenger tilbake i tid.

3.2 TRE PRINSIPPER FOR DATAINNSAMLING

Yin foreslår tre prinsipper for datainnsamling når man utfører case-studie; bruk av flere kilder, opprette en case-studie database, og opprettholde en tydelig beviskjede. Dersom disse følges vil det bidra til å styrke reliabilitet og validitet.

Bruk av flere kilder bidrar til å styrke datamaterialet. Det gir mulighet til å oppdage et bredere sett av historiske og atferdsmessige variabler. Bruk av flere kilder i en undersøkelse betegnes ofte som triangulering. Dersom man for eksempel gjennom skriftlig materiale kan avdekke hendelser, adferd og argumentasjon som sammenfaller med funn fra intervjuer, kan det gjøre funn og konklusjoner mer overbevisende og presise (Yin 2014:120). Det kan også tenkes at det kan gjøres funn som er motstridende med informasjonen som er tilegnet gjennom interjuene. I denne oppgaven utgjør intervjuer med nøkkelaktører hovedkilden til datamaterialet, fordi det er her kunnskapen om evalueringsbruk i IN finnes. Det er i tillegg supplert med enkelte sentrale dokumenter, slik som IN årsrapport 2015, Årlig rapport til oppdragsgivere 2015, og evalueringsrapporter.

Å *opprette en case-studie database* har å gjøre med hvordan man organiserer og dokumentere de dataene som er samlet inn, slik at funn er mulig å etterprøve. Det bidrar spesielt til å styrke reliabilitet (Yin 2014:124) Databasen i denne oppgaven består av en mengde skriftlig materiale, som er fullt mulig å etterprøve. Alle lydfile og de transkriberte dokumentene er tatt vare på. Det må imidlertid hentes inn tillatelse fra informanter dersom det skal gis tilgang til disse dokumentene. Og disse vil etter hvert også bli slettet. Fra dokumentstudiene er årsrapporter tilgjengelig på INs egne nettsider. E-poster og dokumenter jeg har fått tilsendt av IN, er også tatt vare på. I tillegg har jeg i forbindelse med oppgaven hatt et møte med en representant fra DFØ, som blant annet arbeider med evalueringsfaglige spørsmålsstillinger.

Å *opprettholde en tydelig beviskjede* innebærer å tydeliggjøre stegene i forskningsprosessen, slik at stegene fra funn til analyse og videre til konklusjon kan finnes igjen i datamateriale ved inspeksjon (Yin 2014:127). Det handler med andre ord om å styrke reliabilitet ved å sørge for gjennomsiktighet. Jeg har forsøkt å opprettholde en slik beviskjede ved å tydeliggjøre hva som er funn og hva som er egen analyse, refleksjon og kommentarer. I analysedelen forsøker jeg å gjøre det tydelig hvem som har sagt hva, slik at leseren kan følge hver enkelt gjennom analysen og slik også gjøre funnene sporbare, ved eventuelt innsyn. Dette prinsippet for datainnsamling relaterer seg direkte til reliabilitet som blir ytterligere redegjort for nedenfor.

3.3 INTERVJUET

Intervjuet er den dominerende metoden i kvalitativ forskning og kan utformes på varierende måte. Fra det ene ytterpunktet, karakterisert av å følge en strukturert mal hvor alle stilles samme spørsmål i samme rekkefølge, til en mer åpen og samtalebasert struktur. Den klart vanligste

formen er den semi-strukturerte tilnærmingen, hvor temaene som det skal spørres om er fastlagt på forhånd, men hvor det vektlegges fleksibilitet i intervjusituasjonen (Thagaard 2013:99-98).

Jeg har benyttet en semi-strukturert tilnærming, hvor jeg på forhånd hadde utarbeidet en intervjuguide, med ni hovedspørsmål som var ment å styre temaet. Under hvert hovedspørsmål var det videre utformet en rekke underspørsmål. Det var ikke intensjonen at hvert underspørsmål skulle stilles direkte til respondenten, men heller bidra til at samtalen holdt seg innenfor de rammene jeg ønsket. I intervjuene som jeg gjennomførte, forholdt jeg meg relativt passiv til intervjuguiden. Mye av grunnen til det, var at respondentene i stor grad holdt seg innenfor de rammene og den tematikken som jeg hadde penslet ut på forhånd, og at det derfor ikke var behov for å stille alle spørsmålene direkte. En annen viktig grunn er at en slik tilnærming vektlegger opplevelsen til respondenten, noe som bidrar til å forstå de hendelsene og prosessene som utspiller seg rundt evalueringsbruk, fra organisasjonens eget perspektiv (Andersen 2006:282). Jeg ønsket med andre ord at INs egen fortelling om evalueringsbruk skulle avdekkes. Selv om jeg ikke forholdt meg strengt til intervjuguiden, ble den modifisert underveis og tilpasset noe til den enkelte respondent.

Fleksibilitet i intervjusituasjonen er viktig også av andre grunner. Det sørger for en mer sammenhengene flyt, hvor oppfølgingsspørsmål tilpasses samtalen og fungerer som respons på det som har blitt sagt. Dette gir respondenten kontroll over hva som formidles. Oppfølgingsspørsmålene vil da sørge for mer detaljert informasjon og mer nyanserte kommentarer til de temaer, begreper og begivenheter som respondenten selv beskriver (Thagaard 2013: 101). Samtidig som man ved bruk av åpne spørsmål selv kan ta valg om hvilke tilleggsspørsmål (også kalt probing) man velger å stille, etter hvert som intervjuet skrider frem (Berry 2002:680)

Informasjon fra et intervju skapes og konstrueres som et resultat av en sosial prosess (Andersen 2006:281). Åpne og fleksible intervjuer kan gi verdifull innsikt og informasjon, for eksempel ved å få utforsket uventede svar, samtidig som det kan være forbundet med reliabilitet- og validitetsutfordringer (Berry 2002:680). Men løsningen på slike utfordringer består ikke i å minimalisere effekten av den sosiale prosessen, ved kun å stille standardiserte spørsmål. Tanken er heller at man forsøker å innlemme en slik forståelse i datakonstruksjon og analyse (Andersen 2006:281).

I et eksplorerende undersøkelsesopplegg vil aktuelle informanter ofte være personer med spesiell kjennskap til fenomenet som undersøkes (Hellevik 2009). I dette studie er alle

nøkkelinformanter som er valgt ut fordi de «besitter kunnskap om saker, situasjoner, relasjoner og kontekst som ikke er allment tilgjengelig» (Andersen 2006:281). Nøkkelinformantene i denne studiene antas å ha særlig kunnskap og erfaring om evalueringsbruk i IN. Slike intervjuer har også blitt kalt «elite intervjuer», hvor informantene gjerne er offentlige tjenestemenn på høyt nivå (Aberbach og Rockman 2002; Berry 2002). En utfordring knyttet til å spørre eliteinformanter i en gitt virksomhet om egne vurderinger, er at de kan ha tydelige interesser i å formidle et bestemt budskap. Det sitter sannsynligvis langt inne å si at evalueringer er «hylleware», hvis det hadde vært tilfelle. Samtidig er det slik at det er de involverte som har nærhet og mulighet til å observere effekter. Men det er viktig å få frem at slik data kan være subjektive oppfatninger og mulig også strategisk besvart (Madsen, et. al. 2013:71). Sagt på en annen måte:

«Interviewers must always keep in mind that it is not the obligation of a subject to be objective and to tell us the truth. We have a purpose in requesting an interview but ignore the reality that subjects have a purpose in the interview too: they have something they want to say (...) They're talking about their work and, as such, justifying what they do. That's no small matter» (Berry 2002:680).

3.4 UTVALG

I et case-studie velges respondenter basert på kunnskapen de har. Det vil si at det baserer seg på et strategisk utvalg, og det innebærer at deltakere velges på bakgrunn av «egenskaper eller kvalifikasjoner som er strategiske med tanke på problemstillingen og undersøkelsens teoretiske perspektiv» (Thagaard 2013:60). Det er noe annet enn et sannsynlighetsutvalg som skal følge et prinsipp om representativitet, men heller egnet til å utforske problemstillingen. Utvalget og utvalgets størrelse må derfor avgjøres basert på de analytiske målene med studiet (Thagaard 2013: 65).

Dataene fra intervjuene baserer seg på et utvalg av fire respondenter fra IN. Tre av respondentene har tittelen spesialrådgiver, hvor to av disse jobber i avdelingen som har hovedansvar for bestilling og gjennomføring av evalueringer. Én av spesialrådgiverne er overordnet programansvarlig for et av INs programmer. En av respondentene er valgt fra den øverste ledergruppen og er direktør for et av INs virksomhetsområder. I utgangspunktet var det også avtalt et intervju med en sjette informant, som også var direktør på et av INs virksomhetsområder, men vedkommende ble hindret fra å stille og intervjuet derfor avlyst. Tid og ressurser har satt begrensinger på utvalgets størrelse. Jeg vil ikke si at jeg nådde et

«metningspunkt», altså et punkt hvor flere enheter ikke lenger gir ytterligere informasjon (Thagaard 2013:65).

Det hadde klart styrket studiets validitet dersom jeg også hadde kartlagt evalueringsbruk på andre av organisasjonens virksomhets- og programområder.

Alle de fire intervjuene ble gjennomført i perioden mellom 31. mars og 19. april 2016, fordelt på til sammen tre besøk på INs kontor. Intervjuene ble gjennomført på ulike møterom. Den andre gangen ble det gjennomført to intervjuer, med kort mellomrom. Det var satt av én time til hvert intervju og samtlige varte omkring 45 minutter.

Formålet med denne oppgaven er å kartlegge evalueringsbruk i en statlig virksomhet. Som jeg har vært inne på ovenfor, betyr det at jeg også kunne valgt et annet case, eller inkludert flere case i undersøkelsen. Det er likevel slik at når IN var valgt, gjaldt det å få tilgang til sentrale personer i organisasjonen, med kunnskap om det aktuelle fenomenet. Målet med den første henvendelsen var å skape kontakt med en såkalt «gatekeeper», en person som videre kan åpne for tilgang til andre aktuelle respondenter. Den første personen som ble identifisert var en sentral person, med lang erfaring og faglig tyngde på gjennomføring av evaluering i IN. Jeg har vært praktikant i en bedrift som har utført en rekke analyser på oppdrag fra IN og direktøren her rettet en formell forespørsel per e-post på mine vegne, hvor det videre ble avtalt at jeg kunne ta kontakt. Etter intervjuet med «gatekeeperen» ble det identifisert andre mulige respondenter, som på forhånd ble informert fra sin kollega om at jeg ville ta kontakt. Det ble da avtalt to nye intervjuer, mens en tredje ikke ville stille fordi vedkommende var relativt ny i jobben. Det ble opprettet en dialog med denne personen, som også videresendte mine e-poster til andre aktuelle respondenter. På bakgrunn av denne dialogen ble det siste intervjuet avtalt.

Det mest sentrale er at respondenten dekker de ulike nivåene i organisasjonen på en god måte. To respondenter arbeider i avdelingen hvor blant annet ansvar for bestilling og gjennomføring av evalueringer er plassert, altså en evalueringsfaglig avdeling. Én er programansvarlig på det som kan karakteriseres som mellomledernivå på et av INs sentrale programmer, mens én representerer toppledelsen.

I tillegg til å representer et strategisk utvalg, er det også tale om et tilgjengelighetsutvalg. Det vil si at utvalget baserer seg på hvem som har både mulighet og vilje til å stille (Bryman 2001:324). Opplegget for undersøkelsen kan sies å være prosessebasert, hvor informanter ble identifisert fortløpende, og intervjuet ble tilpasset til den enkelte etter hvert som prosjektet skred frem og min kjennskap til INs evalueringsbruk utviklet seg. Et slikt tilgjengelighets- og

prosessbasert metodisk opplegg betegnes gjerne som snøballmetoden (Bryman 2001; Thagaard 2013).

Det var visse logistiske utfordringer med å avtale intervjuer. Å avtale møter med respondenter som jobber langt fra hvor jeg selv bor, og som også har en travel arbeidshverdag, er en tidkrevende prosess. Noe som bidro til totalt tre besøk til INs kontorer. Jeg opplevde imidlertid at de ansatte var interessert i å bidra. Størst utfordring var det å få representanter fra toppledelsen, blant annet fordi det ikke ble avtalt direkte med dem, men gjennom en tredjeperson.

Jeg informerte informantene om oppgavens tema, og hvordan IN skulle brukes som case på evalueringsbruk. Jeg spurte om det var greit at det ble tatt i bruk båndopptager, hvilket samtlige respondenter gikk med på. To av respondentene ønsket sitatsjekk, i begge tilfeller ble det begrunnet med at de tidligere hadde opplevd at sitater var tatt ut av sammenheng og dermed brukt på en feilaktig måte.

Det ble understreket at svarene som ble gitt utelukkende skulle brukes til masteroppgaven. Det ble samtidig nevnt at svarene og sitater ikke skulle knyttes til navn. Selv om informantene er behandlet uten navn, har det vært viktig å kunne knytte uttalelsene til stillingen som informantene har. Nøyaktig informasjon om hvilket fagområdet stillingen dekker er imidlertid utelatt, da det ikke anses som avgjørende. Det må også nevnes at flere av informantene kjenner til hvem i organisasjonen jeg har snakket med. Jeg oppfattet overhodet ikke at INs ansatte var opptatt av anonymisering. Intervjuene var etter min mening heller preget av tillit og åpenhet fra INs side.

3.5 DOKUMENTANALYSE

Dokumentanalyse er en metode hvor man gir visse skrevne tekster status som kilde eller data for selve undersøkelsen og analysen, på samme måte som man bruker data fra for eksempel intervju. Det foreligger ofte svært store mengder tekst som kan være aktuelt å inkludere i en dokumentanalyse. En praktisk tilnærming til dokumentstudier kan derfor være å starte med det som ligger nærmest i tid (Repstad 1998:86). Dokumentene som er benyttet i dette studie er førstehåndskilder, altså INs egne dokumenter.

Det sentrale målet med dokumentanalyse har vært å undersøke om det foreligger skriftlig materiale som enten viser til resultater fra en evaluering, eller på annen måte beskriver hvordan

en evaluering har blitt tatt i bruk i IN. Spørsmål jeg har vært interessert i å belyse i dokumentanalysen er:

Hvorvidt det finnes tilfeller hvor det står at tiltak er iverksatt eller noe er endret på bakgrunn av informasjon fra en evaluering? Dersom man finner eksempler på dette vil det kunne tale for at en evaluering har blitt brukt på en instrumentell eller konseptuell måte.

Hvorvidt det vises til en evaluering på en positiv måte eller om det på annen måte er vist til evaluering for å vise at man har fått til noe/hatt suksess? Dette spørsmålet er undersøkt for å se om det finnes materiale som kan tyde på at evalueringer har blitt brukt på en legitimerende måte.

Når man gjennomfører dokumentanalyser er det viktig å gjøre visse kritiske vurderinger. Det absolutt viktigste i denne sammenhengen er å ta forbehold om at dette er informasjon som IN ønsker å nå ut med. Dokumentene sier lite om hva det faktisk har blitt rapportert om til eierdepartement. Det kan være at det gjennom møter med eier har kommet frem annen informasjon.

Dokumentanalysen må derfor sies å være best egnet for å dokumentere det andre spørsmålet.

Det er ikke gjennomført en systematisk gjennomgang av dokumenter, hvor funnene skal presenteres som en egen analyse. Studier av dokumenter i denne oppgaven har blitt brukt som støttende analyse og som generell kunnskapskilde, hvor det i enkelte tilfeller er referert til dokumenter i analysekapittelet.

3.6 RELIABILITET OG VALIDITET.

Innledningsvis i dette kapittelet ble det gjort klart at en kritisk refleksjon omkring ethvert vitenskapelig arbeid er vesentlig for å kunne vurdere styrker og svakheter ved et studie. Helt sentrale begreper for å vurdere kvalitet i forskning er reliabilitet og validitet (Bryman 2001:270). Redegjørelsen som følger nedenfor er derfor tett knyttet til de vurderingene og diskusjonene som alt er presentert i metodekapittelet. Kort sagt handler reliabilitet om forskningens pålitelighet, mens validitet handler om i hvilken grad resultatene eller grunnlaget for tolkning er gyldig (Thagaard 2013).

3.6.1 Reliabilitet

Et grunnleggende forskningsprinsipp er at man følger metodiske regler for å oppnå mest mulig pålitelige resultater. Det vil si at en undersøkelse skal gjennomføres på en slik måte at andre forskere som anvender de samme metodene ideelt sett skal komme frem til identiske svar om

de forsøker å besvare de samme empiriske spørsmålene (Hellevik 2009:15). I utgangspunktet knyttes begrepet altså til repliserbarhet, men det kan imidlertid diskuteres hvor relevant dette er i kvalitative metoder. Svaret på spørsmålet vil i stor grad avhenge av hvilken vitenskapsteoretisk tilnærming man har, hvor repliserbarhet er mest relevant i en positivistisk tilnærming. Innenfor en sosialkonstruktivistisk logikk snakker man heller om at forskeren må argumentere for reliabilitet ved å redegjøre for hvordan dataen har blitt til i løpet av forskningsprosessen. Argumentasjon skal overbevise om hvor god dataen er (Thagaard 2013:202). I følge Tove Thagaard handler reliabilitet om «*en vurdering av forskerens pålitelighet*», hvor en kritisk vurdering av prosjektet gir inntrykk av om forskningen er gjennomført på en pålitelig og tillitsvekkende måte (Thagaard 2013:202). Hun gjør det samtidig klart at reliabilitet i en kvalitativ sammenheng må knyttes til at forskeren gjør rede for hvordan man har gått frem i prosjektet og at det må være helt tydelig hva som er referat fra et intervju og hva som er forskerens egen vurdering og analyser (Thagaard 2013:203).

Som et alternativ i kvalitativ forskning benytter Bryman seg av begrepet *pålitelighet*, hvor det argumenteres for at forskeren bør velge en revisjonistisk tilnærming, ved å sørge for at et register eksisterer som inneholder notater, lydfiler, transkriberinger etc, som er tilgjengelig og kan ettergås. Videre sikres *bekreftbarhet* ved å tydeliggjøre at forskeren ikke har tillat personlige verdier og meninger påvirke å styre forskningen. Samtidig må man erkjenne at fullstendig objektivitet ikke er mulig (Bryman 2001:272-274).

