



Arbeidsnotat
2018:10

Impact i anvendt forskning: begrepsavklaring og praksis

Magnus Gulbrandsen og Gunnar Sivertsen

NIFU

Arbeidsnotat
2018:10

Impact i anvendt forskning: begrepsavklaring og praksis

Magnus Gulbrandsen og Gunnar Sivertsen

Arbeidsnotat 2018:10

Utgitt av Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU)
Adresse Postboks 2815 Tøyen, 0608 Oslo. Besøksadresse: Økernveien 9, 0653 Oslo.

Prosjektnr. 20986

Oppdragsgiver Forskningsinstituttene fellesarena (FFA)
Adresse Middelthunsgate 27, 0305 Oslo

Fotomontasje NIFU

ISBN 978-82-327-0365-4
ISSN 1894-8200 (online)



Copyright NIFU: CC BY-NC 4.0

www.nifu.no

Forord

Formålet med dette arbeidsnotatet er å drøfte to spørsmål: Hva menes med begrepet impact, og hvordan passer det for anvendt forskning? Arbeidet har vært finansiert av Forskningsinstituttene fellesarena FFA. Magnus Gulbrandsen har vært prosjektleder og hatt hovedansvaret for skrivingen av notatet. Gunnar Sivertsen har bidratt særlig til kapittel 2 og 5 samt til en rekke diskusjoner av temaene som tas opp.

Oslo, 16. oktober 2018

Espen Solberg
Forskningsleder

Innhold

Sammendrag	7
1 Innledning.....	9
2 Begrepsavklaring: impact og anvendt forskning.....	12
2.1 Impact-begrepet: nytte i vid forstand	12
2.2 Hva er nytt med impact?.....	15
2.3 Anvendt forskning i en ny tid.....	17
2.4 Passer impact for anvendt forskning?.....	22
3 Evaluering av impact.....	24
3.1 Hvorfor evaluere impact?.....	24
3.2 Kvantitative målinger.....	26
3.3 Metodeproblemer	27
3.4 Kvalitative/casebaserte evalueringer	30
3.5 Målingenes politiske natur	32
3.6 Oppsummering	34
4 Evalueringer av impact i noen europeiske forskningsinstitutter.....	36
4.1 Fraunhofer	36
4.2 INRA.....	38
4.3 VTT.....	40
4.4 Sammenlignende oppsummering.....	41
5 Oppsummering og prinsipper for gode vurderinger av impact	43
Referanser	47

Sammendrag

Begrepet 'impact' handler om relasjonen mellom forskning og samfunn. På svensk brukes ordet 'samverkan', på dansk 'gavn' av forskning, på norsk 'samfunnseffekter' eller 'nytte' av forskning. Det engelske ordet 'impact' vinner også fram på norsk, særlig etter at Norges forskningsråd de siste tre årene har inkludert dokumentasjon og vurdering av 'impact' i fagevalueringer og i evalueringen av de samfunnsvitenskapelige instituttene.

I engelskspråklige land brukes 'impact' primært i forbindelse med evalueringer av grunnforskning og av universiteter. I denne rapporten diskuterer vi om begrepet også er relevant for anvendt forskning og for forskningsinstituttene som har dette som en hovedaktivitet. Svaret er et tydelig ja med noen modifikasjoner. På den positive siden argumenterer vi at 'impact' mer enn 'innovasjon' og lignende kan bidra til en utvidet forståelse for instituttenes formål og virksomhet ved å inkludere forskningens virkninger på flere samfunnsområder inkludert miljø, helse og kultur. Dette kan for eksempel gi bedre former for evaluering. Men impact-begrepet kan også virke begrensende ved å målbære en lineær forståelse av forholdet mellom forskning og samfunnseffekt, noe som ikke har støtte i litteraturen. Impact skiller seg også fra relevans-begrepet. Mens relevans-begrepet handler om hvilke problemstillinger forskningen adresserer, handler impact om hvilke resultater forskningen bidrar til å skape. Relevans kan derfor være bedre egnet til å vurdere forskningen ex ante, mens impact er svært vanskelig å vurdere før forskningsaktiviteten er i gang – eller lenge etterpå.

Evaluering av impact kan ha flere formål. Revisjon eller ettersyn, ofte kalt summativ evaluering, har ofte som mål å vurdere hvor effektiv en bestemt forskningsinnsats er. Strategi og styring, ofte kalt formativ evaluering, er som regel mer rettet mot læring og forbedring i forskningsorganisasjoner og virkemiddelapparatet. Lobbyvirksomhet er evalueringer som har som hovedhensikt å informere beslutningstakere eller allmenheten om verdien av forskning slik at den kan få fortsatt eller økt støtte. Til sist kan formålet være på en bedre måte å forstå, forklare og fortolke måten forskning tas i bruk på i videre forstand. Det er ulike evalueringsmetoder som egner seg best for de ulike formålene.

Uansett formål er det ikke enkelt å evaluere 'impact', særlig ikke hvis man forsøker å kvantifisere effekter eller å knytte en bestemt samfunnseffekt til et bestemt forutgående forskningsprosjekt på et bestemt sted. Generelle problemer for alle målinger er attribusjon (hvor mye av effekten skyldes forskningen og hvor mye skyldes andre forhold), tidsperspektiv (impact kan ta svært lang tid), kvalitet på data og indikatorer, og perspektiv på impactprosessen (samfunnssiden av relasjonen er i liten grad studert). All forskning er avhengig av annen forskning og er i gjensidig avhengighet med samfunn og næringsliv. Samfunnsrelasjonen må forstås som en interaksjon, som et komplekst samspill over tid. Det kan være mer relevant å evaluere 'impact' på organisasjonsnivå enn på individnivå. Vi henter europeiske eksempler på at forskningsorganisasjoners 'impact' evalueres etter deres formål og samfunnsansvar.

Rapporten avsluttes med blant andre disse rådene til evalueringer av forskningsinstitutter spesielt og som utgangspunkt for diskusjon og refleksjon generelt:

- Legg et bredt samfunnsperspektiv til grunn for forståelsen av impact – det handler ikke bare om instrumentelle og økonomiske effekter.
- Se ikke bare etter positive eksempler på virkninger av forskning.
- Vurder hvordan det er tilrettelagt for samspillet mellom forskning og samfunn.
- Ta hensyn til formål og målsettinger for det som evalueres: Forskningen ved instituttene utføres i organisasjoner med bestemte samfunnsoppdrag som de kan evalueres i forhold til.
- Følg god praksis ved utvikling og bruk av indikatorer. Det finnes ingen perfekte indikatorer for impact, og aspekter som robusthet, transparens og (skjulte) incentiver er viktige å diskutere.

Selv om impactbegrepet kan innebære noen utfordringer for anvendte miljøer, gir det også muligheter til å demonstrere hvordan fag- og institusjonsevalueringer kan gjøres på en klok måte. Her er det mange muligheter for fagmiljøene selv til å ta initiativer.

1 Innledning

Hva og hvor mye får samfunnet igjen fra forskning? Hva kjennetegner forskning som tas i bruk i bedrifter, offentlig sektor og andre steder, og hva kjennetegner forskningsmiljøene bak den? Dette er viktige og komplekse forskningspolitiske spørsmål, og diskusjonen av dem har pågått i mange årtier, sannsynligvis like lenge som vi har hatt offentlig finansiering av forskning. Spørsmålene har ingen enkle svar, og evaluerings- og målemetodene som har vært brukt, har ofte vært kontroversielle.

Mange begreper har blitt lansert for å fange opp forskningens bidrag til samfunnet: nytte, nytteverdi, relevans, avkastning, brukerinvolvering, addisjonalitet, samfunnsbidrag og en rekke andre. «Impact» er det mest populære begrepet akkurat nå, og sammen med «excellence» og «open science» representerer det de viktigste forskningspolitiske prioriteringene i EU og en rekke land:

*I am strongly convinced that the core values of Horizon 2020 and its successor have to be **Excellence, Openness and Impact**. (...) We have an obligation and an incentive to be much better at understanding and communicating the impact of what we do. Not only to ministers of finance, but to the general public! Carlos Moedas, EU-kommisjonær for forskning og innovasjon, 2015¹*

Moedas-sitatet indikerer at diskusjonen om impact har brede målgrupper. Målingene av impact skal – naturlig nok – bidra til bedre prioriteringer og beslutninger innenfor forskningssystemet. Men de skal også anskueliggjøre forskningens samfunnsmessige verdi og betydning for allmenheten og overfor aktørene som sitter aller nærmest den offentlige pengesekken. Impact innebærer en synliggjøring som er ment å røre selv det kaldeste finansdepartements-hjerte.