Tiltak man kan gjøre for å styrke reliabilitet er å gjøre forskningsprosessen transparent, slik at prosessen kan vurderes trinn for trinn (Thagaard 2013:202). Dette har jeg forsøkt ved å gi detaljert beskrivelse av forskningsopplegg og analysemetode. I analysedelen vil jeg også tydeliggjøre hva som er egne vurderinger og kommentarer og hva som er referat fra dokumentstudier og intervjuene. Det har også vært helt sentralt å benytte seg av båndopptaker og transkribering. Dersom man benytter seg av notater vil mye av dataen gå tapt fordi man ikke rekker å skrive ned alt. I tillegg vil det innebære en fortolkning av forskeren allerede i intervjusituasjonen, ved at noe tas med, noe velges bort og det som støtter teorimateriale er kanskje lettest å huske?

Hverken informant eller forsker vil opptre identisk fra et intervju til et annet. Åpne spørsmål kan også bidra til at informasjonen man får kan variere fra gang til gang. Det samme gjelder kontekst og relasjon til intervjuobjektet (Thagaard 2013).

3.6.2 Validitet

Thagaard viser til validitet som «vurdering av grunnlaget for fortolkninger» og handler om i hvilken grad resultater og tolkninger for det utvalget og det fenomenet som er undersøkt er gyldig (2013:204). Litt enkelt sagt snakker vi altså om at vi måler det vi faktisk vil måle (King, Keohane og Verba 1994:25). Det er vanlig å gjøre et skille mellom begrepsvaliditet, ekstern validitet og intern validitet.

Begrepsvaliditet handler om hvorvidt man har identifisert og målt de variablene som best representerer de teoretiske konseptene. Case-studier er forbundet med høy grad av begrepsvaliditet (Georg og Bennet 2005:19). Som vi har sett ovenfor hadde jeg utarbeidet en intervjuguide med en rekke spørsmål (operasjonaliserte variabler), men jeg ønsket samtidig åpne intervjuer hvor samtalen ikke ble styrt i detalj. Tanken var at dersom evalueringer blir brukt på den måten som er begrepsatt og skissert i litteraturen, ville svarene til respondenten reflektere dette begrepsapparatet. Når for eksempel respondentene snakker om legitimerende bruk, uten at jeg selv har benyttet det begrepet i spørsmålstilling kan det bidra til å styrke validitet. Samtidig som åpne spørsmål kan være en utfordring for reliabilitet (Berry 2002), da det kan føre til at andre ikke kommer frem til samme svar.

Kausalitet handler om at visse betingelser antas å føre til et bestemt utfall. Intern *validitet* viser til gyldigheten av en slik slutning, eller om utfall kan antas å være påvirket av andre betingelser som det ikke er kontrollert for (Yin 2014:46). Et slikt begrep vil være aktuelt dersom man søker å etablere kausale-slutninger. Målet med dette studiet har ikke vært å etablere klare kausalslutninger, slik som at y fører til x. Caset er imidlertid valgt ut fra en teoretisk antagelse om at visse kontekstuelle faktorer, slik som «evalueringskapasitet» kan antas å påvirke evalueringsbruk, uten at dette har blitt undersøkt eksplisitt.

Ekstern validitet handler om i hvilken grad resultater og tolkninger kan overføres, og være gyldige i andre sammenhenger (Thagaard 2013:205) Hovedkomponenten i ekstern validitet blir dermed i hvilken grad man kan generalisere studiet til også å gjelde andre situasjoner. I kvalitative studier snakker vi om fortolkning som grunnlag for overførbarhet og ikke beskrivelser av mønstre i dataen (Thagaard 2013:210).

Dette er også essensen i Yins skille mellom analytisk/teoretisk og statistisk generalisering. Analytisk generalisering innebærer at man ikke ser på caset som et utvalg som representerer en større populasjon, men heller som en mulighet til å empirisk belyse teoretiske konsepter (Yin 2014:40). Målet kan likevel være å generaliserer til andre konkrete situasjoner, ikke bare bidra

med abstrakt teoriutvikling. «Lessons learned» fra et case kan potensielt gjelde en rekke situasjoner, enten det dreier seg om et eller flere case (Yin 2014:40-43). Yin mener at analytisk generalisering kan baseres enten på; a) bekrefte, modifisere, forkaste, eller på andre måter fremme teoretiske konsepter som har blitt presentert b) bidra til at nye konsepter kan oppstå (Yin 2014:41). Et viktig poeng da er at funn fra et case-studie vil være på et konseptuelt høyere nivå enn det spesifikke caset (Ibid). Målet med denne studien er ikke å produsere nye teoretiske konsepter eller forkaste gamle, men heller å bekrefte og fremme teorien som har blitt presentert ovenfor. Jeg har med andre ord fokusert på et teoretisk interessant case som gjenspeiler teorien ved å fylle konseptuelle kategorier (Andersen 1997:25).

Målet med dette studie er å kartlegge sentrale trekk ved evalueringsbruk ved hjelp av logiske resonnementer som kobles til teori. Underveis vil jeg forsøke å *argumentere for* at forståelsen utviklet gjennom case-studie av IN, også kan ha gyldighet for andre deler av statsforvaltningen (Thagaard 2013:211). Jeg hevder ikke at funnene her er direkte representativt for en større populasjon. Utfordringen er at betingelser og kontekst i IN kan skille seg fra andre virksomheter, og det er derfor viktig å ta innover seg at vitenskapelige konklusjoner alltid er usikre, men så lenge vi rapporterer om usikkerhet er slike studier nyttig (King et al 1994). I kvalitative studier må forskeren selv argumentere for at en tolkning også er relevant i en større sammenheng. Vesentlig i tolkningen er da å få frem under hvilke situasjoner og betingelser det gjelder (Thagaard:2011).

Styrking av validitet kan også knyttes til gjennomsiktighet. Det innebærer at man redegjør for hvordan analysen gir grunnlag for konklusjoner man har kommet frem til. Det er viktig å gå kritisk gjennom analyseprosessen. Dersom man klarer å argumentere for at alternative forklaringer er mindre relevante, og fokusere på under hvilke betingelser en tolkning gjelder og når det ikke gjelder, vil det bidra til å styrke validitet (Thagaard 2013:205). Overførbarhet i denne sammenheng vil dreie seg om å gi rike og detaljerte beskrivelser som gir andre mulighet til å vurdere hvor overførbart det er, mens troverdighet sikres ved at forskningen er utført etter regler om «good practice» (Bryman 2001:275). I oppgavens analysekapittel vil jeg etterstrebe å få frem slike hensyn.

Det er klart at dersom dette single-case studiet hadde blitt mer direkte sammenlignet med andre lignende studier, eller at det også var samlet inn empiri fra andre deler av forvaltningen, ville det kunne bidratt til å styrke oppgavens validitet. Jeg gjør likevel oppmerksom på at formålet med dette studiet var teoretisk fortolkning ved hjelp av en kartlegging av evalueringsbruk, og ikke primært å foreta generaliserte slutninger. Det at kunnskap ikke formelt kan generaliseres

betyr heller ikke at det ikke bidrar med kunnskapsakkumulering innenfor fagfeltet. Tvert imot kan kontekst-avhengig kunnskap bidra med viktig læring og ekspertise inn på et felt (Flyvbjerg 2006: 223-227). Enten man snakker om generalisering på bakgrunn av et single-case eller på bakgrunn av store utvalg, er generalisering dessuten, ifølge Flyvbjerg, betydelig overdrevet som den eneste viktige kilden til vitenskapelig fremgang. En disiplin uten en systematisk produksjon av eksempler, gjennom case-studier er ikke tilstrekkelig. Sagt på en annen måte; «*the force of example*» er undervurdert som bidrag til vitenskapelig utvikling (Flyvbjerg 2006:228-242).

3.7 ETISKE REFLEKSJONER

I all forskning er det viktig å sette etiske refleksjoner høyt på agendaen. Etikk vil alltid være forbundet med skjønnsvurderinger. Hvem du snakker med, og hvilke opplysninger som innhentes, har betydning for hvordan datamaterialet bør fremstilles og brukes. I denne oppgaven er det ikke innhentet sensitive opplysninger. Respondentene er ressurssterke personer som befinner seg på senior- og ledernivå i statsforvaltningen. Med den bakgrunnen opplever jeg at prosjektet ikke har stått ovenfor særlig etiske utfordringer.

I følge Ryen dreier de viktigste etiske spørsmålene i forskning seg blant annet om *informert samtykke og konfidensialitet* (Ryen 2011:418). Informert samtykke innebærer at respondentene informeres om studiets formål, om hva materialet skal brukes til og at de har muligheten til å trekke seg om de ønsker det (Ryen 2011:418). Gjennom e-postkontakt i forkant av møtene med tre av respondentene fremkom informasjon om prosjektet, i tillegg til at slike opplysninger ble gitt i forbindelse med intervjuet. Det var spesielt viktig å gi en tydelig forklaring til informanten jeg ikke hadde hatt direkte e-postkontakt med i forkant av intervjuet, om hva prosjektet dreide seg om. I forbindelse med den første kontakten jeg gjorde med organisasjonen, ble også prosjektskissen i sin helhet vedlagt e-posten. Jeg tar også utgangspunkt i at ressurssterke nøkkelinformanter forstår hva det innebærer å delta på et intervju, hvor det ble gjort klart at informasjonen skulle brukes i en masteroppgave. Dersom datamaterialet skal brukes i en annen sammenheng, må det i tilfellet bli avklart med samtlige informanter. Alle respondentene samtykket til at det ble brukt båndopptager, og de ble informert om at ingen andre ville få tilgang til hverken lydfiler eller transkriberinger. Ved eventuelt innsyn av tredjepart, må det avklares med den enkelte informant. To av informantene ønsket, og fikk også, sitatsjekk. Tilbakemeldingen besto av presiseringer, og ble i enkelte tilfeller implementert i teksten.

Konfidensialitet innebærer at forskere anonymiserer respondenter for å beskytte deres identitet. (Ryen 2011:419). Prosjektet består hverken av personlig eller kontroversiell informasjon, og

jeg opplever dermed at det ikke var store etiske utfordringer i dette prosjektet. Det ble heller ikke uttrykt noe behov for anonymisering fra informantenes side. Det ble likevel gjort klart at jeg ikke skulle benytte meg av navn, men at kollegaer som kjenner arbeidsoppgavene ville være i stand til å forstå hvem jeg har snakket med. Flere av respondentene var også klar over hvem jeg hadde vært i kontakt med i forkant, basert blant annet på at en ansatt hos IN, sendte informasjon om at jeg kom til å ta kontakt videre til sine kollegaer, slik jeg har beskrevet ovenfor. Mitt inntrykk var at respondentene var interessert i oppgavens tematikk og at de også var interessert i hvilke kollegaer som jeg alt hadde snakket med. Det ble da naturligvis oppgitt.

Når det gjelder en eventuell anonymisering av IN, gir ikke det noe særlig mening. Det er ikke mange offentlige virksomheter som har tilsvarende oppgaver, slik sett vil det være umulig å anonymisere og fremdeles diskutere innholdet i informasjonen som er samlet inn. INs virksomhet og de aktiviteter de engasjerer seg i er dessuten informasjon som er av offentlig interesse. Det ble oppgitt at IN ikke ville bli forsøkt anonymisert. Jeg opplevde for øvrig lite fokus fra respondentenes side omkring spørsmål om anonymisering.

I tilfeller hvor et studentprosjekt skal behandle personopplysninger må det meldes til Norsk senter for forskningsdata (NSD 2016). Den innsamlede dataen består ikke av personopplysninger slik jeg ser det, og prosjektet er derfor ikke meldt inn til NSD.

4. RESULTATER OG ANALYSE - INNOVASJON NORGE OG EVALUERING

I denne delen blir resultater fra intervjuene med INs ansatte presentert, samtidig som det fortløpende tolkes opp mot teorien som er gjennomgått ovenfor. For å gjøre analysen mer lettfattelig og forståelig og for å unngå for mye gjentakelser, ser jeg det som mest hensiktsmessig å presentere funn og analyse som en integrert del. Jeg understreker likevel at jeg har fokusert på å gjøre det tydelig hva som er funn og hva som er tolkning.

Før vi går videre kan det likevel være nyttig å gjenta en meget sentral del av oppgaven, nemlig problemstillingen. Målet med oppgaven har vært å undersøke følgende:

Hvordan tas evaluering i bruk, og hvordan bidrar evaluering til styring og utvikling i en norsk statlig virksomhet?

For å gi en god beskrivelse av hvordan evalueringer blir tatt i bruk i IN, har jeg tatt sikte på en kartlegging av det samlede evalueringssystemet, fordi det er helt nødvendig å se det i sammenheng med hvordan evalueringene brukes. Kapittelet er todelt, hvordan gjennomfører IN evalueringer, og hvordan bruker IN evalueringer? Det viktigste vil være å tolke funnene i rammen av brukstypologien som er presentert ovenfor. Det innebærer et forsøk på å avdekke ulike typer bruk, også innenfor to separate logikker, nemlig bruk av funn fra en evaluering, som kan være instrumentell, konseptuell, og legitimerende, og bruk av prosess som kan være instrumentell, konseptuell og symbolsk. Slik jeg har nevnt i teorikapittelet, er taktisk bruk utelatt fra undersøkelsen. Det bygger på en antagelse av at slik form for bruk vil være svært utfordrende å avdekke (Foss Hansen 2003).

Slik det fremgår i oppgavens metodedel, har jeg benyttet meg av semi-strukturerte intervjuer. Jeg forholdt meg relativt passiv til intervjuguiden fordi respondentene i stor grad snakket om evalueringsbruk på en slik måte at det dekket de teoretiske konseptene, og at det derfor ikke var behov for å stille alle spørsmålene direkte. I tillegg var det viktig å forstå evalueringsbruk fra organisasjonens eget perspektiv. Det er nødvendig for å kunne avdekke alle brukstypene som potensielt er i spill (Patton 1977:144). Funnene som presenteres nedenfor bygger på intervjuene med fire av INs ansatte og materialet blir underveis supplert med empiri fra dokumenter som er gjennomgått.

Jeg vil starte med å beskrive hvordan IN gjennomfører evalueringer, hva slags undersøkelser som gjennomføres og hvilke strategier, rutiner og prosesser som er etablert.

Deretter vil jeg bevege meg over til analysens andre del som diskuterer hvordan IN bruker evalueringene.

4.1 INSTITUSJONALISERING - HVORDAN GJENNOMFØRER INNOVASJON NORGE EVALUERING?

Formålet med dette avsnittet er å danne et overblikk over hvordan evalueringssystemet i IN ser ut. Her vil jeg frem til hvordan arbeidet er organisert, og hvilke rutiner og prosesser som eksisterer. Jeg gjør oppmerksom på at denne delen hovedsakelig bygger på informasjon fra intervjuene. Jeg kan ikke utelukke at det er elementer i evalueringsregime som informanten rett og slett ikke kom på å fortelle om. Derfor har jeg også i enkelte tilfeller supplert med skrevne kilder.

Jeg har tidligere i oppgaven vært inne på at evalueringsevirsomheten i IN er institusjonalisert. Det vil si at evalueringsevirsomheten er organisert i en egen avdeling. Jeg har intervjuet to personer som er ansatt i denne avdelingen (Innovasjonspolitisk avdeling). En av informantene forteller at avdelingen er løsrevet fra de andre tjenestene, dette er viktig for å opprettholde en nødvendig avstand og nøytralitet, slik at man unngår at de som er ansvarlig for evaluering har et eierskapsforhold til tjenestene som evalueres. Informanten fra ledergruppa forteller for øvrig at avdelingen også har ansvaret for det samlede Mål- og resultatstyringssystemet til virksomheten (MRS).

Evalueringsevirsomheten i IN har en lang tradisjon og går tilbake til tidlig 80-tallet. Respondentene peker på at det vanligvis gjennomføres åtte til ti evalueringer årlig, med en samlet kostnadsramme på mellom seks til ti millioner kroner. Årsverk som er satt av til formålet har vanligvis variert mellom to og tre, og disse er tilknyttet Innovasjonspolitisk avdeling som er tillagt overordnet ansvar for bestilling og kvalitetssikring av evalueringene. Virksomheten gjennomfører utelukkende eksterne evalueringer, altså ingen interne. Flere av informantene peker på at det er viktig med et eksternt blick på virksomheten, og at egnevalueringer ikke ville hatt noe særlig legitimitet.

På spørsmål om det er utarbeidet retningslinjer og strategidokumenter som gjelder evaluering i IN svarer en av respondentene i Innovasjonspolitisk avdeling at det tidlig etter at IN ble stiftet kom en detaljert prosessbeskrivelse. I tillegg er det utarbeidet en strategi som ble nedfelt «skikkelig» i 2011. Dokumentet respondenten refererer til har tittelen «*Evalueringsstrategi for Innovasjon Norge*», og er gjort tilgjengelig for meg fra respondenten. Dokumentet beskriver

evalueringssystemet i IN fra 2011-2015. I tillegg er de nå i ferd med å utarbeide en ny strategi som bygger på den gamle. Denne tar for seg 2016-17, og ser mer konkret på hva man ønsker å evaluere i perioden. At de er i ferd med å utarbeide ny strategi fremgår også i *Innovasjon Norges Årlig rapportering til oppdragsgivere* for 2015(139).

En av respondentene fra Innovasjonspolitisk avdeling oppsummerer på følgende måte:

«... så vi har vel rimelig orden på evalueringsarbeidet ja, og har det av systemer som man bør ha tror jeg. Altså den store tunge evalueringsprosess-biten.

Prosessbeskrivelsen er ned i minste detalj».

Fire kilder til kunnskap

På bakgrunn av dokumentet Evalueringsstrategi for Innovasjon Norge og intervjuene er det rimelig å hevde at det totale evalueringsregime IN består av fire hovedkomponenter, eller det respondenten fra ledergruppen kaller fire «kilder». Det er serviceundersøkelsen, kundeeffektundersøkelsen, det som omtales som «øvrige evalueringer» det vil si på tjeneste- og programnivå, og en relativt ny undersøkelse som gjennomføres av SSB.

Serviceundersøkelsen er ifølge informanten i ledergruppa et litt lettere informasjons-tilfang som undersøker mer kvalitative elementer på hvordan kundene opplever samhandlingen med IN. I følge strategidokumentet er ikke hovedfokus i disse undersøkelsene på effekter. I følge en av informanten i evalueringsavdelingen undersøker *Serviceundersøkelsen* om kundene opplever kundehåndteringen på en god måte, og om kundehåndteringen til IN «er i tråd med de verdiene som IN setter høyt». Disse undersøkelsen er ikke publisert i evalueringsportalen og regnes ikke isolert sett som en evaluering.

I evalueringsstrategien oppgis det at en viktig del av datamaterialet som dekker de fleste av INs finansielle tjenester er Kundeeffektundersøkelsene (KEU). Det er en årlig effektmåling av bedrifter som har mottatt støtte fra IN. Det består egentlig av to undersøkelser, en førundersøkelse av bedrifter som har mottatt en tjeneste eller deltatt i et program, for å avklare mottakernes forventninger til prosjektene. Og en etterundersøkelse som blir gjennomført etter at støtte er gitt, for å belyse resultatene i etterkant (Oxford Research 2010:4; Meld.St.22 2012:31). Respondenten i ledergruppen forteller om KEU:

«Kundeeffektundersøkelsen gjennomføres årlig, og undersøker bedriftene først et år etter støtte og tre år etter støtte for å se om vi har skapt effekter. Den er jo en spørreundersøkelse hvor de uavhengig går inn og spør og graver for å finne ut om vi

har skapt effekter. Og den tar en stor bredde av det vi gjør av tjenester (...) den tar for seg 25 tjenester».

Rapportene publiseres på evalueringsportalen, hvor det i skrivende stund ligger 11 rapporter. Disse undersøkelsene blir imidlertid heller ikke eksplisitt omtalt som evalueringer.

En annen informant i evalueringsavdelingen peker på at informasjonen man får fra disse inngår i MRS-systemet som resultatindikatorer, både for å påvise addisjonalitet og kompetanse. Det er videre bred støtte i datamaterialet for at informasjon fra KEU-ene er et vesentlig element i rapporteringen til departementene. En informant sier at de:

«... bruker det inn i virksomhetsstyringen, i tertial rapportene som vi avleverer til departementet henter vi tall fra den store evalueringen som kundeeffekt er».

SSBs undersøkelse er en tredje undersøkelse som blir gjennomført, den er som nevnt implementert relativt nylig. Respondenten fra ledergruppen sier at dette er kvantitative undersøkelser som skal si noe om hvordan kundene som IN har jobbet med, står seg sammenlignet med andre bedrifter som ikke har mottatt tjenester. Undersøkelsen gjennomføres årlig, og har pågått de siste tre årene. SSB tar for seg grupper av bedrifter som er INs kunder, og ser på hvilken omsetningsvekst de har hatt, produktivitet og lønnsomhet. I tillegg analyseres vekst i antall ansatte og verdiskaping. SSB sammenligner med bedrifter som ikke har mottatt tjenester. Dette gjør at man bedre kan se hva slags effekter bedriftene har fått ut av støtten. Videre sier respondenten fra ledergruppen at SSBs undersøkelse *«... forteller "hard facts". De går inn i regnskapene og ser på hva som faktisk har skjedd».*

Respondenten med ansvar for effektmålinger sier at denne undersøkelsen består av *«rene økonomiske design, basert på objektive data».* I følge DFØs *veileder for Mål og resultatstyring i staten*, er det gjennom økonomireglementet pålagt for alle offentlig virksomheter å ha et form for MRS-system, og at man er ansvarlig for å sikre et tilstrekkelig styringsgrunnlag. Det innebærer at man må monitorere og evaluere (DFØ 2010). INs eiere og oppdragsgivere har utviklet et felles MRS-system og effektindikatorer for hvert delmål, som SSB med enkelte modifikasjoner har undersøkt (SSB 2015:8). Undersøkelsen vil da med andre ord kunne benyttes direkte som rapportering på MRS-systemet.

Den fjerde informasjonskilden er det som i strategidokumentet omtales som *«øvrige evalueringer».* Materialet peker på at det da er snakk om evalueringer på det som kan kalles tjeneste eller programnivå. Eller som informanten fra ledergruppe omtaler som en *«separat*

evaluering av en individuell tjeneste» og at evalueringene undersøker i større grad *«det man trenger å gå i dybden på»*. Materiellet peker videre på at dette arbeidet er systematisert i faste frekvenser, og at hver enkelt tjeneste bør gjennomgå en evaluering hvert fjerde år. IN har et stort antall tjenester slik at det blir *selvsagt en del* evalueringer.

En av informantene er hovedansvarlig for INs klyngeprogram. Dette omfatter ulike nettverkstjenester slik som Arena-programmet, med 22 klyngeprosjekter og Norwegian Center of Expertise (NCE) som i dag har 14 prosjekter (IN 2016).

I ansvaret inngår også utvikling av det helhetlige systemet for evaluering, monitorering og rapportering. Respondenten opplyser om at det eksisterer et komplekst evalueringssystem på disse programmene, og gir en oppsummeringen av hvordan det fungerer. Her fremgår det at man har overbyggingen som er selve *«programmet»* i dette tilfellet de to klyngesatsingene. Under det igjen har man et antall Arena-prosjekter som finansiers og støttes av IN, og et antall NCE-prosjekter. Det evalueres både på programnivå og prosjektnivå. Respondenten opplyser om at det er satt opp i alle strategidokumenter at det skal være en ekstern utsjekk av relevans, resultater og effektivitet. På programnivået er dette systematisert i regelmessige hovedevalueringer hvert fjerde-femte år. På Arena-prosjektene, som har tidsramme på tre år, er det systematiserte slutt-evalueringer etter disse tre årene. Det opplyses at dette er *«helt sånn programmert»*. Respondenten forteller:

«Tidligere har det vært slik at disse sluttevalueringene bestilles fra klyngeprosjektene selv. Nå har vi kjørt dette... sentralisert det. Slik at i forrige runde bestilte vi fire sluttevalueringer av Arena- prosjekter og kjørte i et oppdrag».

NCE-programmet støttes innenfor en tiårsramme. Her evalueres det først etter tre år, så etter seks eller syv år. Respondent forteller at dette gjøres for å undersøke:

«... hva har skjedd i forrige periode? Er det grunnlag for å fornye kontrakten? Så da bruker vi evalueringer, rett og slett som direkte premiss for en beslutning om vi skal fornye avtalen. (...) sluttevalueringer av Arena blir jo mer for å si, sjekke ut, fungerte dette, hva lærte vi, hva slags resultater kan vi bokføre? Mens disse underveisevalueringene blir jo mer direkte premiss for en beslutning. Så dette er det store løpet på evaluering».

4.1.1 Hva er evaluering i Innovasjon Norge?

Oppgaven tok innledningsvis for seg at evaluering er et tvetydig begrep som er tillagt mange meninger, og at det ikke er etablert noen fellesforståelse for hva slags type undersøkelser som dekkes av begrepet. Vi har sett at finansdepartementets definisjon viser til evaluering som en *systematisk datainnsamling, analyse og vurdering av en aktivitet* (Finansdepartementet 2005:8). Det er en svært bred definisjon og vil dermed kunne omfatte en mengde ulike undersøkelser.

Virksomheten viser som nevnt til fire kunnskapskilder. Disse undersøkelsene, analysene og evalueringene utgjør til sammen det respondenten i ledergruppen kaller, «*summen av de kunnskapskildene vi har*». Og at disse utgjør «*kjernen i vårt MRS-system*».

Det er imidlertid kun en av disse som omtales eksplisitt som en evaluering. Det er det som i strategidokumentet omtales som «*øvrige evalueringer*». På bakgrunn av finansdepartementet brede definisjon og hvordan respondentene snakker om analysene mener jeg likevel at man må snakke om alle disse prosessene som en samlet del av et totalt evalueringssystemet i IN, og at det i denne sammenhengen gir lite mening å ikke inkludere de øvrige analysene i evalueringbegrepet. Det er for eksempel interessant at to av respondentene omtaler KEU-ene som en form for evaluering. Tre av de fire respondentene tar opp KEU, når de skal svare på et åpent spørsmål om hvordan de går frem når virksomheten skal evaluere. Den eneste respondenten som ikke gjør det er programansvarlig. Det kan tyde at det nettopp for denne personen er de «*øvrige evalueringene*» som er viktigst. Respondenten fra ledergruppa, som kan sies å ha et mer overordnet blikk på systemet, snakker heller om «*summen av de kunnskapskildene man har*» og at det «*er noen grenseganger på hva som kalles evaluering og hva som kalles undersøkelse*».

En av informantene har ansvaret for effektmålinger og snakker om evalueringbegrepet:

«Evaluering er et begrep som kan brukes veldig vidt. Det brukes jo om internrevisjon og det kan brukes på alle mulige områder og på alle mulige nivåer. Vi har prøvd å forbeholde bruken av ordet evaluering til oppdrag knyttet til programevaluering vi da. Sånn at hvert fall vi skal skjønne hva vi snakker om».

Den andre informanten i Innovasjonspolitisk avdeling peker på at KEU er «*jo en evaluering det også*», og sier senere «*den store evalueringen som KEU er*».

En respondent viser forøvrig til at evalueringene er *«et supplement inn sammen med andre analyser, for å drive utvikling og styring av virksomheten»*. En annen sier at det er *«gode støttende analyser»*. Begrepet støttende analyser fremkommer på samme måte i INs Årlig rapportering til oppdragsgivere (2015:8).

I analysens andre del vil bruken av evaluering diskuteres. Respondenten er stilt åpne spørsmål på hvordan evaluering brukes. Svarene de gir må da antas og ikke kun gjelde det som i organisasjonene omtales som evaluering, men heller hvordan alle disse kunnskapskildene benyttes i virksomheten. Inntrykket er for øvrig at det er programansvarlig som virker å snakke mest om de «rene» evalueringene, mens topplederen og de to fra Innovasjonspolitisk avdeling, er like opptatt av SSB-undersøkelsen, KEU-ene og evalueringer.

4.1.2 Evalueringsprosessen

Hvordan går så IN frem når noe skal evalueres? Innovasjonspolitisk avdeling har som nevnt ansvaret for at evalueringer bestilles og gjennomføres. Når et oppdrag lyses ut gjøres det normalt gjennom et konkurransegrunnlag. En av informantene i Innovasjonspolitisk avdeling forteller at å utforme et godt konkurransegrunnlag er en tidkrevende prosess og at det er utfordrende å utforme et dokument man er helt fornøyd med, hvor man har fått med seg alt det man ønsker. Informanten mener derfor at det kan være mer hensiktsmessig å utforme dokumentet litt rundere. Her har praksisen variert noe over tid, hvor det tidligere ble utformet mer spesifikke og detaljerte oppdragsbeskrivelser.

Hvordan tematikken velges ut kan ifølge en av respondentene i Innovasjonspolitisk avdeling ha litt ulike innfallsvinkler. Det ene er at det kan avdekkes ting i KEU-ene, som man ser at det er ønske om å se nærmere på i en evaluering. En annen innfallsvinkel er at det ofte eksisterer krav fra departementet i tildelingsbrevene for en aktivitet *«om at det året vil man ha en evaluering eller etter fire år så vil man ha en evaluering, og at man jevnlig har evalueringer. Formuleringen kan være litt forskjellig»*.

Respondenten i ledergruppen formulerer det slik:

«Men jeg ser jo også at oppdragsgiverne våre vil være opptatt av å få mer hyppig informasjon om de «store» pengene enn om de «små». Sånn som at når de satser veldig på miljøteknologi og bruker masse penger, eller når de øker etablerertilskuddet og gründersatsingen betydelig, så ønsker de hyppigere tilbakemeldinger på det. Selv om

det kanskje ikke er sånn at man syklus-messig skulle sagt at nå kunne vi ha ventet en evaluering, så vil de likevel ha en del informasjon på det».

Samtidig trekkes det frem av samme person:

«I hovedsak så er det IN som selv bestemmer hva vi skal evaluere, og når. Så kan det komme synspunkter og behov hos eierne på det».

Når det gjelder den interne prosessen eller fremgangsmåten, forteller respondenten med ansvar for effektmålinger at man har lagt en evalueringstrategi som man prøver å følge. I tillegg er det slik at de ulike fagdivisjonene også etterspør evalueringer på områder der de har behov for en kunnskapsoppdatering dersom det er en god stund siden forrige evaluering. Eller i tilfeller hvor de er usikre på om et program- eller en arbeidsform er optimal sett i forhold til målsettingen.

Samme respondent forteller videre om at Innovasjonspolitisk avdeling har ansvaret for å legge en strategi, i nært samarbeid med resten av organisasjonen. De kan ikke ta alt på en gang så det må prioriteres å planlegges etter hvor det er størst behov for kunnskap nå. Eksempelvis vises det til at behovet nå er størst på kompetanse- og rådgivertjenestene og at det derfor legges ekstra vekt på disse.

Referansegrupper

Samtlige respondenter opplyser om at det opprettes en referansegruppe når man gjennomfører evalueringer. Dette skal bidra til å styre oppdragstaker inn på rett fokus og dermed sikre at evalueringen svarer på problemstillinger. En respondent sier at det alltid opprettes referansegrupper, med mindre det dreier seg om veldig små oppdrag. Det legges likevel vekt på at man ikke ønsker å kontrollere oppdragstaker i for stor grad, man er ute etter «fri-forskning» som én nevner. Men man er avhengig av at oppdraget løses i henhold til oppdragsbeskrivelsen, og at referansegruppene nedsettes primært for å bistå evaluatør og samtidig sørge for prosjektets fremdrift. Referansegruppene består gjerne av 10-15 personer om det er en viss størrelse på tjenesten og evalueringsoppdraget. Om referansegruppene forteller en respondent fra Innovasjonspolitisk avdeling at det er:

«Fornuftig å hente inn en referansegruppe, eller etablere en referansegruppe for hele prosjektet, som gjerne består av eksterne folk, forskere herfra og derfra og næringslivsfolk og andre som har både eierskap til det, eller som kan se det med litt distanse. Gjerne en litt bredt sammensatt referansegruppe. (...) Og da er jo den

referansegruppen satt opp primært for å bistå evaluator til hjelp. Både i valg av metodikk og kvalitetssikring av det, og vurdere funnene som gjøres. Kvalifisere funnene».

Det er også vanlig at eierdepartementet er deltaker i referansegruppen og man oppfatter dermed dialogen med eier som tett. Det vises til at eierdepartementet i hvert fall i åtte av ti tilfeller har representanter med i referansegruppene. Ofte hentes det også inn representanter fra andre som har levert tilbud uten å vinne oppdraget. Tilbudene som leveres har gjerne ulike styrker og svakheter, og dersom man kombinerer ulike miljøer og trekker inn parter fra konkurrerende tilbud oppleves det som positivt, ved at det tilfører flere «dimensjoner» i prosjektet.

Fra egen organisasjon er programansvarlig en naturlig deltaker i referansegruppen. I tillegg vil det være representanter fra Innovasjonspolitisk avdeling, og de bør være ordstyrer og innta en ledende rolle. De skal la program- og tjeneste ansvarlig komme frem med sitt syn, men ikke på en slike måte at man risikerer og låse diskusjonen i referansegruppen når temperaturen stiger. Respondenten forteller:

«... altså for å skape dynamikk i referansegruppa så er det viktig å ta en coaching rolle da, mentor rolle»

Innovasjonspolitisk avdeling vil også inneha evalueringsfaglig kompetanse, spesielt kan det være viktig når det gjelder valg av metodikk. Respondenten forteller at vurdering av det metodiske opplegget ikke er tjenesteansvarliges kompetansefelt:

«... de vil kunne stille gode spørsmål om valg av metode i etterkant. Men tror ikke de skal være dem som bestemmer det. Det vil ofte være at det er fjernt. (...) hvis du jobber med profilering av reiseliv så har du ikke nødvendigvis god peiling på økonometri, sånn er det bare».

Respondenten fra ledergruppen forteller at referansegruppene bidrar til en avklaring av mål, innretning og behov før man går i gang, og at de er en kilde for å sikre at man forstår virksomheten, muligheter, begrensninger og problemstillinger. I tillegg er det en kilde for å få respons på tolking av observasjoner.

Samlet sett peker materialet på at det er en tett involvering mellom IN og evaluator, men at det kanskje oppfattes som mest relevant og nødvendig i starten av prosessen slik at arbeidet settes i gang i riktig retning. Det er naturlig ettersom en evaluering som skal kunne benyttes, må svare på de problemstillingene som er reist i oppdragsbeskrivelsen. Dette korresponderer

for øvrig med de funnene som ble gjort i Statkonsults undersøkelse av evaluering i norske departementer i 1997, som er omtalt tidligere i oppgaven.

Det anses imidlertid ikke som et topplederansvar å delta i referansegrupper. Direktørene i ledergruppen som representerer de respektive tjenestene som blir vurdert, deltar dermed ikke. Men i noen av de større evalueringene vil lederen for Innovasjonspolitisk avdeling være deltagende, ellers er det hovedsakelig INs fagfolk fra fagdivisjonene og fagfolk fra Innovasjonspolitisk avdeling som sitter i de løpende referansemøtene.

Spredning og forankring

En av informantene fra innovasjonspolitisk avdeling snakker om hvordan de arbeider tett med tjenesteansvarlig når de gjennomfører evalueringer, men også etter at evalueringen er gjennomført og rapporten foreligger er samarbeidet med tjenesteansvarlig tett. Resultater og anbefalinger tas opp til en lederbehandling på tjenestenivå, hvor det vurderes hva som skal gjøres med de ulike resultatene som ligger i rapporten. Tjenesteansvarlig kommer så med en uttalelse, så «løftes» resultater og funn vider til toppledernivå.

Respondent fra Innovasjonspolitisk avdeling poengterer også at det vil variere fra tjeneste til tjeneste hvor strukturert evalueringsregimet er. For eksempel trekkes Arena/NCE, altså nettverkstjenestene, frem som et område hvor man har vært spesielt dyktig på gjennomføring og bruk av evaluering. Et annet eksempel fremgår i dokumentet «*Innovasjon Norges Årlig rapportering til oppdragsgivere 2015*». Fram-programmet er evaluert i 1997, 2006, 2010 og 2014, og i tillegg er tjenestene under dette programmet med i KEU fra 2003 og med i SSB-undersøkelsen fra 2015.

Respondenten forteller videre at det gjøres en vurdering når rapporten foreligger:

«Hvilke resultater og funn har vi her? Hvilke andre anbefalinger har evaluator kommet med, og hvilken vurdering gjør vi som evalueringsfaglige av evalueringen? Er den utført godt nok? Er svarprosenten god nok? Er metodene, empirien solid nok til å ha tro på resultatene? Så vurderer vi de anbefalingene som ligger der. Hva synes vi om de anbefalingene, virker de rimelige, forkaster vi dem, er det andre ting? Synes vi det er bom sånn at vi egentlig synes de burde satt fingeren på andre ting? Ja vi tar nå i hvert fall en vurdering, sånn at ledergruppen også vurderer hvordan vi går videre med de resultatene og anbefalingene som kommer fra evalueringene».

Videre opplyses det eksplisitt i strategidokumentet at lederforankring er en helt sentral del av arbeidet. Topplederen opplyser at Innovasjonspolitisk avdeling er ansvarlig for at det gjennomføres etter en viss plan, og hvordan det faglige innholdet skal være. Videre forteller respondenten:

«... når tingene foreligger så er det jo den som vi kaller tjenesteansvarlig eller programansvarlig som har et ansvar for å si at okay jeg leser dette, jeg vil foreslå ting som kan gjøres endring på og kjøre en type prosess på det. Så fremlegges resultater i ledergruppen hos oss. (...) så trekker man konsekvenser ut av det».