En spesielt viktig drivkraft for diskusjon og vurderinger av impact er det omfattende britiske evaluerings- og finansieringssystemet Research Excellence Framework (REF), som med 6-7 års mellomrom har vært gjennomført siden 1986 (tidligere under navnet RAE). Systemet har vært rettet mot å måle forskningskva-

¹ https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/embracing-era-change_en

litet ved bruk av ekspertpaneler som har gitt en tallkarakter til sentrale publikasjoner fra miljøene. Forut for den seneste evalueringen i 2014 ble REF videreutviklet til også å evaluere det som ble kalt «societal impact» av forskningen ved britiske universiteter og høyskoler. Dette året måtte fagmiljøene levere inn såkalte «impact case», tosiders beskrivelser som dokumenterte hvordan forskningen hadde gitt effekter. Som i evalueringen av forskningskvalitet ble casene gitt en karakter fra en til fire stjerner, basert blant annet på vurderinger av hvor vidtrekkende effektene var, hvor viktige de var for ulike aktører i samfunnet («significance» og «reach»), og hvor tydelig de ble fremstilt og dokumentert. 2014-resultatene om kvalitet og impact ble brukt til å fordele 15 prosent av universitetenes grunnbevilgning. I neste runde av REF i 2021 øker andelen til 20 prosent. Det å dokumentere sin nytteverdi på denne måten får med andre ord stadig større betydning også for den grunnleggende finansieringen av forskning i Storbritannia.

Også i Norge er begrepet impact nylig tatt i bruk i evalueringer av forskningsområder og –miljøer. Som regel blir dette begrepet ikke oversatt til norsk, tross anglisismen. Impactstudier er gjennomført for eksempel i Norges forskningsråds nyeste evalueringer av samfunnsvitenskap, humaniora og ulike grupper av forskningsinstitutter. Selv om ikke de norske evalueringene har noen direkte innflytelse på finansieringen av forskningen, er definisjonene og mange av metodene de samme som i REF med vekt på kortfattede casebeskrivelser etter en mal. Særlig vektlegges det at dokumentasjon av faktiske samfunnseffekter er sentralt for å bevare forskningens legitimitet og betydelige offentlige finansiering, slik også Moedas-sitatet over antyder. Noen tydelig effekt på grunnbevilgning har evalueringene imidlertid ikke.

I Sverige er ordet «impact» oversatt med «samverkan», og i Danmark diskuterer man å bruke ordet «gavn». Nederland og Flandern bruker «valorisation» som et begrep for det å skape verdier av forskning (økonomiske i snever forstand, verdier i bred forstand i andre sammenhenger). Selv om begrepsfloraen er betydelig, samles forskningspolitiske aktører likevel på møter og konferanser, bl.a. arrangert av AESIS-nettverket (Advancing and Evaluating the Societal Impact of Science), som har «impact» i navnet.

Bortsett fra at begrepet er nytt og vektleggingen av kortfattede og egenproduserte eksempler («case») er ny, er det ofte uklart om impact innebærer noe ekstra sammenlignet med nytte, relevans og andre tradisjonelle begreper fra tidlige tiders forskningspolitiske diskusjoner og evalueringer. Siden mye av diskusjonen om temaet baserer seg på det britiske evaluerings- og finansieringssystemet for universitets- og høyskolesektoren, er det også uklart om begrepet fanger de relevante dimensjonene til andre typer forskningsmiljøer som i større grad definerer

seg som anvendte eller brukerorienterte, slik som de norske forskningsinstituttene. Hovedhensikten med denne rapporten er å drøfte disse to spørsmålene – **hva menes med impact og hvordan passer det for anvendt forskning.**

Rapporten baserer seg på nyere litteratur om impact, refleksjoner fra forfatterne som begge er ledere av store prosjekter om temaet (innenfor OSIRIS og R-Quest, hvor vi bl.a. samarbeider med mange eksperter fra utlandet), samt noen beskrivelser av hvordan impact defineres og evalueres ved noen større forskningsinstitutter i andre europeiske land. Sistnevnte er basert på informasjon fra internett, samtaler og eposter med nøkkelpersoner fra instituttene, og presentasjoner fra en konferanse arrangert av organisasjonen EARTO i Brussel 1. mars 2018².

Basert på drøftingen skisserer vi i siste del av rapporten noen prinsipper for god praksis knyttet til måling og evaluering av impact med relevans for norske forhold. Hva slags evaluerings- og målesystemer er rimelige og treffsikre, hva er fallgruvene, og hvordan kan impact støttes av aktører som Norges forskningsråd i sine evalueringer og programmer? Prinsippene er basert på et evalueringsspektiv på impact – som er det aller vanligste – men vi vil også understreke at begrepet kan romme bredere perspektiver på forskningens samfunns effekter og være utgangspunkt for faglige og strategiske diskusjoner.

Vårt grunnsyn er at det er legitimt å diskutere forskningens samfunnsbidrag, og at det er legitimt å lage systemer som søker å evaluere og understøtte dette. Siden impact er særdeles vanskelig å måle, er det innenfor fagfeltet «studier av forskning» ofte betydelig bekymring knyttet til indikatorer og målemetoder som treffer dårlig eller som i liten grad tar hensyn til forskningens egenart. Leiden-manifestet for ansvarlig bruk av bibliometri i evalueringer av forskning³ er et tydelig uttrykk for dette. Det finnes ingen «gullstandard» for evaluering av impact og ingen indikatorer som peker seg ut som svært gode⁴. Feltet er fragmentert, det er ofte vanskelig å sammenligne ulike evalueringer, og metodeproblemene er så betydelige at seriøse evalueringssystemer har gått bort fra å basere seg på indikatorer alene (Donovan, 2011). Vi vil derfor fokusere på refleksjoner rundt godt funderte prinsipper snarere enn større forslag til evalueringssopplegg. Samtidig er det viktig at refleksjonene og prinsippene kan egne seg til å skape forskningspolitisk debatt i Norge med utgangspunkt i landets mangfoldige forskningssystem og tradisjonen for å vurdere og stimulere samfunnsnyten av forskningen, særlig i instituttsektoren.

² Se <http://www.earto.eu/news/detail/article/earto-publishes-new-economic-footprint-study-the-impact-of-9-rtos-in-2015-2016.html>

³ <http://www.leidenmanifesto.org/>

⁴ <https://www.sv.uio.no/tik/english/research/projects/osiris/osirisblog/there%E2%80%99s-no-silver-bullet-for-measuring-societal-im.html>

2 Begrepsavklaring: impact og anvendt forskning

I dette kapitlet presenterer vi først hva som menes med impact og hvordan det skiller seg fra lignende begreper som nytte og relevans. Deretter tar vi for oss «anvendt forskning» og hvordan denne aktiviteten både er kommet under press og har opplevd fornyet interesse. Til sist diskuterer vi om evalueringer og forståelse av impact passer sammen med en moderne forståelse av anvendt forskning. Kapitlet er drøftende og reflekterende – det finnes ingen etablert definisjon av noen av begrepene som alle vil være enig i. Målet med diskusjonen av det flertydige begrepet impact er å bidra til en avklaring av hvordan forskningens samfunnsrelasjon best kan forstås, vurderes, dokumenteres og eventuelt måles, avhengig av formålet med og gjenstanden for evalueringen.

2.1 Impact-begrepet: nytte i vid forstand

Det britiske REF-evalueringssystemet bruker følgende definisjon av impact:

An effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia

Definisjonen innebærer på den ene siden en bred forståelse av impact: det dreier seg om virkninger av forskning innenfor en rekke ulike samfunnsområder, inkludert miljø, helse og kultur. Perspektivet på forskningens samfunnsrelasjon utvides, og begrepet blir mer ambisiøst. På den andre siden uttrykker både definisjonen og den opprinnelige betydningen av ordet impact – virkninger – en såkalt lineær forståelsesmodell for forholdet mellom forskning og samfunn. Samme ord brukes om virkningene når en asteroide treffer en planet eller når en biljardkule treffer en annen: det er en hendelse som katalyserer resultater et annet sted. Muligheten for å forstå og evaluere samspillet mellom forskning og samfunn som interaktivt og gjensidig mulighetsskapende risikerer dermed å bli borte.

Intensjonen bak det brede samfunnsperspektivet er nok først og fremst å gjøre impact til et begrep som passer til alle fagområder, inkludert samfunnsvitenskap