Videre trekkes det frem at man på selve tjenestene har et team som vil være viktig. Det er disse som er de «primære superbrukerne» av evalueringene. På spørsmål om hvordan man sprer resultater og kunnskap fra evalueringer i organisasjonen – svarer samme person:

«Det viktigste er å plassere hos de som har primærkunnskapen og kanskje først og fremst læringspunktene da. Og det skal programansvarlige gjøre sammen med Innovasjonspolitisk avdeling - prøve å trekke ut lærepunktene. Ikke sant, du får 300 sider, hva er det du skal trekke ut som gjør at du kan bruke det».

Videre oppgis det at hver fagdivisjon er representert av en direktør i toppledergruppen. Når det kommer en vurdering på en av tjenestene vil det da naturlig nok være direktøren som er tilknyttet den gjeldende divisjonen som både vil ha ansvaret og interessen for vurderingen. Det vil altså være fagdivisjonsdirektør og programansvarlig som er direkte berørt. Samtidig vil Innovasjonspolitisk avdeling være en sentral aktør. Det er disse tre aktørene «som må være veldig på hugget»

En av informantene i Innovasjonspolitisk avdeling opplyser om at graden av involvering, og opptatthet av evaluering fra toppledelsens side har variert med ulik ledelse. Respondenten gir også et eksempel på en tidligere leder som var klart mer opptatt og nysgjerrig på hvordan virksomheten kunne bruke evaluering i starten av perioden, og mindre de senere årene. Spesielt mente vedkommende at det ikke i alle tilfeller var nødvendig å ta resultater helt opp i toppledermøte, med mindre det dreide seg om de større evalueringene. Videre opplyser respondenten at topplederinvolvering også vil variere etter det som omtales som nivå. Det vil si mer involvering på større og viktigere oppdrag. Ledergruppen vil for eksempel kunne være mer interessert hvis det er et spennende eller nytt tema som er oppe for gjennomgang. Slik som programmer og tjenester som er under utvikling og fremvekst, heller enn de mer rutiniserte gjennomgangene.

En annen respondent i Innovasjonspolitisk avdeling opplyser at det har vært stor utskiftning i toppledelsen og at det en periode derfor har vært mye annet som skal falle på plass på ledernivået, men at det nå er i ferd med å settes høyere på agendaen fra toppledelsens side:

«... de begynner nå å se på dette her som er ting de har visst om at har rullet og gått, men som de ikke har hatt nødt til å ha fokus på i det halvannet året vi har hatt med ny administrerende. (...) de har det liksom ikke sånn tror jeg under huden enda, den nye ledelsen. Departementet minner jo den nye ledelsen på om at dette er saker vi også vil ha fokus på, så dette kommer mer».

Respondenten forteller videre at involveringen mellom både ledergruppe, tjenesteansvarlig og Innovasjonspolitisk avdeling er tett. Alle tjenestene har også en ansvarlig direktør i ledergruppa «så det kan ikke glippe der strengt tatt».

Respondenten fra ledergruppa opplyser om at det er noen i organisasjonene som er «først på ballen og virkelig kan de 100-200 sidene» i en rapport, men at ledergruppen ønsker å få den informasjonene som er destillert ut:

«Toppledergruppen leser ikke gjennom den fra A til Å, den vil ha sitt sammendrag og vi vil gjerne ha en sånn tosidert, hvor de egentlig oppsummerer hva som er læringspunktene og hva vi synes vi bør gjøre med det».

Videre opplyses det om at ledergruppen er mer opptatt av det som er av betydning på strategisk nivå, men at det er fagdivisjonene og leddene nedover som operasjonaliserer konklusjonene og jobber videre med de. Ellers skal det gis status og tilbakemeldinger til ledergruppen underveis, og en evaluering skal alltid via ledergruppen på slutten av prosjektet før konklusjoner er trukket og før noe offentliggjøres.

Respondenten fra ledergruppen forteller videre at det også finnes større undersøkelser som går på grenseflater mellom de ulike organisasjonene som iverksetter offentlig næringspolitikk, og at toppledergruppen i slike tilfeller vil sitte mye tettere på prosessen. Et eksempel på et slikt prosjekt er «Gjennomgang av grenseflatene mellom IN, Siva, NFR og fylkeskommunenes rolle innenfor næringsutvikling», et oppdrag som skal gi anbefalinger på rolledeling mellom de ulike aktørene (referat fra styringsmøte 26.11.2015 mellom NFD og IN)

4.1.3 Kompetanse og evalueringskapasitet

Det som jeg til nå har pekt på i diskusjonen synliggjør at IN er en organisasjon med betydelig kompetanse og kapasitet på evalueringsfeltet. Det er utarbeidet omfattende strategier og prosedyrer i organisasjonene, og de fire nøkkelinformantene har tydelig lang erfaring og kompetanse på feltet. Samtlige har arbeidet i organisasjonens forløper SND. Minst to av respondenten har jobbet der siden 1980-tallet. Programansvarlig har tidligere vært ansvarlig for evalueringsvirksomheten i SND. Det kan tenkes at dette kan bidra til at IN er spesielt godt rustet for å gjennomføre, og bruke, evalueringer.

Det kommer også frem gjennom intervjuene at IN engasjerer seg i faglige nettverk som kan bidra med kompetanseheving. En av respondentene i Innovasjonspolitisk avdelinge oppgir at IN er aktive i begge fagnettverkene i Norge, både i EVA-forum, som er et fagnettverk for evaluering i staten, og Norsk evalueringsforening.

Eva-forum har for øvrig et arbeidsutvalg ledet av DFØ, med deltakere også fra Kunnskapsdepartementet, Norad, Norges forskningsråd, Kommunal- og regionaldepartementet, Direktoratet for forvaltning og IKT og Innovasjon Norge. Temaene som tas opp er for eksempel knyttet til metode, bygging av evalueringskapasitet, bestilling og bruk av evalueringer (DFØ 2015).

I år gikk IN også inn i Norsk evalueringsforening. Noe de hadde avventet fordi de ikke ville stå for tett på noe som blir oppfattet som en kommersiell aktør, hvor ulike tilbydere også er deltakende. Men ettersom samarbeidet mellom de to fagnettverkene er tett, valgte de nå å gå inn som medlem også her.

Programansvarlig peker på at IN også deltar TCI Networks, som er en global forening for fagpersoner og beslutningstakere på klyngesatsinger. I denne sammenhengen deltar IN i TCI Evaluation Group, sammen med andre som jobber med samme problemstillinger, hvor man «forsøker å lære av hverandre sånn metodisk». Respondenten forteller videre om at nettverket består av deltakere fra hele verden, med en «indre krets» fra blant annet Sverige, Danmark og Scotland. Samme person mener det er tydelig at IN har kommet et stykke på vei når det gjelder evaluering:

«...hvor det også er helt klart at vi er nok kanskje de som har kommet lengst da på systematikk og helhet her».

Jeg var tidlig i teorikapittelet inne på begrepet absorberende kapasitet (Cohen og Levinthal 1990) som litt enkelt sagt dreier seg om en organisasjons evne til å produsere ny kunnskap og samtidig være i stand til ta i bruk og implementere denne kunnskapen i organisasjonen. Begrepet evalueringskapasitet kan relateres til dette. En ting er å ha kapasitet til, på en rutinisert måte, å gjennomføre evalueringer. En annen er å ha kapasitet til å benytte resultatene (eller prosessen) til noe. Respondenten fra ledergruppen snakker om at det er betydelige midler som årlig settes av til evaluering. Det ligger som forventning i økonomireglementet og de årlige oppdragsbrevene fra alle oppdragsgiverne, at det skal settes av penger til det hvert år. Samme person poengterer at det går en grense for hvor mange evalueringer og hvor mye kunnskap organisasjonen klarer å konsumere, og at målet er å kunne omsette kunnskapen i den daglige driften. Samme person snakker videre om at organisasjonene må være rustet til å absorbere denne informasjon:

«Vi bruker mange millioner hvert år på å gjøre ulike undersøkelser og evalueringer og det ligger som forventinger i alle oppdragsbrevene fra alle oppdragsgiverne og det settes av penger til det hvert år. Så om vi bruker rundt ti millioner eller mer enn det også i året (...) det ligger rundt der. Men det går en grense for hvor mye du klarere å konsumerer av den slags også».

Når respondenten blir stilt spørsmål om hvordan de ser på grenser for hvor mye som kan gjennomføres sier vedkomne at det ville vært kritikkverdig dersom de ikke var i stand til å ta informasjonen innover seg i organisasjonen. Det er derfor viktig å være tydelig på hva de ønsker å få ut av evalueringen, når man bestiller oppdraget. Respondenten mener videre at man har blitt flinkere med tiden å «tune inn» hva slags informasjon man er på jakt etter, heller enn å prøve å få med seg mest mulig på en gang. Evalueringen må derfor være mest mulig spisset. Det oppgis også at IN har fått betydelig kompetanse selv på hvordan man bør bestille og, hva man skal ha.

På videre spørsmål om man opplever å ha utfordringer knyttet til å absorbere kunnskapen svarer toppleren:

«Nei jeg opplever ikke det. (...) jeg tror at vi er inne på et omfang nå som er gjennomførbart hele veien. Og det kan godt hende at omfanget er litt lavere av evalueringer nå. (...) vi har hatt en ganske stabil tjenesteportefølje over tid. Som da betyr at vi har kunnskap på de fleste tjenestene allerede».

I denne sammenhengen forteller respondenten videre at man nå heller begynner å se på hvordan de ulike tjenestene henger sammen. For eksempel er det gitt rom for at man forsøker å se hvordan finansieringstjenester og rådgivningstjenester henger sammen. Det støttes for øvrig av andre respondenter at evalueringsregimet har utviklet seg i denne retningen.

Topplederen sier videre:

«Så er den store utfordringen vår at når evalueringen kommer og vi ser hva vi gjorde i det året, hvordan henger dette sammen, å gjøre den tolkningen. Det er den virkelige vanskelige delen som vår Innovasjonspolitiske avdeling jobber veldig mye med».

En av respondentene fra Innovasjonspolitisk avdeling forteller også at det er viktig å se ting i sammenheng, men at det:

«... er en stor utfordring for oss å gjøre gode vurderinger basert på alle disse ulike innspillene. (...) det er jo en svær jobb å være oppdragsgiver for en evaluering. (...) Fem- seks evalueringer i året så når vi et tak kapasitetsmessig».

Hvis man skal oppsummere status på evalueringskapasitet og kompetanse i IN, må det sies at organisasjon er godt etablert på evalueringsfeltet, både gjennom lang erfaring og kompetanse hos den enkelte som har dette som sitt felt, men også i organisasjonen som helhet. Samlet sett peker materialet på at IN kan sies å ha plassert analyse og evaluering langt frem på organisasjonsagendaen, selv om de som andre også helt sikkert opplever utfordringer. Hvordan virksomheten står seg sammenlignet med andre offentlige organisasjoner sier imidlertid datamaterialet i dette studiet lite om.

4.1.4 Hvorfor evaluere?

Spørsmålet hvorfor man evaluerer kan, egentlig ikke stilles uten at man samtidig stiller spørsmålet hvordan det brukes, fordi hvordan man bruker en evaluering vil avhenge av hva som var intensjonene i utgangspunktet. Likevel har jeg valgt dette som egen overskrift fordi jeg her er ute etter de mer overordnede beveggrunnene for at evalueringer iverksettes, slik som hvor styrende de formelle kravene er.

Jeg var innledningsvis i oppgaven inne på at evaluering er en plikt etter § 16 i økonomireglementet for staten. Reglementet ble fastsatt i 2003. Her heter det at alle virksomheter som er helt eller delvis eid av staten skal sørge for at det gjennomføres evalueringer (Finansdepartementet 2003). I tillegg fremgår det i § 4 i samme reglement at et

grunnleggende styringsprinsipp er å sikre tilstrekkelig styringsinformasjon og et forsvarlig beslutningsgrunnlag. I et møte med en representant fra DFØ, fremgikk det imidlertid at det fra Finansdepartementets og DFØs side ikke ble ført direkte kontroll med at denne bestemmelsen ble fulgt opp, og at det heller ikke fantes sanksjonsmuligheter. Spørsmålet er dermed hvor styrende økonomireglementet er for hvorfor IN gjennomfører evalueringer?

Som vi har sett ovenfor, gjennomfører IN et stort antall «rene» evalueringer. I tillegg eksisterer det en rekke andre lignende analyser og målinger av effekter.

IN har gjennomført både evalueringer og kundeeffektundersøkelser i en årrekke, lenge før det formelle kravet eksisterte. Det er derfor rimelig å hevde at § 16 ikke kan sies å være en utløsende faktor for INs evalueringsevne. En av spesialrådgiverne i Innovasjonspolitisk-avdeling, som har ansvaret for evalueringer sier at de gjennom tidene har vært et par tre personer som har arbeidet med evalueringer og at det strekker seg langt tilbake i tid. Om INs evalueringsevne sier respondenten:

«... det går jo tilbake til SND-tiden og forløpere før det også, men vi har nå drevet med evalueringsevne siden ja, i alle fall tidlig 80-tall».

Når respondenten snakker om INs evalueringshistorie tar den selv opp at det etter hvert har blitt lovpålagt etter § 16, og at det nå eksisterer et formelt krav til at evalueringer gjennomføres. Respondenten oppgir likevel at det ikke har vært spesielt styrende for virksomheten, men at det kan være en nyttig bestemmelse fordi det bidrar til at virksomheten må fokusere på evalueringer:

«For oss, som hadde et relativt etablert regime på det, så begynte ikke vi å evaluere fordi det reglementet kom, men det er jo skiftende hvor stort trykk ledelsen har på, og interesse for at vi skal evaluere. Så jeg synes jo at det hjalp oss at det reglementet ble etablert. (...) at det ble formalisert fra finansdepartementet sin side, så det hjelper oss i forhold til påtrykk og det å ha det høyt på agendaen i organisasjonen».

Informanten beretter videre at et slikt formelt krav til evaluering også eksisterer fra oppdragsgivernes side.

«Det ligger ofte krav fra departementene i tildelingsbrevene for en aktivitet at det og det året vil man ha en evaluering eller etter fire år så vil man ha en evaluering».

En av informantene har en sentral plassering i toppledelsen. Denne informanten viser også til at evalueringer er pålagt gjennom rammebetingelsene og at det derfor ikke er noe som kan velges bort:

«Der er jo sånn at som offentlig, eller en organisasjon, som bruker offentlige penger. Skattebetalernes penger som vi kaller det. Så skal vi kunne redegjøre både for hva vi bruker pengene til og hvilken effekt pengene gir og det følger som en del av det regelverket som offentlige institusjoner må ha. Det er den ene rammebetingelsen. Den andre rammebetingelsen er at når vi skal bruke noen virkemidler, eller tjenester som vi kaller det ut i markedet. Så skal de oppfylle åtte punkter, prinsipper for god virkemiddelbruk. Se setter krav til at vi skal kunne si noe om effekten, om det er en markedssvikt der ute. En del sånne konkrete ting, og for å kunne gå i retning av å mene noe om det, så må vi ha en evaluering på plass».

Her peker respondenten på at det er viktig å oppfylle krav fra omgivelsene. Man må kunne dokumentere at ressurser utnyttes på en fornuftig måte. Løpende målinger, slik som kundeeffekundersøkelsene, er kanskje spesielt godt egnet for å oppfylle et slikt krav. Kravene eksisterer både fra offentligheten, men også formelt sett gjennom krav til rapportering, som følger av tildelingsbrevet fra NFD og oppdragsbrevene fra departementene som bevilger penger til IN. Evaluering blir da et verktøy for å sikre kontroll fra overliggende departementer. På bakgrunn av dette er det tydelig at en viktig grunn for at analyser og evalueringer gjennomføres er formelle krav. Det er ikke noe som kan velges bort. Når det er sagt må ikke dette poenget overspilles. Det er også andre grunner for å gjennomføre evalueringer som er helt sentrale.

Det samlede datamaterialet peker på at de ansatte i IN er mer opptatt av helt andre ting enn formelle krav. Selv om formelle krav er en grunnleggende premisse. Samtlige informanter uttrykker at evalueringer er helt nødvendig, og at man uten hadde styrt i blinde, slik én uttrykker det. Evalueringene spiller en viktig rolle i tjenesteutvikling og forbedring og at det dekker et informasjonsbehov som ligger i det å styre etter et MRS-system. I evaluerings-strategien til IN heter det at «evaluering er et verktøy for å fremskaffe kunnskap som skal være grunnlag for strategiske valg og faglig utvikling» Det er spesielt det siste som blir tydeliggjort gjennom datamaterialet, det skal vi komme tilbake til i diskusjonen om bruk.

Før det, skal det trekkes frem et siste poeng i denne delen. Det er at IN som leverandør av næringsrettede virkemidler er i en særstilling sammenlignet med mange andre offentlige virksomheter. Grunnen til det er bestemmelsene som ligger i EØS-avtalen setter begrensninger

på muligheten for offentlig støtte til det private næringsliv. Dette gjør at IN må kunne dokumentere at de justerer for markedssvikt. Respondenten fra toppledelsen forteller at IN bruker penger inn i det «frie» markedet, og at man da kan si at man «tukler» med markedet. I henhold til EØS-reglementet kan man ikke gjøre det. Unntaket er når virkemidlene er notifiserte og godkjent av EU. Virkemidlene legger til grunn at det foreligger en markedssvikt, som gjør at den norske stat har lov til å gå inn å korrigere:

«... og som følge av det så må du dokumentasjon på om det er markedssvikt og om virkemidlene har effekt».

IN er med andre ord i en særstilling som gjør at caset muligens ikke er like sammenlignbart med andre deler av statsforvaltning, hvor krav til å dokumentere at virkemidlene har effekt ikke eksisterer på samme måte som det gjør for IN. Likevel vil behovet for faglig utvikling og kontroll, som formidles som meget viktig fra respondentenes side, også melde seg hos andre aktører i offentlig forvaltning.

4.2 HVORDAN BRUKES EVALUERINGER?

I denne delen skal jeg redegjøre for hvordan Innovasjon Norge bruker evaluering og hvordan dette samsvarer med brukstypologiene som er presentert tidligere. Et overordnet inntrykk er at evaluering er et viktig verktøy for styring og utvikling av virksomheten, og at måten evalueringer brukes på, samsvarer med de teoretiske konseptene.