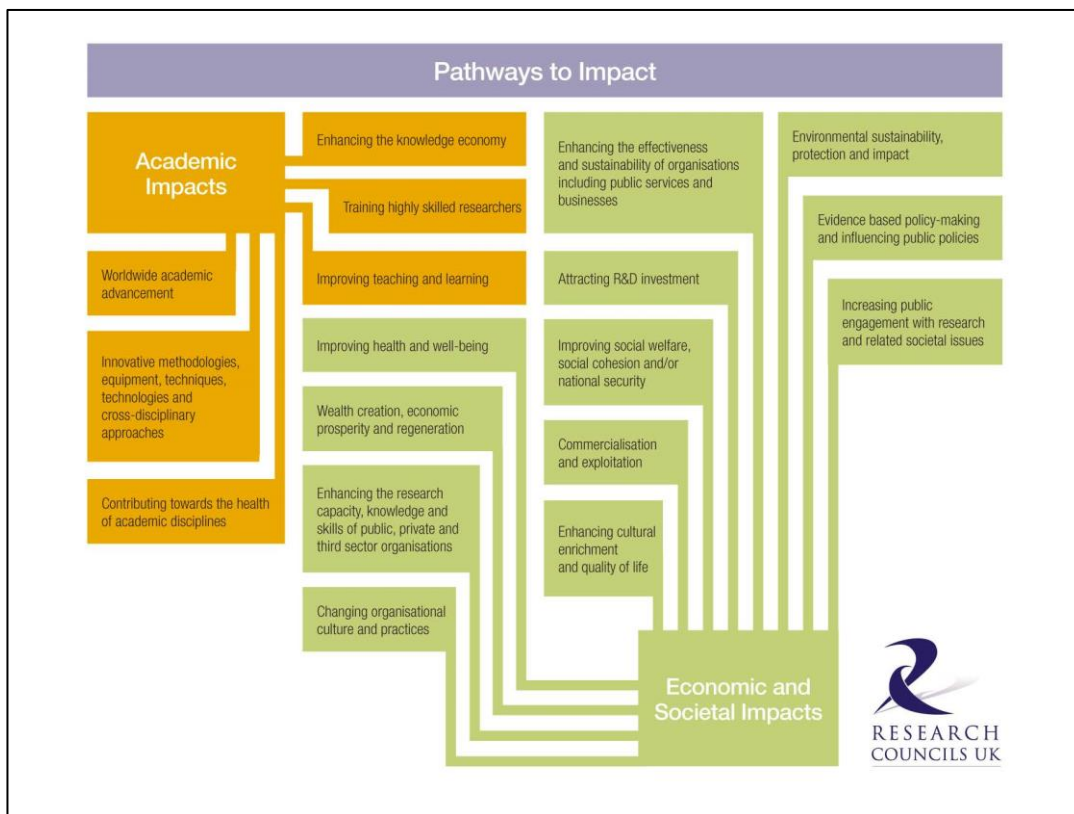
og humaniora. Bakgrunnen for dette er at begrepet ble utviklet for å dekke all forskning ved britiske universiteter, som er evalueringsobjektet til REF-systemet. Ellers ser vi at det bare er virkninger utenfor forskningssystemet som er av interesse. Interne effekter innen forskningen antas å kunne undersøkes tilfredsstillende gjennom tradisjonelle indikatorer knyttet til vitenskapelig publisering og sitering. Begrepet «societal impact» var opprinnelig ment å supplere begrepet «citation impact» da det ble introdusert i den britiske diskusjonen av forskningsevaluering for tolv år siden.

At definisjonen samtidig er begrensende ved å gjenspeile en lineær forståelsesmodell, har nok også sammenheng med at den ble utviklet for å som grunnlag for å evaluere og finansiere bare den ene siden i samhandlingen, universitetene. Slik ble det ble fokus på impact som en aktivitet blant forskerne og på dokumentasjon av at resultatene henger sammen med identifiserbar forskning.

Nyere definisjoner i faglitteraturen om impact er svært like den til REF-undersøkelsen, noe som sannsynligvis skyldes at mye av faglitteraturen kommer fra miljøer som arbeider direkte med forskningspolitikk og evaluering. Dette er ett eksempel:

“The direct and indirect ‘influence’ of research or its ‘effect on’ an individual, a community, or society as a whole, including benefits to our economic, social, human and natural capital” (Harland & O’Connor 2015)

Forfatterne bruker hermetegn rundt «influence» og «effect on», kanskje fordi dette er komplekse og flertydige begreper som kan ha ulik betydning for ulike interessenter i forskningspolitikken, men kanskje også for å markere avstand til en lineær forståelsesmodell. Effekt kan for eksempel oppfattes som noe direkte og instrumentelt – hvordan forskningens resultater og ideer bidrar til å løse et problem – eller som noe mye mer indirekte knyttet til forskningens lærings-, politiske eller symbolske betydning. Uansett forståelsesmodell er også denne definisjonen svært bred og vil være mulig å ta som utgangspunkt for å vurdere alle typer forskning fra alle fagområder. Figur 1 viser et eksempel på hvordan de britiske forskningsrådene tenker rundt mangfoldet i impactbegrepet.



Figur 1. Veier til impact slik de britiske forskningsrådene definerer det.

Kilde: www.ukri.org

Begge definisjonene nevnt i dette avsnittet åpner for at en kan se på impact som noe negativt – at forskning leder til mindre velferd, skader på miljøet eller annet. Begrepet «grimpact» er lansert for å dekke dette. I en del tilfeller kan det være slik at forskning er bra for noen grupper i samfunnet men dårlig for andre (se Gaunand m.fl. 2012 og Joly m.fl. 2015). Vurderinger av impact innebærer dermed ofte en (som regel implisitt) vurdering av hva slags verdier forskningen bidrar til å dekke, svekke eller styrke (Bozeman & Sarewitz 2011). Det er imidlertid svært få evalueringsmetoder som søker å belyse problematiske effekter av forskning, og diskusjonen om «grimpact» foregår mest på sidelinjen.

Ofte beskrives impact som langsiktige ringvirkninger eller siste ledd i en kjede (Penfield m.fl. 2014). Fra forskningen kommer det «output» i form av publikasjoner, doktorgradskandidater, patenter og annet. Dette gir ulike «outcomes», slik som når en publikasjon siteres eller tas i bruk utenfor forskningen eller en doktorgradskandidat finner arbeid utenfor academia. Dette vil igjen kunne gi ulike former for impact, sisteordenseffekter slik definisjonene over uttrykker. Skillet mellom disse trinnene kan synes kunstig og lineært, men det har betydning særlig for evaluering av forskning fordi det uttrykker hva som kan evalueres og hvilke indikatorer som kan brukes.

Siden det ofte er snakk om lange tidsforløp, vil impact også kunne være noe som endrer seg betydelig over tid. Et tenkt eksempel kan være et nytt forskningsbasert plastprodukt som løser ulike behov men som senere skaper forurensning i havet. Et mye brukt historisk eksempel er legemiddelet Thalidomid, mest knyttet til en av tidenes største legemiddelskandaler ved at det i mange tilfeller ga store skader på fosteret hos de gravide kvinnene middelet skulle hjelpe å bli kvitt morgenkvalme. Senere har Thalidomid vist seg å være lovende i behandlingen av noen kreftformer (Penfield m.fl. 2014). Tidsperspektiv og valg av måletidspunkt kan med andre ord ha stor betydning for resultatet (både størrelsen og valøren på effektene), noe vi kommer tilbake til i kapittel 3.

2.2 Hva er nytt med impact?

Det finnes altså mange begreper som uttrykker hva forskningen bidrar til utenfor forskningen, slik som (samfunns)effekter som omhandler faktiske resultater av forskning, og relevans som betegner at det er en tematisk eller annen forbindelse mellom forskningen og et behov i samfunnet. Relevans handler på mange måter om å styre forskningen i bestemte retninger, selv om forskningens iboende usikkerhet og ofte uventede resultater (det er mange eksempler på at svært nyttig forskning kommer fra uventede områder) gir noen begrensninger.

For de fleste i Norge er impact et nytt begrep som bærer med seg mindre bagasje enn for eksempel innovasjon, kommersialisering og nytte, som for mange benevner økonomiske og instrumentelle effekter av forskning. Mange vil se på de sistnevnte som for snevre dersom dette er det eneste som vektlegges av effekter, og som vist, representerer impact nytte i mye videre forstand. Selv om nytte er et velkjent ord også fra forskningspolitikken, står impact likevel for noe annerledes enn det en har sett i tidligere begreper og evalueringer – på tre ulike måter.

For det første representerer begrepet større ambisjoner med evalueringene. Tidligere nøyde myndigheter og forskningsfinansierende organisasjoner seg med å se på de første trinnene i den ofte lange prosessen som leder til impact – altså det som en på engelsk kaller outputs. Det handlet om å gjøre relativt enkle undersøkelser blant oppdragsgivere og andre brukere, se på siteringer til publikasjoner og patenter, og se på samarbeid mellom forskningsmiljøer og nærings- og samfunnsliv. Dette ga indikasjoner på relevans og på et potensial for nytte ved at forskerne var i kontakt med miljøer som kunne ha bruk for forskningen. Nyere målinger av impact går mye lenger og representerer mer eller mindre systematiske og omfattende forsøk på å synliggjøre og/eller tallfeste de totale effektene av forskning på mange ulike samfunnsområder. I motsetning til outputs kalles dette for outcomes eller impact (noen skiller mellom disse, andre ikke).

Impact er for det andre – som vist over – et begrep som defineres slik at alle fagområder og typer av forskning skal oppleve det som relevant. Slik det er tatt i bruk f.eks. av Norges forskningsråd, formuleres det som en forventning til alle forskningsmiljøer som får offentlig støtte. Det er derfor mye videre enn en tradisjonell oppfatning av at noe er nyttig eller gir nytte. I Norge har vi sett at impact-case er brukt som metode i evaluering av naturvitenskapelig, samfunnsvitenskapelig og humanistisk forskning, og knyttet til miljøer i universiteter, høyskoler og forskningsinstitutter. Også næringslivet er blitt opptatt av impact, og evaluering av innsats i og effekter av forskning, utvikling og innovasjon er blitt et satsingsområde for store konsulent- og revisjonsselskaper.

Erfaringene fra REF i Storbritannia og fra den norske evalueringen av humaniora tyder på at forskere i miljøer som har følt seg lite truffet av «innovasjon» og lignende begreper, i en rekke tilfeller har opplevd impact som mer presist og også mer motiverende. Forskere med en sterk faglig profil som av ulike årsaker ikke passer inn i excellence-begrepet, har kunnet styrke sin posisjon i fagmiljøet med gode tydelige impact-eksempler, noe som personer i Research England opplyser er en villet og ønsket effekt av evaluerings- og finansieringssystemet i seg selv. En tilsvarende erfaring i Norge fikk man særlig med Humaniora-evalueringen (2016-17), hvor fagmiljøene fikk anledning til å dokumentere samfunnseffekter som ikke direkte kunne knyttes til innovasjon eller samfunnsøkonomisk gevinst. Disse er blitt presentert grundig på mange av lærestedenes nettsider. Også fra andre evalueringer ser vi at impact-casene i etterkant blir brukt for å markedsføre betydningen av forskningen.

For det tredje innebærer impact i større grad (høye) forventninger til forskningen, kanskje også en ny tilnærming til å uttrykke og styre forventningene. Her er det et tydelig skille til «relevans». Begrepet relevans uttrykker at forskningen bør strebe etter å adressere viktige problemstillinger i samfunnet eller bestemte gruppers behov for forskningsbasert kunnskap. Dette kan en vurdere på tidspunktet hvor forskningen finansieres, ut fra en tematisk vurdering eller ved å se på om aktivitetene involverer ulike organisasjoner som skal ta forskningen i bruk. Relevans kan dessuten stilles som et rimelig krav til enkeltprosjekter og enkeltforskere innenfor mange finansieringsordninger med anvendt orientering. Om dette faktisk leder til at forskningen får ulike effekter, ble kanskje tatt mer for gitt tidligere, eller ble vurdert som noe som i stor grad er utenfor forskningsmiljøenes kontroll og derfor mindre relevant å ta med i evalueringer.

Her er impact dermed svært forskjellig fra relevans. Det er vanskelig å vurdere impact på et tidspunkt hvor forskningen er på skisse- og søknadsstadiet, ex ante, og noen vil hevde at det er urimelig å stille krav om impact til enkeltprosjekter på finansieringstidspunktet. Relevans uttrykker et potensial og i mange tilfeller krav til enkeltprosjekter, men innebærer også at ikke all forskning (uavhengig av hvor

relevant den er) vil lede til noen vesentlige resultater. Impact representerer på den annen side tydeligere forventninger om at på lang sikt så bør det komme «et eller annet» ut av all forskning. I alle fall hvis forskning defineres som noe som angår en gruppe, en avdeling eller et institutt.

Impact er tett koblet sammen med måling og evaluering. Begrepet dukker oftest opp etter at forskningen er utført, mens «relevans» tradisjonelt er brukt for å velge ut forskning som skal få støtte i form av prosjekter eller andre bevilgninger. Det er likevel tegn til at impact også må redegjøres for på søknadsstadiet i en rekke land og hos en rekke forskningsfinansierende organisasjoner, selv om det varierer hvor mye dette faktisk teller som utvelgelseskriterium (se Langfeldt & Scordato 2016).

Oppsummeringene og refleksjonene i denne delen viser at impact er et mangefasettert begrep som nok kan fortolkes på ulike måter. På den ene siden kan diskusjonen om impact bidra til at mange andre effekter av forskning vektlegges enn bare økonomiske ringvirkninger. For de fleste forskningsmiljøer vil dette innebære mer realistiske og kanskje mer motiverende forventninger – motiverende i betydningen at det oppleves som meningsfullt å svare på evalueringene. På den annen side er impact også i en del tilfeller brukt for å finne ut hva som er den mest effektive forskningen i form av den innsatsen som gir mest igjen per krone. Om impact betraktes i hovedsak som et styringsverktøy eller som et uttrykk for å formidle brede samfunnsvirkninger av forskning, kommer mye an på hva slags evalueringemetoder som legges til grunn. Dette kommer vi tilbake til i kapittel 3.

2.3 Anvendt forskning i en ny tid

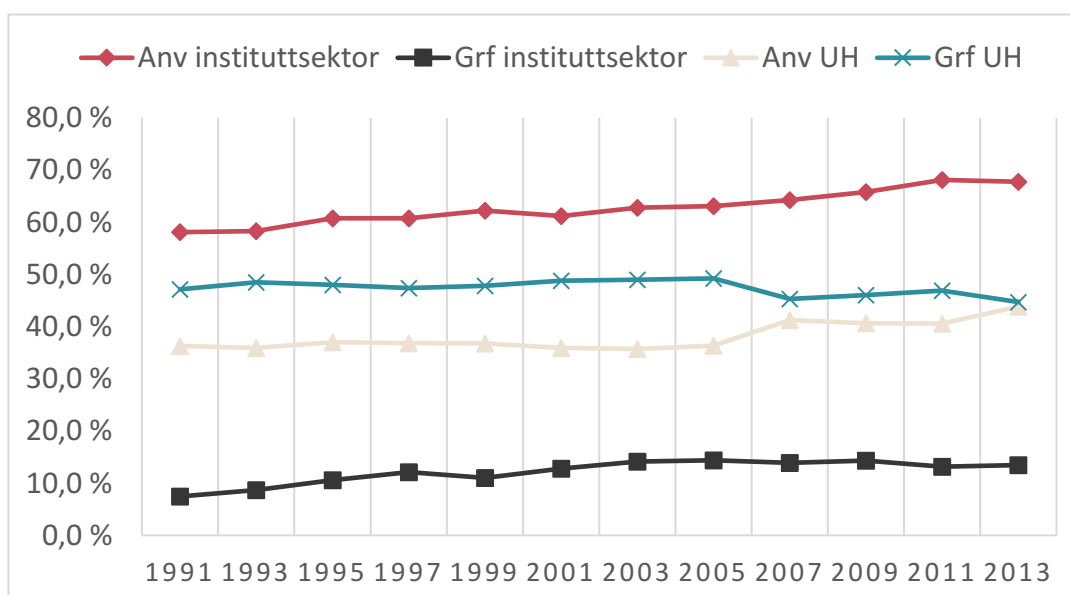
De vanligst brukte definisjonene av ulike typer forskning finnes i Frascati-manualen, et OECD-dokument som beskriver prosedyrer for innsamling av statistikk om forsknings- og utviklingsarbeid (FoU). Her er følgende definisjoner relevante:

Grunnforskning er eksperimentell eller teoretisk virksomhet som primært utføres for å skaffe til veie ny kunnskap om det underliggende grunnlag for fenomener og observerbare fakta, uten sikte på spesiell anvendelse eller bruk.

Anvendt forskning er også virksomhet av original karakter som utføres for å skaffe til veie ny kunnskap. Anvendt forskning er imidlertid primært rettet mot bestemte praktiske mål eller anvendelser.

Frascati-manualen tar med andre ord utgangspunkt i hensikten med forskningsaktiviteten. For mange forskere er det andre forhold som kan være relevante (Gulbrandsen & Langfeldt 2004). Eksempler er om forskningen har teoretiske snarere enn bare empiriske ambisjoner (selv om den kan ha et overordnet praktisk siktemål) og om den ønsker å utvikle fagområdet gjennom bestemte former for vitenskapelig publisering.