I arbeidet med å systematisere empirien har jeg møtt på utfordring ved å dele opp i brukskategorier nettopp fordi det er så mange brukstyper i sving samtidig. Funn som peker direkte på konseptuell bruk kan i enkelte tilfeller også sies å peke vel så mye på instrumentell og legitimerende bruk. I dette studiet er det naturligvis ikke mulig å tallfeste hvordan andelen fordeler seg. Jeg vil likevel hevde at det er en tendens til at spesielt to bruksområder utkrystalliserer seg som viktige, krav fra omgivelsen (legitimering), læring og utvikling (konseptuell). Men at det vil også avhenge av hva slags type analyse og evaluering man snakker om. Datamaterialet støtter også at virksomheten heller tar i bruk funn enn bruk av prosess, selv om det også finnes støtte for det siste.

4.2.1 Prosess eller funn?

I teoridelen har vi sett at spesielt Patton har satt fokus på prosessuelle bruk, noe som viser til at aktiv deltagelse i evalueringsprosessen kan føre til kognitive og atferdsmessige endringer hos

ansatte og endring i programmer og organisasjon. Man lærer altså å tenke på en evaluerende måte (Patton 2008:108). I tillegg har vi sett at det finnes to grunnleggende måter å bruke en evaluering på. Bruk av prosess og bruk av funn. Jeg var også inne på at bruk av prosess kan være på en instrumentell, konseptuell eller symbolsk måte. Mens bruk av funn kan være på en instrumentell, konseptuell og legitimerende måte (Alking og Taut 2002).

For det første må jeg nevne her at det er svært utfordrende å avdekke kognitive og atferdsmessige endringer på individnivå, og at jeg heller ikke har forsøkt det. For det andre er det utfordrende og skille datamaterialet som peker på instrumentell og konseptuell bruk tilbake til om det stammet enten fra prosessen eller fra resultatene. Hvorvidt IN benytter seg av prosess, funn eller begge deler, har derfor blitt forsøkt avdekket ved at jeg har stilt spørsmål om hvor viktig virksomheten opplever at det er å være tett involvert og en deltagende part, i evalueringsprosessen? Deltar de i det hele tatt i prosessen? Opplever de at det er en læringsverdi i selve prosessen?

Samlet sett peker datamaterialet på at bruk av resultater er mer fremtredende enn bruk av prosess. IN har i stor grad vektlagt at evaluering skal bidra med et eksternt blikk på organisasjonen, og at man derfor har forsøkt å holde en viss avstand fra evaluatør. Det er tydelig at det er viktig del av legitimiteten til selve evalueringen at de ikke står for nært prosessen. Som vi skal se finnes det altså mest støtte for at IN bruker funn, men det finnes likevel støtte for at prosessen i seg selv, i det minste til en viss grad, også benyttes.

Grunnen til at det er rimelig å hevde at bruk av prosess er mindre viktig er at tre av respondentene oppgir at IN ikke ønsker å være for tett involvert og styre evaluatør. Flere av respondentene omtaler også gjerne evalueringer som «*eksterne evalueringer*», selv om de faktisk ikke har interne evalueringer. Noe som i seg selv signaliserer at legitimiteten som et blikk utenfra gir, er sentralt.

En av respondentene fra Innovasjonspolitisk avdeling forteller at når oppdraget er gitt, legger oppdragsbeskrivelsen og evaluatørs fortolkning av den i tilbudet, rammen for evalueringen. Deretter ønsker IN så mye «*fri forskning som mulig*». Respondenten forteller at de ikke ønsker

å styre en evaluering. Jobben blir da å følge opp evaluator på milepæler i forhold til fremdrift og leveranser og avrapportering som avtalt.

Den andre respondenten fra Innovasjonspolitisk avdeling argumenterer på lignende måte:

«... det er jo evaluator som skal stå for den siste skrivingen. Vi skal bare prøve å si noe om hvordan de skal ha fokus på forhånd. (...) men du jobber jo videre med evaluator når du begynner å jobbe i referansegruppen etterpå. Vi følger jo hver evaluering ganske tett.

På spørsmål om det er en viktig del av gjennomføringen å involvere seg tett med evaluator, svarer programansvarlig:

«Vi prøver å være veldig presise på selve bestillingen og at vi skal ha svar på de og de tingene der. Så pleier vi også naturlig nok å si at metodikken, hvilke data dere samler inn, det må dere selv ta ansvar for. Vi går ikke inn i det, vi skal ikke gå inn å mene noe om det».

Videre forteller samme respondent at involveringen heller dreier seg om det praktiske, slik som å forsyne evaluator med nødvendige dokumenter, eller å legge til rette for en survey ved å gi deltakerlister etc. Respondenten sier at *«det blir på en måte på det nivået der»*, og at det ellers er opp til evaluator å avgjøre hvordan man skal gå frem. Likevel oppgis det at det nok finnes eksempler på at man har bedt om en mer utdypende metodebeskrivelse før arbeidet går fullt i gang og at de har ønsket mer diskusjon om det i forkant av gjennomføring:

«... men det viktigste for oss er å reise spørsmålene, vi vil ha svar på det og det og det. Prøver å være ganske presise der, og selvsagt heller si at i den grad dere finner noe annet spennende så kom med det også».

Ovenfor var jeg inne på at det i de fleste tilfeller blir nedsatt en referansegruppe. Referansegruppene trekkes som regel frem når jeg stiller spørsmål om prosess. Dette kan potensielt være en arena hvor prosessuell bruk kan utspille seg. Datamaterialet peker imidlertid på at disse møtene i større grad brukes til å sørge for en utsjekk i startfasen av et prosjekt, slik at funnene skal bli mer matnyttig for organisasjonen. Som jeg var inne på ovenfor i kapittel 4.1.2, forteller respondenten fra ledergruppa at referansegruppen heller bidrar med en «avstemming» av mål, innretning og behov før man går i gang med prosjektet. Respondenten opplyser videre at evaluator har en faglig frihet til å kjøre prosessen, og at det er nødvendig for å oppnå tilstrekkelig uavhengighet:

«Jeg mener bestemt at det er en veldig «avstemming» før man går i gang helt i starten, når det gjelder «kravspekken», og hva man skal gjøre. (...) Ellers så er jo evaluator helt fri til å kjøre den prosessen, for vi må ha tilstrekkelig uavhengighet».

Et annet element som kan bidra til å styrke antagelsen om at det hovedsakelig er funnene snarere enn prosessen man bruker, er at KEU-ene og SSB-undersøkelsen ofte trekkes frem av respondentene, og dermed oppfattes som svært viktige. Disse undersøkelsen er kvantitative effektmålinger, slik at IN får informasjon om hvorvidt virkemidlene treffer. Det er med andre ord tallene man er ute etter, altså funn. Det er videre støtte for at disse tallene brukes aktivt i rapportering på MRS-systemet til eierdepartementet og andre oppdragsgivere. Det må sies å være sentralt for legitimering av virksomheten. Funnene fra analyser og evalueringer benyttes med andre ord til å få kunnskap om, og formidle videre, at organisasjonen leverer.

I forbindelse med arbeidet på denne oppgaven hadde jeg et møte med en representant fra DFØ. Vedkommende oppga at inntrykket var at det eksisterte to skoler på evalueringsfeltet om hvorvidt man benyttet seg av prosess eller funn. IN må nok da sies å være innenfor bruk av funn. Det er nok rimelig å anta at en tett involvering ikke ses på som forenlig med prinsippet om at man skal ha et eksternt blikk på organisasjonen, og at man derfor velger å begrense det.

Det finnes imidlertid også noe støtte i materialet for at selve prosessen benyttes på en konseptuell måte. Konseptuell bruk er når evaluering brukes til å påvirke kunnskap eller tenking om et tema. Med andre ord dreier det seg om opplysning og læring (Weiss 1979). Læring er ikke helt enkelt å operasjonalisere, men jeg har forsøkt å finne støtte for at IN opplever det som lærerikt å delta i prosessen eller om de eksterne utførerne bidrar med kunnskap utenom presentasjon av selve funnene. Et eksempel til støtte for at prosessen brukes på en konseptuell måte er følgende uttalelse fra topplederen:

«Vi har på en måte innovert litt selv. I mål og resultatstyring gjennom dette SSB-opplegget så vi går litt foran, og så er det sånn at vi vet jo ikke hva som er den beste måten å gjøre det på, så vi lar forskningsavdelingen i SSB ta steget videre. Så bruker vi Oxford og Menon og alle disse andre ikke sant. For vi sier at noen må holde fanen høyt på hva som er den beste faglige måten å gjøre ting på. Og så blir det jo kritikk og diskusjon, også lærer forskningsinstitusjonene dette og vi lærer, også tar vi steget videre. Så det er på en måte en læringsprosess og tolke dette også».

For øvrig vises det også i INs rapport til eierne for 2015 at man arbeider med å utvikle ny metode for effektmålinger på kompetanse og rådgivningstjenester i samarbeid med forskningsavdelingen i SSB. Dette kan tolkes innenfor rammen av prosessuell bruk. De har gjennom dialog og ekstern ekspertise knyttet til seg en mulighet for å ikke bare bruke resultatene som SSB kommer med, men en mulighet for at den eksterne kompetansen bidrar til å utvikle nye indikatorer og metoder som virksomheten tidligere ikke har benyttet seg av i analyser.

Videre snakker en respondent fra Innovasjonspolitisk avdeling om at referansegruppene:

«... er veldig konstruktive møter og veldig viktige læringsarenaer også for oss. Vi er ikke allvitere og vi trenger å få oversatt det som blir sagt til noe som blir matnyttig for oss (...) de milepælene tror jeg er veldig viktige sånn læringsmessig for vår del, ikke bare sluttproduktet».

Videre snakker respondenten om at et svar kan gi ti nye spørsmål, slik at man gjennom referansegrupper også kan vri prosjektet i en ny retning, og at man dermed gjennom møter underveis i prosessen har muligheten til å skape en form for tilleggsverdi, hvor nye problemstillinger kan reises med alternative forslag til hvordan de kan besvares.

I det siste tilfellet kan man kanskje i tillegg til læring snakke om instrumentell bruk av prosessen, ettersom man reorientere seg og tilpasser prosjektet som en direkte følge av diskusjoner i referansemøtene. Grensegangene mellom ulike brukstyper må imidlertid sies å være uskarpe, og som vi har sett i teoridelen kan mange former for bruk utspiller seg samtidig.

Det siste sitatet som skal trekkes frem i denne delen illustrerer et viktig poeng - mange av brukstypologiene er i sving samtidig. I sitatet finnes støtte for at prosessen benyttes og at det er en læringsverdi i det. Man ber samtidig om konkrete anbefalinger som det jobbes systematisk med i etterkant (bruk av funn). Her formidles det både prosessuell bruk, bruk av funn, konseptuell bruk og instrumentell bruk:

«Ja det er et viktig kunnskapsgrunnlag og det er viktig at dette er et eksternt blikk. Så det er på en måte den prosessen da. Det andre er læringsaspektet og jeg vil kanskje trekke frem disse programevalueringene. (...) vi kjørte altså et evalueringsoppdrag men to parallelle løp, gjennomført da av to ulike evalueringsmiljøer. Menon evaluerte Arena og Damvad Evaluerte NCE, der hadde vi en fellesprosess. Vi starta opp hele

den prosessen med et fagseminar med evaluator og en del ressurspersoner for å diskutere litt ja... hvordan river vi opp dette her? Hva er de relevante spørsmålene? Hva er de relevante metodene? Så kjørte vi en utveksling eller et samspill underveis i evalueringsprosessen og så bestilte vi på slutten en felles oppsummerende evalueringsrapport hvor begge de to miljøene var bidragsyttere. (...) der ba vi også evaluator om å gi konkrete anbefalinger til forbedringer og da jobbet vi systematisk med de anbefalingene».

4.2.2 Evaluering som direkte premiss for beslutning

I teorikapittelet var jeg inne på at instrumentell bruk av evaluering er tilfeller hvor evalueringer brukes til å endre det som er gjenstand for evaluering og at det gjerne referer til direkte handlinger som har blitt gjennomført på bakgrunn av en evaluering (Henry og Marks 2003). Eller som Rich sier, handler det om spesifikke og dokumenterbare tilfeller hvor evaluering har blitt brukt til å fatte beslutninger (Rich 1977:200).

Dette kan sies å være et kjernepunkt i evalueringsfeltet, men enkelte har tatt til ordet for at avgjørelser og aktiviteter i en organisasjon sjeldent kan spores direkte tilbake til en spesifikk evaluering (Patton et. al. 1977:161). Vi har også sett Weiss (1980) understreke at evalueringer ofte ikke benyttes på en instrumentell måte fordi evalueringer sjeldent leverer så tydelige svar som aktørene kan anvende direkte. Denne delen vil forsøke å belyse om Innovasjon Norge kan vise til og dokumentere spesifikke tilfeller hvor evalueringer har blitt brukt til å ta beslutninger? Hvilke direkte handlinger er blitt gjennomført som resultat av evaluering? Hvordan brukes informasjonen fra en evaluering direkte i beslutningstaking og tiltaksforbedring?

Det empiriske materialet støtter langt på vei en antagelse om at det viktigste med evalueringer ikke nødvendigvis er at de utgjør et direkte grunnlag for beslutninger. Men som vi skal se finnes det kanskje noe overraskede betydelig støtte for instrumentell bruk.

Samlet sett er imidlertid respondentene mer opptatt av at evalueringer bidrar med retning og utvikling av tjenestene. Innledningsvis i dette kapittelet så vi også at informanten fra ledergruppen snakket om at de ulike undersøkelsene utgjør et samlet kunnskapsgrunnlag. Respondenten understreker også at det «først og fremst er læringen som er viktig». Videre forteller toppleren at det er utfordrende og tolke resultater fra en evaluering:

«Det er jo ikke så svart hvitt at det enten er sånn, eller så er det sånn. Dette er tolkningsspenn hele tiden».

Videre har vi sett at organisasjonsadferd ikke alltid følger en rasjonell styringsmodell, men kommer i stand gjennom deliberasjon. Beslutninger vil da være basert på summen av kunnskap og dømmekraft som deltakerne bærer med seg (Weiss 1988). Evaluering vil da være en av flere informasjonskilder som man trekker på, uten at den gir helt klare svar. Lederen nevner at:

«Det er aldri noe ja og nei svar, det er ikke noe svart hvit i disse undersøkelsene. Det er alltid litt greit der og litt ugreit, men det er aldri håpløst. Så du må liksom velge hvor toleransegrensen går hen».

«Hvis det kommer opp at noe treffer dårlig så er ikke det første vi gjør å si at da nedlegger vi det. Da er spørsmålet hva er galt? Hva kan vi vri å gjøre? Så teoretisk mekker vi den til, så setter vi den på lufta igjen videre».

Respondenten som er programansvarlig er opptatt av at evalueringer bidrar med viktig informasjon, og at de også kan fungere som et grunnlag for at beslutninger fattes, men at det ikke alltid foregår på en direkte og entydig måte:

«... det kan også være at evalueringen (...) viser at det har ikke vært så imponerende resultater. Men med et strategisk løft for neste periode, så kan vi kanskje si noen ganger at, ok det har ikke vært all verden det dere har gjort så langt, i henhold til evalueringen, men vi har trua på dere videre. Så vi bruker jo ikke evalueringen sånn direkte. Det gir ikke direkte svaret på en beslutning da. Hvis du skjønner».

En av informantene fra Innovasjonspolitisk avdeling sier at det vil variere i hvor stor grad en evaluering kan sies å ha fungert som direkte premiss for beslutninger, og at dette også har en sammenheng med kvaliteten:

«... hvis du tar en evalueringen og legger den på bordet og sier hva brukte vi denne til så er jeg litt sånn, nja nei kanskje ikke så mye. Mens for en annen evaluering kan du si ganske klart at vi har fulgt anbefalinger slik og slik, eller vi har valgt å ikke gjøre det fordi sånn og sånn. Det vil variere fra evaluering til evaluering».

Samme respondent trekker samtidig frem at kvaliteten på en evaluering har betydning for om den blir brukt eller ikke. En evaluering som treffer virksomheten og deres virkelighetsforståelse vil naturligvis ha større effekt.

Som jeg nevnte ovenfor er det likevel støtte for at evalueringer brukes instrumentelt. Det er spesielt interessant at alle informantene kan vise til helt konkrete tilfeller hvor en tjeneste er lagt ned på bakgrunn av resultater fra en evaluering. Selv om respondenten fra ledergruppen understreker at de ikke har noe særlig «track record» på at man gjennom undersøkelser har identifisert noe som har fungert så dårlig at det har blitt lagt ned, så finnes det enkelte eksempler. Et eksempel som trekkes frem er det KMD-oppdrag, hvor det på forhånd gjennom utredning var knyttet store forventninger til at tilførsel av ekstra midler ville utløse verdiskapning og effekter. Evalueringen som fulgte ga tilbakemeldinger om at det ikke fungerte etter hensikten og tjenesten ble følgelig lagt ned. I tillegg viser både topplederen og en informant fra Innovasjonspolitisk avdeling at evalueringen av etablerer.no, var en direkte premiss for at tjenesten ble lagt ned. Sistnevnte sier følgende

«Det viste seg at det var ikke interaktivt nok det nettstedet til at det ga utbytte nok for kunder eller potensielle kunder, så det ble lagt ned, for eksempel. Vi har også lagt ned bit-programmet og du kan si enten blir det lagt ned eller så innlemmer vi det i andre ting. Men at evalueringer har blitt brukt til å styre vekk og bort og til å omforme det er helt sikkert».

I tillegg opplyser topplederen at IN har noen frihetsgrader på hvordan de kan fordele midlene på enkelttjenester:

«Da er det sånn skal vi putte dem hit eller dit? Og da kikker vi jo på effekter som kommer ut av det. Våre oppdragsgivere ser veldig på det. De styrer veldig tydelig. I så måte så er det et eksempel i innværende statsbudsjett hvor det var en tjeneste som forsvant. Hvor evalueringen viste at det ga ikke noe særlig effekt. (...) det var landsdekkende innovasjonstilskudd som næringsdepartementet fremmet. Vi så på undersøkelsen (...) så at det var behov for å prioritere bort noe. Da ble det vektlagt ganske kraftig, og så gikk den ut».

Den andre respondenten fra Innovasjonspolitisk avdeling viser til at evalueringer kan gi signaler om at «det er nok nå» og viser til at et tidligere Nettverksprogram ble lagt ned på bakgrunn av evaluering. Man så at etterspørselen begynte å avta. På tidspunktet for nedleggelse hadde programmet vart i seks-sju år og, et par år etter det igjen startet man isteden å utvikle nye konsepter på nettverkstjenester, henholdsvis Arena og NCE. Respondenten forteller videre i denne sammenhengen: «Så det er viktig å bruke evalueringer til læring, men ikke nødvendigvis som domsavsigelse».