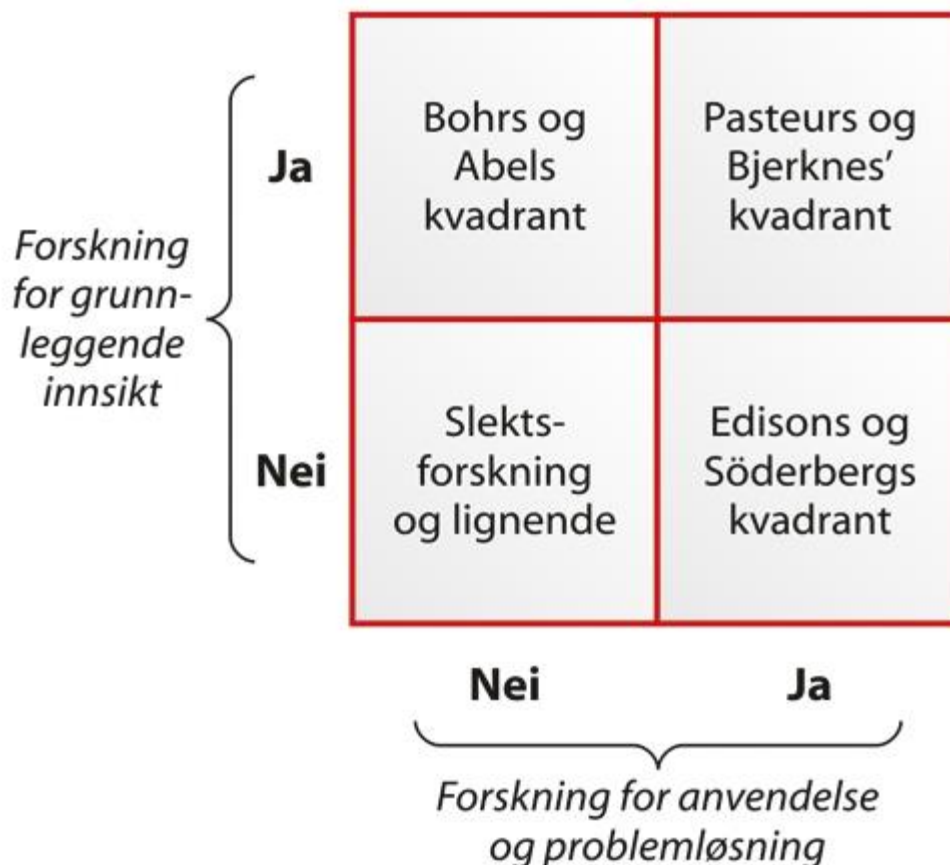
Vi vil trekke fram fire sentrale aspekter ved hvordan anvendt forskning har utviklet seg som begrep de siste tiårene. For det første brukes det mindre i statistiske målinger. Dersom en ser på den internasjonale FoU-statistikken er det mange land som har stoppet å rapportere fordeling mellom grunn- og anvendt forskning i ulike sektorer. I Norge gjøres dette fortsatt: her viser målingene at nivået på anvendt forskning har vært økende i både forskningsinstitutter og i universitets- og høyskolesektoren (UH) – se Figur 2. I instituttene har andelen anvendt forskning økt fra litt under 60 prosent til litt under 70 prosent siden begynnelsen på 1990-tallet. I samme periode har denne andelen økt fra 36 prosent til 45 prosent i UH-sektoren, og er i siste måling på omtrent samme nivå som grunnforskningsandelen. Anvendt forskning er med andre ord en dominerende og sannsynligvis identitetsskapende aktivitet i instituttene, men det er også en sentral aktivitet i UH-sektoren.



Figur 2. Anvendt forskning og grunnforskning i instituttsektoren og UH-sektoren. Kilde: NIFU/FoU-statistikken.

For det andre er skillet mellom grunnforskning og anvendt forskning blitt mye kritisert, og det er foreslått en rekke begreper som skal erstatte dem eller som kan uttrykke en kombinasjon av dem eller en annen form for forskningsvirksomhet. Skarpest kritikk finnes hos Stokes (1997), som mener at skillet gir et ideologisk, organisatorisk og praktisk skille – grunnforskning i universiteter, anvendt forskning andre steder – som gjør at samfunnet får mye mindre igjen for forskningen enn det burde fordi perspektivet blir overdrevent lineært. Hans ideal er det han kaller «Pasteurs kvadrant», en inkluderende og samfunnsengasjert form for forskning (i fagmiljøer, hos enkeltpersoner) som kombinerer en søken etter grunnleggende innsikt med et mål om å løse praktiske problemer. Figur 3 viser Stokes'

idealtyper av forskning med noen norske alternativer. Den er hentet fra en tidligere Stortingsmelding om forskning, som argumenterer for at mange av de mest kjente eksemplene på god norsk forskning (objektorientert programmering, havforskning, meteorologi) passer inn i Pasteurs del av skjemaet.



Figur 3. Pasteurs kvadrant med norske eksempler.

Bildet er hentet fra Stortingsmelding 20 (2004-05), Vilje til forskning.

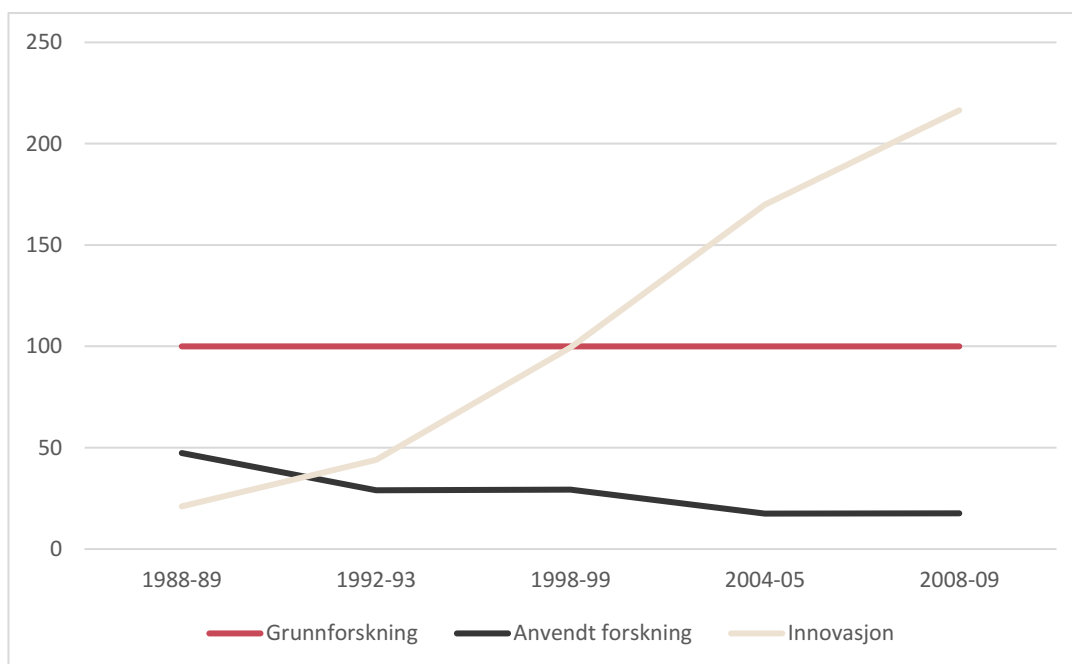
Også «Mode 2»-begrepet (Gibbons m.fl. 1994; Nowotny m.fl. 2001) uttrykker noe lignende, hvor forfatterne postuler at fremtidens kunnskapsproduksjon i mye større grad vil skje i samspill på tvers av sektorer og med et overordnet praktisk siktemål. Denne formen for kunnskapsproduksjon vil vokse mye mer enn «Mode 1», som refererer til tradisjonell disiplinorientert grunnforskning. En rekke empiriske undersøkelser av forskningsmiljøer viser dessuten at de beste miljøene som regel er engasjerte i flere typer forskning, med andre ord at de scorer høyt både på «grunnleggende» og «anvendte» dimensjoner (Gulbrandsen 2000).

Selv om flere av disse begrepene er omdiskuterte, ikke minst hvordan de er blitt tatt i bruk av ulike forskningspolitiske aktører – uttrykker de nok endrede forventninger til forskningen. Anvendte miljøer forventes å publisere og på andre måter

bidra til fagets utvikling, og grunnforskningsmiljøer kan ikke underslå seg forventninger om nytte – særlig ikke hvis det defineres så bredt som i diskusjonene om impact. Det britiske evaluerings- og finanseringssystemet REF definerer at i dokumentasjonen fra fagmiljøene, utgjør vurderingene av kvalitet en indikasjon på «Mode 1», mens impactcasene er en indikasjon på «Mode 2». Denne forståelsen av begrepene er nok ikke helt i tråd med deres opprinnelige definisjon, som er kritisk til nettopp et slikt skille.

Imidlertid er selve begrepet anvendt forskning – for det tredje – blitt borte fra forskningspolitikken. Vi har gått gjennom de fem siste Stortingsmeldingene om forskning (fra 1988-89 til 2008-09) og søkt etter relevante begreper for ulike typer aktiviteter. Dette er oppsummert i Figur 4. Her ser vi at «Anvendt» og relaterte ord er så å si blitt borte annet enn i referanser til definisjonene i Frascati-manualen. «Grunnforskning», «grunnleggende forskning» og lignende omtales relativt mye i alle fem meldingene. I alle fall gjøres det forsøk på å diskutere det, presisere betydningen av det og hvordan det skal fremmes.

Selv om en skal være forsiktig med å fortolke slike enkle tellinger, kan det synes som om begrepet «innovasjon» har erstattet «anvendt forskning». Vi går ikke mer inn på det her, men vi bemerker at innovasjon og anvendt forskning er svært forskjellige aktiviteter, og at hvis endret språkbruk reflekterer en forventning om at anvendt forskning alltid eller ofte skal lede til innovasjon, er det meget problematisk. Det kan hevdes at det er utfordrende at den største aktiviteten i det offentlige forskningssystemet – anvendt forskning – enten ikke omtales eller omtales på nye og potensielt diskutabile måter i sentrale politiske dokumenter.



Figur 4. Antall ganger ulike begreper er brukt i stortingsmeldinger om forskning, relativt til hverandre (begreper for grunnforskning satt til 100, en score på 50 for et av de andre begrepene betyr da at det nevnes halvparten så ofte).

På den annen side, og for det fjerde, kan det hevdes at anvendt forskning er blitt revitalisert gjennom andre begreper og diskusjoner fra de siste par års forskningspolitikk. Sarewitz' (2016) sterke kritikk av forskningssystemet, særlig den delen av det som foregår uten kontakt med brukere og anvendte perspektiver, er et godt eksempel. Sarewitz, en kjent forsker innenfor vitenskapsstudier, uttrykte blant annet at forskningen blir bedre av slik kontakt:

“Scientific knowledge advances most rapidly (...) when it is steered to solve problems — especially those related to technological innovation (...) When science is not steered to solve such problems, it tends to go off half-cocked in ways that can be highly detrimental to science itself. (...) Science will be made more reliable and more valuable for society today not by being protected from societal influences but instead by being brought, carefully and appropriately, into a direct, open, and intimate relationship with those influences.”

Disse perspektivene er ikke ukjente i norsk sammenheng, og Sarewitz var en av hovedinnlederne på Norges forskningsråds store forskningspolitiske konferanse i 2017.

Et på mange måter lignende budskap, om enn fra et helt annet perspektiv, er Mazzucatos (2018) argument om at Europa bør styrke sin forskning rettet mot store, tydelig definerte samfunns-utfordringer, såkalte missions. Mazzucato, som er innovasjonsøkonom, argumenterer for at ulike «kombinasjoner av grunnforskning og anvendt forskning» er nødvendig for å løse nåtidens komplekse samfunnsproblemer, og at disse aktivitetene organiseres under et klart definert mål med klart definerte milepæler. Sentralt i en slik «mission-oriented» forskningspolitikk er tanken om å bruke prinsipper fra tidligere tiders store teknologi-prosjekter (slik som månelandingen) på mer komplekse samfunns-problemer. Hvordan dette skal gjøres er ikke helt klart, men EU har i flere sammenhenger uttrykt at det skal satses mer på denne formen for finansiering av forskning. Mazzucatos rapport diskuterer tre eksempler – CO2-frie byer, fjerne plast i havet og minske byrden ved demens – og hvordan grunnleggende og anvendt forskning kan kombineres og styres innenfor hvert av disse. I EUs neste rammeprogram, kalt Horizon Europe og presentert i juni 2018, er missions en sentral del.

I sum kan det hevdes at anvendt forskning som begrep kanskje er blitt borte fordi hele forsknings-systemet styres i mer anvendt retning, til dels med nye overskrifter og mål. Statistikken fra Norge viser i alle fall at anvendt forskning som aktivitet lever i beste velgående både i institutter, universiteter og høyskoler. En kan kanskje se på forskningspolitikken som en balansegang eller pendelbevegelse mellom et syn om at forskningen bør være mest mulig fri fra direkte koblinger til nytte og relevans og et syn om at forskningen overveiende bør være rammet inn

av konkrete samfunnsbehov. Resultatet blir to hovedtyper av offentlige forskningsmiljøer. Samtidig er impact-begrepet definert og brukt for at det skal være relevant for begge typene av forskning.

2.4 Passer impact for anvendt forskning?

Siden impact i hovedsak har sine røtter i diskusjoner om hvordan universitetsforskningen kan evalueres og styres, er det likevel relevant å spørre om det passer for anvendt forskning. Det umiddelbare svaret er ja, og som regel er det den motsatte bekymringen som uttrykkes, altså at det er vanskeligere å dokumentere effektene av grunnleggende og langsiktig forskning. Den anvendte er lettere å evaluere fordi det er kortere tid fra FoU-aktiviteter til effekt, noe som også gjør det enklere å vurdere attribusjon, altså hvor mye av effekten som skyldes forskning (se kapittel 3).

Men sett fra et anvendt perspektiv er det nok noen skjær i sjøen. For det første gir impact muligens et mindre tydelig skille mellom ulike typer forskning på finansieringstidspunktet, noe som kan være en ulempe for miljøer som har profilert seg sterkt på å være relevante for bestemte brukere. Impact er et løsere begrep enn relevans, og selv om det finnes forskningsprogrammer som krever at søkerne sier noe om impact, gis denne delen av søknaden som regel liten eller ingen vekt i vurderingen (jf. Langfeldt & Scordato 2016). Med andre ord kan introduksjonen av impact i styringen av forskning gjøre at behovet for å styre etter relevans bli mindre, og som vist, er begrepene ikke de samme.

For det andre er tanken om at en skal dokumentere impact gjennom enkeltcase hvor en effekt kobles direkte til et prosjekt eller en publikasjon, potensielt mer problematisk for anvendte miljøer enn andre. I anvendte miljøer arbeider personalet som regel sammen i prosjekter, og det er summen av en rekke aktiviteter over tid som gir effekter. Selv om det nok er slik også for grunnforskning, vil nok mange si at en likevel lettere kan plukke ut de viktigste bidragene ved å se på sentrale publikasjoner. Hvordan noen europeiske instituttmiljøer har løst denne utfordringen, presenteres i kapittel 4.

Det kan for det tredje legges til at REF-undersøkelsens rangering av impact-casene (fra en til fire stjerner) kanskje også er fjernere fra kvalitetsbegrepet i anvendt forskning. Der hvor grunnforskningen kan vurderes på en skala med «excellence» i den ene enden (og dermed lar seg rangere numerisk), er anvendt forskning noe som er hensiktsmessig for formålet eller ikke, uavhengig av eventuelle vurderinger av hvor fremragende den er (se Gulbrandsen 2000 for en diskusjon).

Det fundamentale budskapet i dette kapitlet er at de ulike måtene å definere anvendt forskning på også representerer ulike tilnærminger til å styre og evaluere impact. Den klassiske definisjonen er FoU-aktiviteter som har et bestemt siktemål.

For forskningsfinansierende organisasjoner betyr det at de bør velge ut prosjekter med en direkte kobling til dette siktemålet, og deretter evaluere de aktuelle fagmiljøene ut fra om de bidrar til det. Nyere perspektiver knyttet til store samfunnsutfordringer («missions») som en overordnet styringsheuristikk åpner for mindre direkte måter å vurdere relevans på. I praksis er det likevel usikkert om dette vil være tilstrekkelig for forskningsrådene og lignende organisasjoner.

I tillegg kan det nevnes at forskningsinstitutter og andre fagmiljøer som i stor grad vektlegger anvendt forskning, ofte definerer sine mål ut fra mer enn bare om FoU-aktiviteten deres er grunnleggende eller ikke. Som regel har de andre og ofte mer sektorspesifikke mål, f.eks. om å bidra til kunnskap om eller utvikling av bestemte næringer eller politikkområder. I Lov om universiteter og høyskoler er den såkalte «tredje oppgave», samfunnsbidraget, uttrykt så generelt at det kan passe i enhver fagsammenheng. Anvendte forskningsinstitutter er opprettet med mer spesifikke formål og vil ha strategiske målsettinger som angår samfunnsbidraget. Regelen om at en virksomhet bør evalueres i forhold til hva den vil oppnå, kan anvendes mye mer konkret på forskningsinstitutter.

Nåtidens definisjon av impact, som åpner for en rekke ulike samfunnseffekter ved siden av praktisk-økonomiske, kan vise seg å være godt tilpasset anvendte forskningsinstitutter. Samtidig er det ofte uttrykt stor bekymring for at impact-evalueringer i siste instans skal styre forskningsmidlene mot fagfelt og -miljøer som lettere kan demonstrere sin nytte, ikke minst i økonomisk forstand. Dette avhenger imidlertid av hva slags metoder for evaluering av impact som velges, som er tema for neste kapittel.

3 Evaluering av impact

Dette kapitlet beskriver ulike generelle tilnæringer til å evaluere samfunnseffekter av forskning – som regel kalt *impact* i dagens forskningspolitikk. Vi skiller grovt mellom kvantitative og casebaserte tilnæringer, og i tillegg diskuterer vi hensikten med evalueringer og hva de sentrale problemene ved ulike metoder er. Kapitlet er en vesentlig omskrevet og utvidet versjon av en kort artikkel som ble publisert i Indikatorrapporten 2017 (Gulbrandsen 2017) og en blogg publisert på nettsidene til forskningssenteret OSIRIS⁵.

Det kan legges til at litteraturen om impact er svært omfattende. Det er mye grålitteratur – rapporter og lignende produsert for en enkelt evaluering – men også mye akademisk litteratur. Tidsskriftet *Research Evaluation* er det mest sentrale for generelle diskusjoner av impact og metoder, men mye av litteraturen finnes i spesialiserte tidsskrifter innenfor store forskningsområder som medisin, landbruk og teknologi. Et godt eksempel på hvor omfattende litteraturen er, er en gjennomgang av nylige studier av impact av landbruksforskning (Weißhuhn m.fl. 2018). Her fant forskerne nesten 200 vitenskapelige artikler som utelukkende handlet om impact av landbruksforskning og som var publisert i perioden 2008-2016. Med et lengre tidsperspektiv – i dette kapitlet viser vi at målinger av impact strekker seg minst 50 år tilbake i tid – snakker vi sannsynligvis om et firesifret antall akademiske studier for dette fagfeltet alene.