Respondenten som er programansvarlig på Arena/NCE viser til samme evaluering:

«... det var en evaluering som fikk toppoppslag i DN, at her var det mye offentlige penger og lite resultater, og sammen med alle andre tin, så ga nok det grunnlag for å si at nå stopper vi her, vi får tenke oss om, så har vi senere reetablert dette her. Og det kan godt hende at IN har flere sånne eksempler, men det var en veldig klar sammenheng».

Respondenten som er programansvarlig viser også til at det på dette området ikke er så mange tilfeller hvor evalueringer har ført til at tjenester er tatt ut. I sitatet ovenfor nevnes det også at evalueringen sammen med andre ting ga et grunnlag for beslutning.

Respondenten forteller videre at det i visse tilfeller finnes konkrete eksempler på noen «gode betraktninger» fra evalueringer, som har fått direkte konsekvenser for beslutninger. I et tilfellet var tilbakemeldingene fra en evaluering «*såpass klare, og det var såpass mange som pekte på et problem, og da løftet vi dette frem og det må vi gjøre noe med. Så har vi gjort noe med det.*».

Respondenten vektlegger likevel at evaluering som regel inngår som en av flere begrunnelser for at beslutninger fattes, og ikke alene vil være styrende. Flere av respondentene snakker da også om at evalueringer er viktig som «*støttende analyser*». Det er forøvrig forenlig med Weiss antagelse (1980) om at evalueringer «*sniker*» seg inn i beslutningstaking på mer indirekte måter.

Ut ifra teorien kan man forvente at det er utfordrende for beslutningstakere og være i stand til å vise til direkte beslutninger som er fattet på bakgrunn av en evaluering. Vi skal også være klar over at informantene er stilt spørsmål om evaluering generelt, og ikke om spesifikke evalueringer. Det kan være en stund siden de er gjennomført, noe som gjør det vanskelig å vise til konkrete tilfeller av direkte bruk. Informantene er imidlertid i stand til å vise til flere avviklinger som de mener er direkte konsekvens av evaluering. Men det må da også sies at det er en stor endring, inkrementelle endringer kan også være instrumentelle, men da er det kanskje spesielt utfordrende å kunne vise til spesifikk og dokumenterbare bruk. I tillegg vil respondentene forsøke å lete etter tilfeller hvor bruken kan sies å være direkte, når det stilles spørsmål om det. Uten at det nødvendigvis vil være representativt for hvordan evaluering vanligvis brukes.

4.2.3 Læring og utvikling

Evalueringer gir et kunnskapsgrunnlag som sammen med andre faktorer bidrar til læring og utvikling. Det finnes bred støtte i datamaterialet for at det oppleves som et viktig aspekt ved å gjennomføre evalueringer i IN. Jeg var også tidligere inne på at det finnes en viss støtte for at det er et læringsaspekt i selve gjennomføringsprosessen. Selv om man hovedsakelig er mest opptatt av funn. I tillegg så vi ovenfor at selv om beslutninger til dels fattes som en direkte følge av evaluering, er det hovedsakelig mer indirekte bruk som blir tydeliggjort gjennom intervjuene. Evaluering blir dermed en av flere kunnskapskilder som bidrar til å fastsette retning og utvikling.

I dette tilfellet kan man altså snakke om bruk av evaluering som tenking rundt en tematikk og som utdanning av beslutningstakere (Cousins og Leithwood 1986:345). Konseptuell bruk er en brukstype som er nært relatert til læring. Evaluering peker gjerne ut styrker og svakheter ved en tjeneste og bidrar dermed til å «opplyse» beslutningstakere ved å påvirke oppfatninger om den nåværende tilstanden, og ideelle måter å strukturere et tiltak på (Ibid). I dette perspektivet er evalueringer kilde til ny kunnskap og informasjon på feltet, som setter beslutningstakere og ansatte i stand til å ta kunnskapsbaserte avgjørelser om hvordan tjenestene bør utformes.

Det er utfordrende å avdekke hvordan evalueringer brukes til læring og utvikling, nettopp fordi det er tilfeller hvor informasjonen blir brukt til påvirkning, og ikke «*til spesifikk og dokumenterbar bruk*» (Rich 1977:200, egen oversettelse). Dersom respondentene ikke viser til spesifikk bruk, blir det naturligvis utfordrerne å undersøke. Dette er derfor operasjonalisert ved at jeg gjennom et åpent spørsmål - hvordan brukes evaluering? - Har sett etter hvordan informantene snakker om læring og utvikling. I enkelte av intervjuene stilte jeg også mer direkte spørsmål om hvor viktig læringsaspektet var når virksomheten gjennomførte evalueringer?

Begrepene læring og læringsaspekt blir brukt hyppig av informantene, uten at jeg nødvendigvis har stilt spørsmål om det på forhånd. Læring er heller ikke så lett å synliggjøre gjennom et intervju på annen måte enn at respondenter snakker om at evaluering bidrar med ny kunnskap og læring. For eksempel utaler informanten fra toppledelsen at det «*viktigste for oss er læring og forbedring*» og at «*læring det har vært hovedhensikten og det har vi flagget hele tiden*». Respondenten argumenterer videre for at de vektlegger læringsaspektet ved at man bruker evaluering til å gjøre ting annerledes.

Det siste er et viktig poeng, fordi læring i en organisasjon skal føre til noe. Man skal gjøre det man gjør bedre. I teoridelen og under diskusjonen om evalueringskapasitet var jeg inne på

begrepet absorberende kapasitet som viser til at en organisasjon både må være i stand til å skaffe seg ny kunnskap, men også være i stand til å implementer kunnskapen slik at den får praktiske konsekvenser (Cohen og Levinthal 1990). Alle respondentene er tydelig på at kunnskapen som erverves gjennom evaluering er viktig for å endre på og utvikle tjenestene, noe følgende uttalelse fra en informant i Innovasjonspolitisk avdeling illustrerer:

«... tror læringsaspektet er desto viktigere. Altså læring for flere hensyn. En ting er å gjøre programmet mer effektivt som da er objektet for evalueringen, men også som idé til nye tjenester, nye programmer, vridninger, utfasing av eksisterende kanskje oppbygging av noe nytt».

En annen respondent fra Innovasjonspolitisk avdeling argumenterer for at man er spesielt opptatt av den faglige utviklingen, og at hele poenget med å gjennomføre prosessevalueringer (formative evalueringer) er å videreutvikle tjenestene. Alle undersøkelsene og analysene som IN gjør skaper et kunnskapsgrunnlag som også bidrar til å utvikle nye tjenester:

«Man har et kunnskapsgrunnlag i forhold til hva som virker og ikke virker ikke sant. Sånn at når man skal utvikle nye tjenester så vet du hvordan du skal sy opp en ny tjeneste. Hvilke målgrupper og kunder har størst suksessrate? Hvilke kriterier bør man stille til kundene for at suksessen skal bli størst mulig? Så vi bruker definitivt kunnskapen vi erverver ved evalueringer til å etablere nye tjenester, det gjør vi absolutt».

I tillegg mener samme respondent at virksomheten nå prøver å fokusere på hvordan «bundlingen» av tjenester fungerer. Man ser altså nå på hvilken pakke av tjenester kunden har fått for å oppnå gode resultater og bli en «superinnovatør», altså en kunde som har høy score på mange av INs indikatorer.

«Når vi bruker den kunnskapen som suksesskriterier og kombinasjoner av tjenestepakker og sånn, så kan det være nyttig i forhold til å utvikle nye tjenester, sette de sammen på nytt og kanskje lage en tjeneste av en som har bestått av tre tjenester før eller noe sånt».

Respondenten fra ledergruppen forteller at evaluering gjennomføres for å få informasjon om effekter, men at det er vesentlig å bruke informasjonen til å styre tjenestene i en ønsket retning:

«Ja det er det altså i forhold til å vite hvordan vi treffer også da for å korrigere kursen, men læringen bruker man da til å sette eventuelt en ny retning eller noen nye

mål da ikke sant. Og det er nok det å ha kontroll fra statens side, er jo også da å kunne vise at, ok, du kom med en evaluering og du gjorde noe med den. Så det er styring altså, det at du tar en konsekvens av det du ser som er det avgjørende».

Når det gjelder læring og utvikling virker respondenten som er programansvarlig spesielt relevant. I et tilfell viser programansvarlig til følgende:

«... også har vi også veldig positive erfaringer med at vi selv føler at vi har ganske bra overblikk nå, så vil vi med en god profesjonell evaluering alltid få nye innspill og betraktninger. Det fikk vi i den siste runden, spesielt i Menons evaluering av Arena. Den plukket opp en del ting som vi nok ikke var så bevisst på og ikke det at den Menon-evalueringen anbefaler noe... sånn må det være, men at det hvert fall reiste flere diskusjoner»

Jeg var tidligere inne på at Arena/NCE er et felt hvor det blir oppgitt at man har et særskilt godt system på evaluering.

«Vi er jo veldig tett på, tettere på enn de fleste programmene her i IN, så vi vet mer om feltet enn de fleste. Sånn sett er det kanskje færre overraskelser for oss. Men likevel er erfaringene at vi får alltid verdifull informasjon som vi kan bruke til noe. Og nå, når vi begynner å forberede den neste programevalueringen så har vi også på en måte hele veien med oss noe, jo vi trenger å skjønne noe mer av deler av... vite noe mer om det... vi trenger et eksternt blikk på de og de tingene. Så vi har samlet opp en del sånne spørsmål som vi har lyst til å få belyst i evalueringen. Vi tenker absolutt sånn. Det gjør vi. Så det er ikke noe sånn at vi har pålagt oss noe, også gjør vi det bare for at vi må gjøre det».

Som nevnt kan det være utfordrende å få tak på hva læring nøyaktig brukes til. Læring er abstrakt og kan bestå av både «taus» og «eksplisitt» kunnskap. Individene i en organisasjon tenker, vurderer og argumenterer basert på den samlede erfaringen og kunnskapen de besitter. Programansvarlig ble stilt spørsmål om hva man bruker læring til og hvorvidt vedkomne opplever at det er utfordrende å vise direkte til evalueringer når det gjelder læringsperspektivet:

«Jo men jeg mener nok det at som jeg har gitt eksempler på, den prosessen vi hadde i forkant av forrige programevaluering, så brukte vi evaluering veldig direkte inn i en lærings og utviklingsprosess, da var det på en måte planlagt. Da skulle vi gjennomføre den evalueringen. Så skulle vi gå løs på en strategirevisjon... programrevisjon, og det

gjorde vi. Men så er det klart at den samlede læringen fra og rundt programmet er jo satt sammen av en god dose løpende erfaringslæring en sånn... snakke med folk... høre med folk, og vi vet jo det og det skjer og sånn og sånn gjør man det. Supplert da med mere disse analytiske tilnærmingene. Hvor evalueringene er ett instrument, men vi har jo også andre analyser».

Siste del av sitatet ovenfor understreker et poeng om at det er en samlet erfaringslæring som danner grunnlaget for at beslutninger fattes, slik vi har sett Weiss argumentere for. Videre til støtte for at evalueringer påvirker på subtile og tvetydige måter, sier programansvarlig i et tilfellet følgende:

«Vi hadde da antagelig gjort den samme beslutningen uten den direkte referansen i den evalueringen, men det er i noen tilfeller hvor det er på en måte litt sånn komplekst og litt vanskelig å se det totale bildet, så vil en ekstern evaluering ofte være veldig nyttig for da får du dette eksterne blikket, og du får stort sett, hvis den er profesjonelt gjennomført en god og systematisk oversikt. Du får noen sammenhenger og hva slags virkningsanalyse også videre. Så det er absolutt viktig for å systematisere begrunnelser og ting da. Selv om i bunn og grunn kanskje konklusjon hadde vært den samme. Så er veldig profesjonelt utført evaluering nyttig da i sånne situasjoner».

Selv om det kan være utfordrende på en presis måte å fastslå hva evalueringer har bidratt med, hevder Weiss (1999:468) at evalueringer utvilsomt har reelle konsekvenser. De kan utfordre gamle ideer, tilby nye perspektiver å gjøre om på agendaen. I den forbindelse kan det nevnes at programansvarlig sier at de er helt avhengig av å kunne stoppe opp underveis, og at de ikke kunne klart seg uten dette verktøyet.

I denne delen har jeg presentert et knippe uttalelser fra informantene som peker i retning av at evalueringer brukes til læring, forbedring og utvikling. I den forbindelse må det nevnes at det er en betydelig mengde med informasjon fra datamaterialet som kan tolkes i retning av konseptuelt bruk. Samlet gir det et inntrykk av at informantene opplever at analysene som virksomheten gjennomfører er helt sentrale verktøy for styring og utvikling. Funnene støtter også langt på vei en antagelse om at evalueringer bidrar *«in a more subtle, clarifying, reinforcing, and reorienting way»* (Patton et al 1977:161).

4.2.4 Evaluering - et verktøy for å legitimere innsatsen?

En av de mulige brukstypene som ble gjennomgått i teoridelen var legitimerende bruk. Dette ble presentert som en egen brukstype samtidig som vi så at enkelte heller kategoriserte legitimerende, taktisk og symbolsk bruk under én fane. Slik som Højlund med samlebetegnelsen «justificatory uses». De har det til felles at det dreier seg om rettferdiggjøring av organisasjonen ovenfor omgivelsene (Højlund 2014:27).

Vi så i tillegg at Vedung (2009) argumenterte for at det er to situasjoner hvor legitimerende bruk kan komme til uttrykk. Det første er i tilfeller hvor man ønsker faktabasert støtte for en eksisterende overbevisning, det andre er hvor man anser egen posisjon som rimelig, men det eksisterer en viss tvil, og man ønsker argumenter som støtter posisjonen. Nå har ikke denne oppgaven undersøkt hva som var virksomhetens standpunkt forut for evaluering, slik at det ikke vil være mulig å gjøre en slik direkte kobling.

Inntrykket er at respondentene bruker begrepet legitimitet som man gjør i dagligtalen, altså at man er opptatt av støtte og aksept fra omgivelsene, og at evalueringer bidrar til det. Ovenfor har jeg også vært inne på at man som offentlig organisasjon er forpliktet til å kunne dokumentere at midlene brukes på en forsvarlig måte. Evalueringsresultater brukes dermed som «ammunisjon» eller evidens for å vise at man oppnår gode resultater (Weiss 1977:14). Hva er så omgivelsene? Som for andre organisasjoner kan det både være kunder, media og opinionen. Det fremkommer for eksempel i datamaterialet at man ønsker å kommunisere ut gode resultater gjennom; «*kunnskapsbaserte uttalelser ut i brosjyremateriell eller presse*», slik en informant i Innovasjonspolitisk avdeling formulerer det. Et eksempel på det er brosjyren «*Innsats gir resultater*» hvor det vises til at man gjennom et samarbeid med SSB er i stand til å presentere tall som gjør at man vet hvordan man lykkes (IN 2014:3).

Slik det fremgår av intervjuene oppfatter jeg imidlertid at det viktigste i denne sammenhengen er legitimitet ovenfor eierdepartementet og de andre oppdragsgivende- og bevilgende myndigheter.

Sentralt her er derfor også kontroll. Et MRS-system baserer seg i all enkelhet på at man har satt noen mål for virksomheten og deretter utformer indikatorer som tar sikte på å måle, effekter. Uten evaluering og andre analyser ville man ikke hatt særlig å rapportere om. Når man rapportere til eiere, vil det kunne dreie seg om legitimering, samtidig som den samme informasjonen gjerne også kan benyttes som direkte grunnlag for beslutninger, til læring og utvikling.

To av respondentene trekker frem legitimerende bruk på eget initiativ. To andre trekker det frem på spørsmål om evalueringer har en bredere interesse enn styring?

Til en respondent i Innovasjonspolitisk avdeling ble det også stilt spørsmål om det gjennomføres evaluering hvor man hadde en forventning om hva resultatene kom til å være på forhånd, og på den måten fikk bekreftet noe man allerede mente, tenkte eller viste?

Respondenten svarer «*ja, legitimering er jo ett aspekt ved evalueringer*». Videre viser vedkommende til at spesielt de nye effektmålingene basert på objektive registerdata med SSBs Forskningsavdeling som operatør, vil kunne gi økt legitimitet til større deler av virksomheten.

Samme respondent trekker også frem a legitimeringseffekten, når det gjelder enkeltevalueringer, har blitt mindre viktig nå enn hva det var tidligere. Vedkommende mener at grunnen til det er at viktigheten av innovasjon og fornyelse i næringslivet er høyere i folks bevissthet nå enn hva det var før. Hvor viktig legitimeringseffekten oppfattes å være vil også avhenge av nivået det evalueres på. Legitimeringseffekten er viktigere når hele systemet evalueres på tvers, enn hva det er på enkeltevalueringer av tjenester. Et eksempel på en slik undersøkelse er «*Evalueringen av Innovasjon Norge*» fra 2010. Den ble bestilt direkte av NFD, og ikke av IN selv. En annen respondent fra Innovasjonspolitisk avdeling nevner forøvrig at evalueringer av hele organisasjonen skal være vært femte år.

Det som er nevnt ovenfor kan signaliser at KEU og SSB-undersøkelsen som er brede kvantitative effektmålinger, er viktigere for å legitimere innsatsen enn hva som er tilfellet for evalueringer av enkelttjenester. Sistnevnte synes viktigere som utviklingsverktøy.

I tillegg er det ifølge en av respondentene forskjell på hva man får ut av en sluttevaluering og en prosessevaluering, hvor selve formålet med sistnevnte er at det skal være et utviklingsverktøy. Sluttevaluering er da kanskje heller et instrument for å vise at det man har gjort fungerte.

Respondenten fra toppledergruppen sier følgende som peker mot at analyser og evalueringer brukes på en legitimerende måte:

«... det er med å skape legitimitet på hva vi faktisk oppnår. For hvis det er slik at vi ikke kan vise til at vi skaper effekter, så vil det ofte slå ut i at man vil være mer betenkt i statsbudsjettene og bevilge mindre penger til formålet. Sånn at det er med på å skape legitimitet for bevilgende myndigheter».

Videre sier samme person:

«... vi viser til resultatene i vår årsrapport og vår oppdragsgiverrapport som er to hovedkilder tilbake til departementet og oppdragsgiverne. Så det er jo liksom både læring ja, men det er hele legitimiteten (...) på hvorvidt vi har skapt gode resultater (...) Det er det vi måler på, vi måler liksom ikke på så mye annet».