3.1 Hvorfor evaluere impact?

Det kan grovt skilles mellom fire ulike hensikter med å evaluere impact (modifisert fra Penfield m.fl. 2014):

1. **Revisjon**, ofte kalt *summativ* evaluering eller «accountability». Innbakt i mer og mer av offentlig finansiering ligger det en systematisk oppfølging av hva pengene har gått til i videste forstand. Evalueringen av impact er dermed knyttet til systemer som er utbredt i all offentlig forvaltning. I de fleste

⁵ <https://www.sv.uio.no/tik/english/research/projects/osiris/osirisblog/measuring-impact.html>

tilfeller vil det dreie seg om kost-nytte-analyser, vurdering av sosioøkonomiske effekter eller «addisjonalitet» – altså hva pengene har gitt i tillegg til det som en kunne forventet uten offentlig støtte til FoU (eller annet). Positive resultater fra evalueringene bidrar til å rettferdiggjøre innsatsen. Denne typen evaluering brukes jevnlig i Norge for ulike offentlige virkemidler, ofte med en vurdering i etterkant av om det aktuelle virkemidlet skal fortsette eller ikke. Den nylige evalueringen av SkatteFunn-ordningen er ett eksempel (Benedictow m.fl. 2018).

2. **Oversikt og styring**, ofte kalt *formativ* evaluering. Her skal forskningsmiljøene eller finansieringssystemet gis et innblikk i hva slags resultater aktivitetene deres har skapt, slik at de involverte miljøene på ulike måter kan bli enda bedre til å skape samfunnseffekter (eller kvalitet eller annet). Målet er læring mer enn kontroll, og i noen tilfeller er det fagmiljøene selv som setter i gang slike evalueringer. De norske fag- og institusjonsevalueringene passer inn i denne tradisjonen, samt en del større evalueringer av virkemidler slik som den nylige av ordningen med Sentre for forskningsdrevet innovasjon (Damvad 2018).
3. **Lobbyvirksomhet**, altså med hovedhensikt å informere beslutningstakere eller allmenheten om verdien av forskningen slik at den kan få fortsatt eller økt støtte. Ulike aktører har interesse av å demonstrere impact på en systematisk måte, som er et kraftfullt politisk argument i et system som ønsker å være evidensbasert (Donovan 2011). Bak slike evalueringer ligger også en antakelse om at beslutningstakere ønsker å gi forskningsmidler til miljøer, sektorer og institusjoner hvor de opplever at de får «nye igjen for pengene». Den europeiske interesseorganisasjonen for teknisk-naturvitenskapelige forskningsinstitutter EARTOs «economic footprint study» er et godt eksempel på dette (Bilsen m.fl. 2018). Vi bruker lobbyvirksomhet som et nøytralt ord her, ikke som noe som betegner en lyssky eller illegitim aktivitet.
4. **Innsikt**. Det er grunn til å understreke at det å forstå, forklare og fortolke måten forskning tas i bruk på og i videre forstand relasjonen mellom forskning og samfunn, er en interessant og kompleks problemstilling for mange forskere. Spesialiserte fagfelt som vitenskaps- og teknologistudier (STS), innovasjonsstudier og studier av forsknings-, innovasjons- og teknologipolitikk (STI) berører alle på ulike vis nøkkelspørsmål rundt impact. Forskningen på disse feltene bruker omtrent alle tilgjengelige metoder, men ofte advarer forskerne mot enkle former for måling og evaluering. Et godt eksempel er den britiske rapporten *The Metric Tide* som er skrevet av en stor gruppe forskere og praktikere (Wilsdon m.fl. 2015). Samtidig er en rekke av de nyeste evalueringsmetodene sterkt influert av fagfeltenes perspektiver.

Et eksempel er ASIRPA-systemet som er utviklet for å evaluere samfunnseffektene av forskningen til det store franske primærnæringsinstituttet INRA, og som er basert på innsikter fra innovasjonsstudier og STS. ASIRPA beskrives i mer detalj i kapittel 4.

Poenget med denne distinksjonen er at hensikten med evalueringene avgjør hvilke metoder som bør benyttes. Kost-nytte-analyser eller indikatorer kan være hensiktsmessige både i summative evalueringer og for lobbyvirksomhet, men de gir få verktøy for organisasjonsutvikling og svært begrenset innsikt i prosessen som leder til impact. Det finnes også evalueringssystemer som søker å forene flere hensikter. Det beste eksempelet er britiske REF, som både ønsker å utvikle universitetene og å bidra til «accountability», ikke minst ved at evalueringene er direkte koblet til finansieringen av sektoren. Dette komplekse eller hybride utgangspunktet kan være en årsak til at ikke andre land har importert REF-tilnærmingen (jf. Sivertsen 2017). Koblingen av to formål – evaluering og finansiering – gir manglende fleksibilitet i metodevalg og store kostnader.

3.2 Kvantitative målinger

Den vanligste metoden for å evaluere impact handler om å kvantifisere de (mulige) økonomiske effektene av investeringer i FoU. Effektene av både offentlig- og privatfinansiert FoU kan belyses ved å se på komplekse indikatorer som innovasjon, vekst, omsetning, sysselsetting og produktivitet. I bunn og grunn studeres samfunnseffektene her i form av et matematisk uttrykk for sammenhengen mellom «input» (som regel penger brukt til FoU eller spesielle satsinger) og en eller annen indikator på ønsket «output». Slike økonometriske effektstudier er mest brukt i summative evalueringer og i mange former for lobbyvirksomhet.

Videre er det to hovedtilnærminger i slike effektstudier. Den første innebærer bruk av ulike typer offentlige databaser eller registre. Målingen innebærer da å se på relasjonen mellom indikatorer for forskning på den ene siden og effekter på den andre siden. Eksempler er undersøkelser av hvordan og i hvilken grad kommersielt vellykkede patenter baserer seg på publisert forskning, og i hvor stor grad bedrifter som mottar bestemte typer offentlig støtte til FoU kan vise til bedre økonomiske resultater en viss tid etterpå. For sistnevnte kan en sammenligne med bedrifter som ikke har mottatt den aktuelle FoU-støtten men som ellers ligner. Dette er f.eks. hovedtilnærmingen i evalueringen av næringsdepartementets virkemidler (Fjærli m.fl. 2016).

Den andre hovedtilnærmingen baserer seg på en mer aktiv innsamling av data særlig gjennom spørreskjemaer, noen ganger kun for evalueringens skyld, andre ganger med bredere formål. Her kan bedrifter eller andre organisasjoner spørres om erfaringer med bruk av forskning, samarbeid med eksterne forskningsmiljøer

og annet. Den store innovasjonsundersøkelsen (Community Innovation Survey, CIS) i Norge og andre land er det mest sentrale eksempelet på denne tilnærmingen. Spørreskjemaer og tidsseriedata er spesielt relevante for å vurdere effektene av forskning innenfor enkeltforetak, men kan også brukes til å estimere videre samfunnseffekter.

Felles for de kvantitative tilnærmingene er at de ofte finner svært høye tall for avkastningen på investeringer i forskning, ofte 20 prosent eller mer, for bedriftene som investerer i eller mottar finansiering til FoU. Mange undersøkelser finner enda høyere tall på avkastningen for samfunnet, gjerne 50-100 prosent eller enda mer.

Den nylige økonometriske undersøkelsen blant ni store europeiske teknologiske forskningsinstitutter (Bilsen m.fl. 2018; SINTEF var ett av instituttene) er et godt eksempel. Her fant evaluatørene at en arbeidsplass finansiert ved et forskningsinstitutt ga fire arbeidsplasser andre steder, og at myndighetene fikk tre euro tilbake (i form av økte skatteinntekter og annet) for hver euro de investerte i instituttene i form av grunnbevilgning. Slike tall er imponerende, men også faglig omdiskuterte på grunn av de store metodologiske utfordringene og usikre resultatene ved denne typen målinger av impact.

3.3 Metodeproblemer

Som nevnt i avsnitt 2.1, er impact-begrepet bredt i perspektivet på samfunnseffekter, men snevert ved å bygge på en lineær forståelsesmodell for samspillet mellom samfunn og forskning. Den som ønsker å vise at en bestemt aktivitet i forskningen får virkning et annet sted i samfunnet, står overfor en rekke metodeproblemer som skal omtales her. Et eksempel på en alternativ tilnærming hvor man unngår disse problemene, er å evaluere hvordan man har tilrettelagt for interaksjon mellom forskere og samfunn mer enn å måle resultatene av interaksjonen.

Attribusjon er en helt sentral utfordring ved evaluering av impact. Forskning leder vanligvis ikke til effekter på egen hånd eller automatisk. Det er kombinasjonen av FoU og en rekke andre faktorer – sosiale, organisatoriske, politiske og ofte ikke direkte knyttet til forskningen i seg selv – som utgjør forskjellen for en bedrift eller for ulike sektorer i samfunnet som skole, helse og politikkutforming. Utfordringen er å finne ut hvor mye av «æren» som skal gis, eller tilskrives, til forskningen eller til de individene og miljøene som en ønsker å evaluere. Hvis andre faktorer som er nødvendige for å skape effekter, ikke inkluderes i evalueringene, er det lett å få et feilaktig inntrykk av at impact beror mer eller mindre på forskningens bidrag og aktiviteter alene.