Bruk av analyser og evaluering er altså viktig som legitimerende instrumenter ovenfor departementene. Det hører også med at «INs Årlig rapportering til oppdragsgivere 2015» i hovedsak dreier seg om å informere om resultater og effekter som er samlet inn gjennom de ulike undersøkelsene.

Programansvarlig snakker også om at det er tydelig og formalisert hvordan man skal rapportere på de ulike effekt- og resultatindikatorerne, og at årsrapportene hovedsakelig inneholder data fra «monitoreringsopplegget» (de kvantitative effektmålingene), men at evalueringer også brukes i dette:

«... ja også legitimere innsatsen og ha noen å vise tilbake til departementene som finansierer klyngeprogrammet. Så vi skal også melde tilbake, og det gjorde vi selvsagt også, og si at her har vi gjennomført denne evalueringen med de og de konklusjonene. Så ble det tatt videre til begge de to departementene som kom med sine pressemeldinger. At nå er klyngesatsingen evaluert og vi er veldig fornøyd med det, så også en del av legitimeringsprosessen».

Samme respondent sier på spørsmål om man viser direkte til evalueringer i rapporteringen:

«Ja absolutt, absolutt, og for å ta et annet eksempel fra et bedriftsnettverk som ligger nært inn til klyngesatsingen som ble evaluert i 2013. Årsrapporten da for 2014, refererte veldig konkret til den. Så det er jo også noe av dette her som vi har avtalt med departementene. Vi utnytter eksterne evalueringer inn i disse årsrapportene. Så de evalueringene (...) er særdeles viktig for å legitimere og få aksept for videre finansiering fra departementene, det er det ingen tvil om».

Sitatene ovenfor peker tydelig på evalueringer brukes som faktabasert støtte for å dokumentere at det man gjør fungerer ovenfor omverdenen. Evaluering brukes dermed for å legitimere innsatsen.

Selve kontrollbegrepet kan imidlertid diskuteres litt videre. Vi kan si at rapportering til departementer er sentralt i styring på en «armlengdes avstand», og at det eksisterer en

formalisert form for kontroll ovenfra, gjennom et MRS-system, hvor IN legitimerer virksomheten ved å kunne vise til at man skaper effekter.

Men det kan være like relevant å se legitimerende bruk av evaluering som en metode IN har til å kontrollere egne omgivelser. Man får «politisk ammunisjon» som Weiss har kalt det (1977:14), og er dermed i stand til å fremme egne synspunkter og overbevisning, og på den måten påvirke omgivelsene. For eksempel ved å sikre at midlene blir bevilget dit IN mener at de får størst nytte. Legitimerende bruk av evaluering kan derfor potensielt være et viktig instrument for å «tøyle» departementene og dermed få gjennomslag for egen vilje.

Helt til slutt kan jeg også trekke frem at de samme analysene og evalueringene som IN bruker som legitimerende instrumenter ovenfor sine oppdragsgivere, naturligvis også kan bidra til å legitimere oppdragsgiverne selv. Uten at det har vært gjenstand for undersøkelse i dette studiet.

4.2.5 Evaluering er en trend?

I følge Vedung er evaluering en «megatrend» som organisasjoner gjennomfører for å vise modernitet (Vedung 2009). Symbolsk bruk av evaluering knyttes til det institusjonelle perspektivet på organisasjoner, som ble omtalt innledningsvis i teorikapittelet. Det bygger på en forståelse av at organisasjoner utvikler en felles oppfatning av verdier, «oppskrifter» og rutiner, som blir reproduert og tatt for gitt (Dhaler-Larsen 2003:57). Organisasjoner må også søke legitimitet fra omgivelsene ved å demonstrere at man lever opp til fundamentale modernitetsnormer. Disse normene kommer gjerne til uttrykk gjennom generelle ideer og tydelig oppskrifter (Røvik 2008:26).

Når det gjelder symbolsk bruk av evaluering, vil det til forskjell fra legitimerende bruk, dreie seg om bruk av selve prosessen (Alkin og Taut 2003). Det er svært utfordrende å avdekke at evalueringprosessen blir brukt på en rituell/symbolsk måte. For det første, fordi det er tydelig at evalueringer også brukes både instrumentelt, konseptuelt og legitimerende. Dersom opplevelsen var at evaluering ikke ble tatt i bruk på andre måter, og man fremdeles gjennomførte det, kunne det ha pekt i retning av at evalueringer ble gjennomført som et ritual. For det andre snakker man om at organisasjonsoppskrifter er tatt for gitt som en viktig og riktig måte å organisere på, og at symbolske handlinger derfor gjerne fremstår som grunnleggende rasjonelle handlinger (Røvik 1998:36). Det innebærer at evalueringer kan være rasjonaliserte

symboler, altså at det oppfattes av organisasjonens egne som et effektivt redskap for forbedring og modernisering (Ibid).

Spesielt det siste gjør at det vil være utfordrende å fastslå symbolsk bruk av evaluering, på annen måte enn at enkelte utsagn kan tolkes i den retningen. For eksempel sier topplederen følgende:

«Det er større og større behov for å underbygge at du skaper resultater og hva du bruker pengene til. Så det ligger i trenden, eller i utvikling av samfunnet vårt at man skal være flinkere til å vise til resultater og utvikling».

Samme respondent snakker om at hele legitimiteten ligger i det at man kan vise til at man har skapt gode resultater, og at det er det man faktisk måler på. I den forbindelse sier respondenten videre: *«forresten synes jeg at vi måler på alt for mye»*, uten at det gis noen nærmere redegjørelse for det. Om dette kan sies å peke på symbolsk bruk eller om det bare er et uttrykk for et faktisk behov for kontroll, som følge av det å styre på en armlengdes avstand, er ikke helt enkelt fastslå (Askim et. al. 2013:72).

Vi har sett at evaluering i IN er styrt av formelle prosedyrer og at det til dels består av faste sykluser. Symbolsk/rituell bruk ble forsøkt undersøkt ved å se i hvilken grad evalueringer ikke blir brukt. Eller om informantene i visse tilfeller oppfattet at evaluering ikke var nødvendig. På spørsmål om man har opplevd at det også gjennomføres evalueringer som kanskje ikke har vært så viktig å gjennomføre - svarer respondenten fra ledergruppa:

«Ja det kan jo godt være at det har vært noen av de der at man lurer på hvorfor man evaluerte denne. Av alle ting man kunne gjøre hvorfor var dette valgt, kontra ting som man kanskje hadde større behov for informasjon på».

Samtidig som det presiseres følgende:

«... men der tror jeg hvert fall de siste årene, at vi har vært ganske tydelige i premissene på hva vi velger å evaluerer, sånn at jøss har vi gjort det? (...) den tiden tror jeg er forbi».

Programansvarlig sier noe som kan peke på at man ser større nytte i formative evalueringer, men at man uansett gjennomfører sluttevaluering:

«...så gjør vi noen ganger sånne... bestiller sånne sluttevalueringer av Arena-prosjekter som vi tenker at det... ja... dette fungerte ikke så bra, at det stort sett ikke er så veldig

mye mer vi kan trekke ut av dette her, men vi må gjennomføre en sluttevaluering for at det.. ja... det er sånn vi gjør det».

Samme respondent sier følgende på spørsmål om det er noe de kunne klart seg uten:

«Nei vi kunne jo ikke det, det her vil det sikkert være litt sånn forskjellig svar på. Jeg vil nok si at det har vært veldig viktig, så vil jeg også si at det kanskje var viktigere i en tidligere fase og så lenge dette her var nytt, et område som vi strengt tatt prøvde å utforske, da trengte vi den typen, skal vi si støtte da, analysestøtte så har vi... vi føler vel kanskje at vi blir mer og mer fortrolig med det området vi skal jobbe innenfor, så det er ikke slik at disse eksterne evalueringene gir oss noen sånn stor øyeåpner».

Samtidig som at det opplyses at de har et bevisst forhold til hva som er behovene og at de tar konsekvenser av det, ved at man i det siste har:

«... begrenset litt omfanget av disse evalueringene. Prøvd å gjøre det litt mer effektivt, mindre ressursbruk som også er en konsekvens av dette. Vi trenger kanskje ikke så tunge evalueringer som vi har hatt. Vi ønsker fortsatt da, og det tror jeg er viktig for programmet å ha den systematikken. Men vi tar nok ned litt det samlede omfanget. Vi gjør det. I og med at vi egentlig vet ganske mye mer nå enn for fem år siden, enn for ti år siden. Evalueringene eller analysene var kanskje viktigere da».

Det er viktig å presisere at dette er sagt i sammenheng med det spesifikke programmet som vedkommende er ansvarlig for og ikke i sammenheng med INs evalueringsvirksomhet generelt. Likevel er det slik at behovet kanskje ikke i alle tilfeller føles like presserende, men det skal uansett utføres evaluering. Til dels fordi det er formalisert gjennom faste-sykluser, til dels fordi det eksisterer krav fra departementene. Det er viktig at dette ikke tas ut av sammenheng. Jeg kan ikke si ut i fra det empiriske materialet at det er en oppfatning i organisasjonen om at det gjennomføres for mange evalueringer, eller at de ikke tas i bruk. Informantene uttrykker snarere at det eksisterer et kontinuerlig behov for evaluering, fordi markedet alltid er i endring, nye tjenester skal utvikles, og det skal rapporteres. Kort sagt uttrykkes det at virksomheten er avhengig av fortløpende «feedback» i styring og utvikling av tjenester, og at man uten denne «feedbacken nærmest hadde styrt i blinde.

Samlet sett må det sies at det gjennom det empiriske materialet ikke er mulig å trekke slutninger om at evaluering brukes på en symbolsk/rituell måte. Det må imidlertid nevnes at det både er utfordrende å operasjonaliser, men også at det i stor grad vil dreie seg om et rent

tolkingsspørsmål, fordi instrumentene man tar i bruk gjerne fremstilles og forstås som rasjonelle verktøy av deltakerne selv, slik vi har sett Røvik argumentere for. Det at det ikke gjøres synlig gjennom empiri, betyr ikke at det ikke finnes. I følge Røvik eksisterer det en grunnleggende tvetydighet i det at organisasjoner som etterspør organisasjonsoppskrifter, (slik som evaluering) oppfatter seg selv som en rasjonell aktør som er ute etter å løse reelle problemer. Samtidig som at organisasjonsoppskrifter kan ha en selvstendig kraft, som et «tatt for gitt» symbol på det å være moderne og effektiv (Røvik 2007:54). Selv om symbolsk bruk ikke har blitt tydeliggjort gjennom dette studiet, er det ikke dermed sagt at evalueringer ikke har en slik funksjon, i tillegg til de andre bruksmåtene.

5. OPPSUMMERING OG AVSLUTTENDE BETRAKTNINGER

Innledningsvis så vi at evaluering i økende grad har blitt tatt i bruk av norske myndigheter, og at det nå er pålagt å gjennomføre gjennom Statens økonomireglement. Det eksisterer imidlertid lite kunnskap om hvorvidt og hvordan evalueringer tas i bruk.

I denne oppgaven har jeg derfor belyst hvordan evalueringer brukes i en norsk offentlig virksomhet. Analysen har tatt sikte på å kartlegge evalueringsbruk gjennom et case-studie av Innovasjon Norge, knyttet til en brukstypologi som er sentral i litteraturen på feltet. Oppgaven har søkt å svare på problemstillingen: *Hvordan tas evalueringer i bruk og hvordan bidrar evalueringer til styring og utvikling i en norsk statlig virksomhet?*

Det ble ikke utformet hypoteser som skulle testes. Studiet har snarere hatt en eksplorerende tilnærming, hvor de teoretiske konseptene har blitt undersøkt med en kartlegging av Innovasjon Norges samlede evalueringssystem og bruksformer.

5.1 ET MANGFOLDIG FENOMEN MED EMPIRISK ORIENTERING

Den første delen av resultat- og analysekapittelet tok for seg hvordan IN gjennomfører evalueringer. Her ble det presentert et system bestående av fire kunnskapskilder; kundeserviceundersøkelsen, kundeeffektundersøkelsen, SSB-undersøkelsen og til slutt det som blir karakterisert som øvrige evalueringer. Det siste er en separat undersøkelse av en individuell tjeneste.

Det eksisterer formelle krav til gjennomføring av evaluering som IN er underlagt. Evalueringstakten avgjøres dels av IN selv, dels gjennom krav fra oppdragsgiverne. Virksomheten må sies å ha god kompetanse, lang erfaring og veletablerte strategier på gjennomføring og bruk av evaluering. Evalueringsevne er med andre ord godt forankret i organisasjonen.

Selv om kun en av undersøkelsene formelt karakteriseres som evaluering, har oppgaven ikke ekskludert de andre analysene fra diskusjonen om bruk. Det ser ut til at de forskjellige undersøkelsene brukes på ulike måter. Dataene peker på at KEU og SSB-undersøkelsen har en viktig funksjon når det gjelder legitimerende effekter, mens rene evalueringer i større grad er verktøy for utvikling og læring. I sistnevnte tilfellet vil funksjonen også kunne variere ut i fra om det dreier seg om en underveisevaluering eller en sluttevaluering.

Andre del av resultat- og analysekapittelet har vist at flere av de seks brukstypene som er presentert i teorikapittelet er synlige i det empiriske materialet. Det er utfordrende å se klare mønstre i et relativt lite utvalg. Det er likevel enkelte brukstyper som virker å være fremtredende.

I teoridelen gjorde jeg det klart at taktisk bruk av evaluering ikke ville bli undersøkt. Grunnen til det er at det oppfattes som en kontroversiell bruksform, og at det derfor er utfordrende å dokumentere gjennom intervjuer. Både fordi det er forventet at informantene ikke ønsker å snakke om bruk som en ren strategisk handling, men også begrunnet i at en slik «spektakulær» form for bruk, som Foss Hansen kaller det, må ses i lys av at evalueringer er rasjonaliserte «fortellinger» om hva som fungerer (2003:89, 188). Det samme gjelder til dels evaluering som en symbolsk/rituell handling. I sistnevnte tilfellet ble det likevel gjort et forsøk på operasjonalisering ved å undersøke hvorvidt evalueringer ble gjennomført, til tross for at man ikke opplevde et reelt behov eller om man opplevde at funnene ikke ble benyttet. Det fantes lite støtte i empirien for denne bruksformen, selv om enkelte utsagn ble trukket frem. At man ikke finner empirisk støtte for symbolsk/rituell bruk kan gjerne være et uttrykk for at evaluering er «rasjonalisert». Dersom evaluering brukes på en symbolsk måte er det imidlertid ikke til hinder for at evalueringer har en viktig funksjon også på andre måter.

Når det gjelder instrumentell, konseptuell og legitimerende bruk er disse tydelig å finne i det empiriske materialet. Inntrykket er at det særlig gjelder de to siste. Når det gjelder bruk av funn versus bruk av prosess, peker dataene i retning av at det i hovedsak er funnene som tas i bruk snarere enn prosessen. Det begrunnes blant annet i at det legges stor vekt på at evaluator skal ha en tilstrekkelig faglig uavhengighet.

Det er altså tale om en blanding av bruk. I flere tilfeller kan både instrumentell, konseptuell og legitimerende bruk finnes, i et og samme resonnement, slik vi så i et eksempel i delkapittel 1.2.1. I andre tilfeller kan et utsagn både støtte bruk av funn og bruk av prosess. Det ene utelukker ikke det andre.

Det at det i samme resonnement kan uttrykkes flere former for bruk, gjør det utfordrende i praksis å skille de enkelte teoretiske konseptene fra hverandre. Det er derfor i stor grad et tolkingsspørsmål å vurdere hva som er til støtte for hva.

Brukstypologien oppleves som generell. Noe som kan tale for at det er nødvendig med en presisering av begreper. Alternativt kan det være et argument for at det er nødvendig med en empirisk orientering til fenomenet, fordi evalueringer påvirker organisasjoner på tvetydige

måter. Både Weiss og Patton har med andre ord et viktig poeng når de argumenterer for at evalueringer påvirker beslutninger på diffuse og indirekte måter, og dermed fører til oppklaring, modifisering og reorientering av programmer og tjenester.

Det kan ikke sies å være overraskende at materialet nettopp peker mot at konseptuell bruk er mest fremtredende. Konseptuell bruk kan man si er tilstede så lenge evalueringsrapporter ikke blir lagt rett i skuffen, fordi det innebærer «opplysning» og påvirkning. Gjennom empirien er det enklere å se det generelle, enn det presise. Læring og utvikling vekker også, positive konnotasjoner. Som vi så i metodekapittelet, skal man ifølge Berry (2002) ikke undervurdere betydningen av at informanter har et bestemt budskap som de ønsker og formidle, noe som kan bidra til usikre slutninger.

Dersom det er fruktbart med en empirisk tilnærming til fenomenet, kan det være interessant og trekke med et perspektiv på organisasjoner som ikke har vært adressert i teoridelen, men som etter at funnene er presentert, virker desto mer relevant. Røviks pragmatiske posisjon i mellom et rasjonelt og institusjonelt perspektiv på organisasjoner, som tar høyde for at populære ideer og oppskrifter både kan være redskaper som gir instrumentell nytte på varierende måte, samtidig som det er uttrykk for sosialt konstruerte symboler, med legitimerende effekter (Røvik 2007:53). Denne posisjonen er gjerne rettet mot å gjøre empiriske casestudier for å finne ut hva som *faktisk* skjer når redskaper, slik som evaluering blir benyttet på praksisfeltet (Røvik 2007:55).

5.2 BIDRAG, BEGRENSNING OG FORSLAG TIL VIDERE FORSKNING

Dette studiet er av eksplorerende karakter. Grunnen til det er at det i liten grad er utviklet testbare modeller om evalueringsbruk, og at det i liten grad har blitt gjennomført tidligere empiriske undersøkelser fra norsk forvaltning. Det tilsier at det på forhånd eksisterte få klare empiriske forventninger, annet enn forventninger utledet fra de teoretiske begrepene. Det gjør at en overordnet kartlegging av evalueringssystem- og bruk i IN, er et relevant «første spadestikk» til kunnskap på feltet.

Jeg mener at dette studiet er et bidrag til kunnskap om hvordan evalueringer blir brukt i en norsk sammenheng. Det bidrar i tillegg til å belyse de teoretiske konseptene, og dermed bekrefte teorien. Studiet har vist at evalueringer bidrar på diffuse og indirekte måter, og dermed inngår som en del av et samlet beslutningsgrunnlag. Samtidig har det blitt vist til at

evalueringer bidrar til at konkrete og dokumenterbare beslutninger fattes, slik som at et program blir lagt ned eller innlemmet i et annet. Studiet har også bidratt til en styrket antagelse om at ulike analyser og evalueringer er et helt sentralt virkemiddel for å legitimere virksomheten.

Det er klart at dersom dette studiet hadde blitt direkte sammenlignet med andre lignende studier, eller at det var samlet inn empiri om fenomenet fra andre caser, ville det bidratt til å kunne trekke generaliserte slutninger. Formålet var imidlertid primært kunnskapsakkumulering og teoretisk fortolkning, snarere enn et mål om å generalisere. Men som vi har sett i oppgavens metodekapittel skal man ikke undervurdere «the force of example» (Flyvbjerg 2006:228-242). Funn fra et single-case studie vil være på et konseptuelt høyere nivå enn det spesifikke caset som er undersøkt (Yin 2014:41).

Oppgaven har kartlagt kontekstuelle faktorer som også kan eksistere i andre deler av forvaltningen. I hvilken grad dette tilfellet er overførbart til andre offentlige institusjoner, må ses i lys av disse.

For det første vil jeg trekke frem det som er felles for alle statlige virksomheter. Det eksisterer et formelt krav til å analysere, evaluere og rapportere, slik at andre offentlige virksomheter også må ha systemer og prosedyrer for gjennomføring og bruk av evaluering på plass. Evaluering er ikke noe man kan velge bort. Det gjør at det er rimelig å anta at lignende funn også kan forventes andre steder. For det andre kan man forvente lignende funn der hvor virksomhetene har erfaring og kompetanse på gjennomføring av evaluering, og hvor evalueringspraksisen er institusjonalisert, ved at ansvaret er plassert i en egen avdeling, og at det er utarbeidet planer, strategier og systemer.

I dette tilfellet virker evalueringspraksisen og stå høyt på agendaen, og være godt forankret i organisasjonen. Hvorvidt dette er tilfellet andre steder, er ikke undersøkt her. Det er imidlertid rimelig å anta at organisasjonen må være av en viss størrelse for å ha lignende systemer og prosedyrer på plass.

Denne oppgaven er begrenset i den forstand at det er et single-case studie med intervjuer av relativt få nøkkelinformanter i én enkelt virksomhet. Det innebærer at man skal være varsom med å trekke slutninger om i hvilken grad funnene og tolkningene er gyldige i andre situasjoner.

En annen viktig begrensning, er at det er utfordrende å se tydelig mønstre i et datamateriale med et begrenset utvalg av informanter. Underveis i prosessen ble «superbrukerne» identifisert som programansvarlig og andre i nær tilknytning til denne. Noe som indikerer at ansvarlige på andre programmer også burde vært inkludert i studien. Det må likevel nevnes at jeg opplever å ha snakket med svært relevante nøkkelinformanter, og at disse er blant de helt sentrale i organisasjonen når det gjelder evaluering. Jeg opplevde også i liten grad at informantene kom med motstridende informasjon.

Som nevnt er dette et første «spadestikk» til mer kunnskap på feltet. Det finnes forsvinnende få empiriske undersøkelser på fenomenet, det til tross for at det er et sentralt styringsverktøy for statlige virksomheter.

Den nåværende kunnskapsstatusen tilsier at videre forskning på feltet er viktig. Gitt denne oppgavens begrensning, kan det både være komparative case-studier av et større utvalg av virksomheter og bredere kvantitative undersøkelser, som tar for seg større deler av forvaltningen. Det gjelder både på departements- og direktoratnivå. Det mest sentrale kan likevel sies å være undersøkelser av underliggende direktorater, da det i Askim .et al. (2013) har blitt trukket frem at disse aktørene ofte selv er oppdragsgivere.

Fremtidig forskning kan også ta for seg et utvalg av spesifikke evalueringsrapporter å se hvordan de har blitt brukt. Det kan legge til rette for å undersøke hvordan bakgrunnsvariabler slik som lederforankring, organisatorisk beredskap, kvalitet etc. har en innvirkning på om, og hvordan, evalueringer blir brukt.

6. REFERANSELISTE

- Aberbach, J. D., & Rockman, B. A. (2002). «Conducting and Coding Elite Interviews», *Political Science & Politics APSC*, 35(04), 673-676.
- Alkin, M. C., & Taut, S. M. (2003). «Unbundling evaluation use», *Studies in Educational Evaluation*, 29 (1), 1–12.
- Andersen, S. S. (1997). *Case-studier og generalisering: Forskningsstrategi og design*. Bergen: Fagbokforlaget.
- Andersen, S. S. (2006). «Aktiv informantintervjuing», *Norsk statsvitenskapelig tidsskrift*, 22(3), 278-298
- Askim, J., Døving, E., & Johnsen, Å. «Evalueringspraksis i den norske statsforvaltningen 2005-2011», I Halvorsen, A., Madsen, E. L., Jentoft, N. (Red). *Evaluering: Tradisjoner, praksis, mangfold* (72-87) Bergen: Fagbokforlaget
- Berry, J. M. (2002). «Validity and Reliability Issues In Elite Interviewing», *Political Science & Politics APSC*, 35(04), 679-682.
- Bryman, A. (2001). *Social Research Methods*. Oxford: Oxford University Press.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). «A Garbage Can Model of Organizational Choice», *Administrative Science Quarterly*, 17(1), 1-25.
- Cohen, W. M., & Levinthal, D. A. (1990). «Absorptive Capacity: A New Perspective on Learning and Innovation», *Administrative Science Quarterly*, 35(1), 128-152.
- Cousins, J. B. & Leithwood, K. A. (1986). «Current Empirical Research on Evaluation Utilization», *Review of Educational Research*, 56(3) 331-364.
- Dagsavisen.no (2016). «Kritisert for feil og mangler». Hentet fra <http://www.dagsavisen.no/oslo/kritisert-for-feil-og-mangler-1.670433>
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford Business Books, Stanford University Press.
- Dahler- Larsen, P. (2013). «Evaluering i Norge set Udenfra». I Halvorsen, A., Madsen, E. L., Jentoft, N. (Red). *Evaluering: Tradisjoner, praksis, mangfold* (43-57) Bergen: Fagbokforlaget
- DFØ (2010) «Veileder resultatmåling: Mål- og resultatstyring i staten». Hentet fra http://www.dfo.no/Documents/FOA/publikasjoner/veiledere/Maal_og_resultatstyring_i_statens.pdf. Sist besøkt 29.05.2016.
- DFØ (2011) «Strategisk og systematisk bruk av evaluering i styringen». Hentet fra http://www.dfo.no/Documents/FOA/publikasjoner/veiledere/Strategisk_og_systematisk_bruk_av_evaluering_i_styringen.pdf. Sist besøkt 25.05.2016
- Eliassen, K. A. & Sitter, N. (2008). *Understanding public management*. London: Sage.
- Ervik M. R. Horrigmo Å. M. J. (2015). «Trenger ikke flere men bedre evalueringer», *Stat og styring*, 2015(3), 2-5.
- Evalueringsportalen (2015). «Om Evalueringsportalen». Hentet fra <http://evalueringsportalen.no/om-evalueringsportalen>. Sist besøkt 25.05.2015

Finansdepartementet (2005). «Veileder til gjennomføring av evaluering». Hentet fra https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/veileder_til_gjennomfoering_av_evalueringer.pdf. Sist besøkt 25.05.2016

Finansdepartementet (2003). «Reglement for økonomistyring i staten». Hentet fra https://www.regjeringen.no/globalassets/upload/FIN/Vedlegg/okstyring/Reglement_for_ekonomistyring_i_staten.pdf. Sist besøkt 25.05.2016

Flyvbjerg, B. (2006). *Five Misunderstandings About Case-Study Research*. *Qualitative Inquiry*, 12(2), 219-245.

Foss Hansen, H. (2003). *Evaluering i staten: Kontroll, læring eller forandring?* Fredriskberg: Samfundslitteratur.

Furre, H., Horrigmo, Å. M. J. (2013). «Nærhet og distanse. Om forholdet mellom bestiller og utfører av evaluering». I Halvorsen, A., Madsen, E. L., Jentoft, N. (Red). *Evaluering: Tradisjoner, praksis, mangfold* (180-194) Bergen: Fagbokforlaget

George, A. L. & Bennet A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press

Gerring, J. (2004). «What Is a Case Study and What Is It Good for? » *American Political Science Review*, 98(2), 341-354.

Gildemyn, M. (2014). «Understanding the Influence of Independent Civil Society Monitoring and Evaluation at the District Level: A Case Study of Ghana», *American Journal of Evaluation*, 35(4), 507-524.

Halvorsen A. Madsen E.L. (2013). «Styring og læring gjennom evaluering», I Halvorsen, A., Madsen, E. L., Jentoft, N. (Red). *Evaluering: Tradisjoner, praksis, mangfold* (15-27) Bergen: Fagbokforlaget

Hellevik, O. (2009). *Forskningsmetode i sosiologi og statsvitenskap* (7. utg.). Oslo: Universitetsforlaget.

Højlund, S. (2014). «Evaluation use in the organizational context – changing focus to improve theory», *Evaluation*, 20(1), 26-43.

Henry, G. T. & Mark, M. M. (2003). «Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions», *American Journal of Evaluation*, 24(3) 293-314.

Herbert J. L. (2014). «Researching Evaluation Influence: A Review of the Literature», *Evaluation Review*, 38(5), 388-419.

Howlett, M., Ramesh M & Perl, A. (2009). *Studying public policy. Policy cycles and policy subsystems*. (3 utg.) New York: Oxford University Press.

Innovasjon Norge (2015). «Årsrapport 2015». Hentet fra <http://www.innovasjonnorge.no/aarsrapport/2015/full.html>. Sist besøkt 25.05.2016

Innovasjon Norge (2015). «Årlig rapportering til oppdragsgivere». Hentet fra <http://www.innovasjonnorge.no/PageFiles/3262657/Innovasjon%20Norges%20%C3%85rlig%20rapportering%20til%20oppdragsgivere%202015.pdf>. Sist besøkt 25.05.2016

Innovasjon Norge (2016). «Norwegian Centres of Expertise (NCE)» Hentet fra <http://www.innovasjonnorge.no/nce/>. Sist besøkt 25.05.2016.

- Innovasjon Norge (2016). «Arena programmet». Hentet fra <http://www.innovasjon norge.no/nce/>. Sist besøkt 25.05.2016
- Johnson, R. B. (1998). «Toward a theoretical model of evaluation utilization», *Evaluation and Program Planning*, 21(1) 93–110.
- King, G., Keohane, R. O. & Verba, S. (1994). *Designing Social Inquiry. Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press
- Kirkhart, K. E. (2000). «Reconceptualizing Evaluation Use: An Integrated Theory of Influence», *New Directions for Evaluation*, 2000(88), 5-23.
- Knorr, K. D. (1977) «Policymakers Use of Social Science Knowledge: Symbolic or Instrumental?» I C. H. Weiss (Red.), *Using social research in public policy making* (165-182) Lexington, KY: Lexington Books
- Levinton, L. C. & Hughes E. F. (1981). «Research on the Utilization of Evaluation: A Review and Synthesis», *Evaluation Review*, 5(4) 525-548.
- Madsen, E. L., Alsos, G. A., Clausen, T. H., Ljungren, E. (2013) «Evaluering av næringsrettede virkemidler- et spørsmål om addisjonalitet?» I Halvorsen, A., Madsen, E. L., Jentoft, N. (Red). *Evaluering: Tradisjoner, praksis, mangfold* (58-71) Bergen: Fagbokforlaget
- March, J. G., & Olsen, J. P. (1984). «The New Institutionalism: Organizational Factors in Political Life», *Rev American Political Science Review*, 78(03), 734-749
- Mark, M. M., & Henry, G. T. (2004). «The Mechanisms and Outcomes of Evaluation Influence», *Evaluation*, 10(1), 35-57.
- Meld. St. 22 (2011-2012). «Verktøy for vekst – om Innovasjon Norge og SIVA SF». Hentet fra <https://www.regjeringen.no/contentassets/e574980451b94330b833dc7da563f4f9/no/pdfs/stm201120120022000dddpdfs.pdf>. Sist besøkt 25.05.2016
- Meltevik, S. Hansen, T. B., Brastad, B., Borthen, H. (2010) «Hurtigvirkende kur- Kundeeffektundersøkelse av bedrifter som mottok støtte i 2009», Oxford Research AS
- Norad (2016) «Evalueringsavdelingen» Hentet fra <https://www.norad.no/om-norad/ansatte/evaluering-savdelingen/>. Sist besøkt. 03.04.2016.
- Nordbakken, L. P. (2015) «Virker Innovasjon Norge?» Civita Notat, 28/2015. Hentet fra https://www.civita.no/?download=40782&file=Civita-notat_28_2015.pdf. Sist besøkt 25.05.2016
- Norsk Evalueringsforening (2015). «2015 blir evalueringsåret». Hentet fra <http://norskevalueringsforening.no/2014/08/2015-blir-evaluering-saret/>. Sist besøkt 25.05.2016
- NRK (2016). «Fremtidens arbeid». <https://tv.nrk.no/serie/debatten>.
- NSD (2016). «Må prosjektet meldes?» Hentet fra <http://www.nsd.uib.no/personvern/meldeplikt/>. Sist besøkt 13.04.2016.
- Patton, M. Q. (1988). «The Evaluator's responsibility for utilization», *Evaluation Practice*, 9(2), 5-24.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4.utg.) Thousand Oaks, CA: Sage Publications

- Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D. & Blyth, D. A. (1977). «In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research», i C. H. Weiss (Red.), *Using social research in public policy making* (141-163) Lexington, KY: Lexington Books
- Repstad, P. (1998). *Mellom nærhet og distanse: Kvalitative metoder i samfunnsfag* (3. utg.). Oslo: Universitetsforlaget.
- Rich, R. F. (1977). «Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding», i C. H. Weiss (Red.), *Using social research in public policy making* (199-230) Lexington, KY: Lexington Books.
- Ryen, Anne (2011) «Ethics and Qualitative Research», i D. Silverman (Red.), *Qualitative Research* (3. Utg.) (416-438) Sage Publications
- Røvik, K. A. (1998). *Moderne organisasjoner: Trender i organisasjonstenkingen ved tusenårsskiftet*. Bergen-Sandviken: Fagboklaget.
- Røvik, K. A. (2007). *Trender og translasjoner: Ideer som former det 21. århundredets organisasjon*. Oslo. Universitetsforlaget
- Scott, R. W. (2014). «Crafting an Analytic Framework I: Three Pillars of Institutions», i *Institutions and Organizations: Ideas, Interests, and Identity* (4 utg.). Sage Publications.
- Scott, R. W. & Davis, G. F. (2007). «Organizations as Rational Systems», I Scott, R. W. & Davis, G F, *Organizations and organizing: Rational, natural and open systems Perspectives*. Pearsons Education, Prentice Hall
- Statskonsult (1997). «Evalueringsspraksis i departementa», Oslo: Statskonsult, Rapport 1997:13. Hentet fra <https://www.difi.no/sites/difino/files/R97-13.PDF>. Sist besøkt 25.05.2016
- Thagaard, T. (2013). *Systematikk og innlevelse: En innføring i kvalitativ metode* (4. utg.). Bergen: Fagbokforlaget
- Utdanningsdirektoratet (2016) «Organisasjonskart». Hentet fra <http://www.udir.no/Stottemeny/Om-direktoratet/Organisasjon/organisasjonskart1/>. Sist besøkt 03.04.2016
- Vedung, E. (2009). *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur
- Vedung, E. (2010). «Four Waves of Evaluation Diffusion», *Evaluation*, 16(3) 263-277.
- Weiss, C. H. (1977). «Introduction», i C. H. Weiss (Red.), *Using social research in public policy making* (s. 1-22) Lexington, KY: Lexington Books.
- Weiss, C. H. (1980). «Knowledge Creep and Decision Accretion» *Knowledge: Creation, Diffusion, Utilization*, 1(3) 381-404.
- Weiss C. H. (1988). «If Program Decisions Hinged Only On Information: A Response to Patton», *American Journal of Evaluation*, 9(3), 15-28.
- Weiss, C. H. (1999). «The Interface between Evaluation and Public Policy», *Evaluation*, 5(4) 468-486.
- Yin, R. K. (2014). *Case Study Research: Design and Methods* (5. utg.) Los Angeles:Sage Publications

7. VEDLEGG

Interjuguide

1. Hvordan går dere frem når dere bestiller en evaluering?

- Hvordan identifiseres behov for hva som skal evalueres? Hvordan velges tema?
- Hvilke rutiner har dere etablert for bestilling og gjennomføring? Finnes det for eksempel retningslinjer i form av strategi/policydokumenter?

2. Hvorfor gjennomfører dere evalueringer?

- Hvilke formelle krav til gjennomføring av evalueringer er dere underlagt?
- I hvor stor grad bestemmer dere selv hva som skal evalueres?
- Har dere noen gang gjennomført evalueringer som du mente ikke var viktig/nødvendig?

3. Hvordan bruker dere evalueringer?

- Har du konkrete eksempler?
- Har du eksempler på at beslutninger er tatt som du tror ikke hadde blitt tatt uten evaluering?
- Har evalueringsbruk betydning for Innovasjon Norges resultater/måloppnåelse?
- Har du opplevd at en evaluering ikke har blitt tatt i bruk?
- Har dere foretatt store endringer på bakgrunn av kunnskap fra en evalueringsrapport?
- Har dere avsluttet, lagt ned et program, tiltak eller ordning basert på anbefalinger fra en evaluering?
- Hvor viktig er evalueringer, kunne dere klart dere uten?

4. Hvordan spres/formidles informasjon fra en evaluering ut i organisasjonens avdelinger?

- Hvordan sørger man for at de tas i bruk? Hvem har ansvaret?
- Hvordan følger dere opp?

5. Hvor viktig er læringsperspektivet for dere ved gjennomføring av evaluering?

Bidrar evalueringer til kunnskapsutvikling på området den omfatter?

6. Hva slags rolle har dere under gjennomføringen?

- Hvor tett er involvering mellom dere og oppdragstaker i en evalueringsprosess?
- Er det viktig for dere å være en deltagende part underveis?
- Er dere med på å justere eller endre underveis?

7. Oppfatter du evaluering som en del av et kontrollapparat fra eiere?

- Opplevde du at det var ubehagelig at virksomheten/tiltaket ble evaluert?
- Her evalueringer noen gang satt virksomhetens i et dårlig lys?
- Har dere vært uenig i resultatene av en evaluering?

8. Har evalueringsresultater en bredere interesse en til styringsformål, feks i omdømmebygging?

- Kommuniseres evalueringsresultater til eksterne interessenter/og eiere?
- Kommuniseres gode resultater/funn til media? For eksempel i form av pressemeldinger?
- Er det vanlig å referer til evalueringer i rapportering til eiere?
- Har du noen gang vist til en evalueringer for å imøtegå kritikk?

9. Andre kommentarer?