For å ta et populærkulturelt eksempel: i svært mange intervjuer har filmregissøren George Lucas uttrykt at forskningen til Joseph Campbell innenfor fagfeltet

komparativ religionshistorie var helt essensiell for skapelsen av Star Wars-universet. Campbell skrev overbevisende og eksplisitt om at verdens store fortellinger hadde en felles struktur, kalt monomyten, med helt spesielle stadier, hovedpersoner og bipersoner, som Lucas brukte som, mer eller mindre, blåkopi for sine skikkelser og narrativ. Men hvor mye av Star Wars skal Campbell og hans bok *The Hero with a Thousand Faces* egentlig ha æren for? Det er åpenbart at det måtte mange andre krefter til (samt at Campbell neppe tenkte seg denne bruken selv om han visstnok som eldre mann uttrykte begeistring for Star Wars). Og bør i så fall noe av æren tilfalle andre forskere som Freud, Jung og Maslow, som Campbell satte svært høyt og brukte ekstensivt i sine argumentasjonsrekker?

Dette eksempelet illustrerer også en annen hovedutfordring: det å finne en god indikator for effekten det er snakk om. Hva skal telle som «størrelsen» på Star Wars? Skal en telle salg og billettinntekter knyttet til filmene eller alt leketøy og andre produkter? Skal en vurdere videre kulturelle eller andre effekter? Noen vil kanskje si at Star Wars er et unikt – og gigantisk – eksempel hvis en aksepterer at det i det hele tatt er et tilfelle av impact av forskning.

Skjevfordelt impact og fortjeneste på innovasjon er imidlertid snarere regelen enn unntaket. Litteraturen om kommersialisering av forskning viser at de aller fleste ideer og prosjekter går med tap, men noen få har en eventyrlig avkastning på tusenvis av prosent. Dersom vi for diskusjonens skyld aksepterer noen av estimatene for innsparingene fra flerfaseteknologien i Nordsjøen (med opprinnelse i forskningsinstituttet IFE på Kjeller), har denne forskningen hatt en avkastning som betaler for nesten all norsk offentlig forskningsinnsats i et tiår. I slike tilfeller kommer konvensjonell statistikk og kost-nytte-gjennomsnittsbetraktninger til kort. Også i slike eksempler er attribusjonsproblematikken til stede. Selv om det er åpenbart at FoU i norske institutter og universiteter har bidratt til suksessen i den norske petroleumsnæringen, har jo næringen også representert et sterkt motaktsmiljø med høy kompetanse og store ressurser til å ta forskning i bruk. Slike rammebetingelser – som helt sikkert varierer mellom sektorer og samfunnsområder – gjør spørsmålet om attribusjon komplisert.

Det kan legges til at det finnes en bevegelse som aktivt ser etter alternative effektindikatorer – både til de tradisjonelle bibliometriske (vitenskapelige publikasjoner og siteringer) og økonomiske (produktivitet, innovasjoners andel av omsetningen). Under navn som «StarMetrics» og «Altmetrics» samles det f.eks. data om forskeres karriereløp og forskningens synlighet i sosiale medier. Selv om det kan være verdifullt og nødvendig med flere og alternative indikatorer, har disse likevel mange av de samme problemene som de tradisjonelle. I tillegg kommer at indikatorene er påvirkelige av aktørene innenfor sosiale medier, og at synlighet i sosiale medier bare i liten grad kan sies å uttrykke samfunnseffekter. Et alternativ

til å undersøke generell synlighet i sosiale medier, kunne være å undersøke synlighet og innflytelse i bestemte samfunnssektorer, for eksempel innflytelsen fra anvendt samfunnsforskning i sosialsektorens policy-dokumenter.

Et annet mye omtalt problem er latenstid – eller tidsforsinkelse. Ofte kan tiden mellom forskning i form av prosjektstart, nye oppdagelser eller konkrete resultater, og impact i form av at forskningen er tatt i bruk på en målbar måte, være svært lang. Systematiske undersøkelser av landbruksforskning, sannsynligvis det fagfeltet som er mest studert når det gjelder impact, indikerer at gjennomsnittstiden mellom FoU og dokumentert resultat ofte er flere tiår (Alston m.fl. 2010; se også Donovan 2011). Det omfattende evalueringssystemet ASIRPA (jf. kapittel fire) finner i sine case at gjennomsnittstiden er 19 år. Med andre ord er det i de fleste tilfeller urealistisk å forvente store effekter innenfor tidsrammen til et enkelt FoU-prosjekt. Problemet med latens vil selvsagt avhenge av hva slags effekt det er som evalueres. Tidsforsinkelse skaper uansett betydelige problemer i alle evalueringssystemer og indikatorer – og for spørsmålet om attribusjon. Jo lenger tid det tar mellom forskning og impact, jo vanskeligere blir attribusjon og jo viktigere blir andre forhold enn forskningen for å skape en effekt (Hughes & Martin 2012).

Et siste problem som er mye diskutert, handler om impactprosessens natur. Mange har argumentert for at forholdet mellom samfunnseffekter og forskning er vesentlig mer komplisert enn at sistnevnte leder til førstnevnte, slik vi også har nevnt i de to foregående kapitlene. I mange tilfeller er det behov og utfordringer i samfunnet som initierer og påvirker forskningsinnsatsen. Impact blir dermed et resultat av en gjensidig innflytelse mellom forskningen og de som bruker den, fortolker og oversetter den, overfører den og videreutvikler den. Selv om det selvsagt finnes mange eksempler på vitenskapelige gjennombrudd eller forskningsbaserte innovasjoner som leder til konkrete produkter og effekter, er impact bedre forstått som en mer kompleks, sosial og indirekte prosess enn noe som kan fanges av en enkel form for kausalitet. Prosessen kan være så indirekte at forskningen blir usynliggjort. Det finnes bl.a. en rekke studier som viser at resultater og perspektiver fra forskning «kryper» ut i samfunnet og tas i bruk i politikktutforming og for andre formål uten at brukerne er klar over hvor kunnskapen kommer fra, og hvor dokumentasjon av direct impact er svært vanskelig (Penfield m.fl. 2014). Innenfor mange fagområder, noen vil hevde alle, er det nok slik at den bredeste formen for impact kommer over tid når forskningen tas i bruk i undervisningen av studenter. Dette perspektivet finnes imidlertid sjelden i evalueringer av forskning, og er nok også mindre relevant når temaet er anvendte forskningsinstitutter – selv om personalet der også kan være engasjert i mange former for opplæring og formidling.

Denne gjennomgangen har vært rettet mot å diskutere overordnede problemer og feilkilder. For enkeltindikatorer vil det være relevant å diskutere forhold som

validitet (måler indikatoren det fenomenet den er tenkt til å måle) og reliabilitet (vil indikatoren være konsistent over tid). Hensikten med gjennomgangen her er ikke å si at evaluering av impact er umulig. Men det er metodisk svært komplisert og krever betydelig ekspertise og kjennskap til kontekstuelle forhold. Forskning er nok på mange måter en relativt unik aktivitet, særlig knyttet til dens langsiktige, usikre, indirekte og (i mange tilfeller) allmenne natur.

3.4 Kvalitative/casebaserte evalueringer

Undersøkelser av impact basert på casestudier har søkt å håndtere disse metodeutfordringene på andre måter enn de rent kvantitative metodene. Casestudier er typisk brukt for å evaluere et bestemt forskningsmiljø, fagområde, virkemiddel eller institusjon, og ofte brukes en blanding av kvalitative og kvantitative data. En hovedmetode er det som kalles «foroverskuende case». Evaluatører starter med den aktuelle forskningen og forsøker å kartlegge systematisk hva den har ledet til. Selv om metoden kan basere seg på historiske data, f.eks. ved å starte med forskning som ble utført for en tid tilbake, oppfattes impact her som noe som kan spores framover i tid. Felles for metodene som følger stien eller banen fra forskning mot impact er at de i hovedsak brukes i formative evalueringer, og de fleste av dem er utviklet av forskningsmiljøer som har vært opptatt av å belyse impact-prosessen.

Flere av de nyere metodene for å evaluere impact er basert på denne tilnærmingen. En mye brukt metode kalles Payback og ble først utviklet for å vurdere helseforskning i Storbritannia (se f.eks. Donovan & Hanney 2011). Den er senere blitt tatt i bruk i en rekke andre engelskspråklige land, særlig av medisinske forskningsråd for å evaluere forskningsprogrammer og fagmiljøer. Payback baserer seg på en systematisk kartlegging av ulike faser i prosessen som skaper samfunnsnytte, og metoden er avhengig av at det er god interaksjon med de aktuelle miljøene som skal evalueres. Noen interessante funn fra metoden er en problematisering av en enkel forståelse av sammenhengen mellom forskningskvalitet i form av «excellence» og senere impact. Metoden har avdekket en rekke eksempler på forskning med lav akademisk prestisje som har hatt store effekter bl.a. på behandling av pasienter (Buxton 2011; Donovan & Hanney 2011; Penfield m.fl. 2014). Bornmann (2011) nevner, basert på Smith (2001), feltet «apoptose» («selvmord» blant celler) som et eksempel på forskning vurdert som svært høy kvalitet men med ingen målbare effekter på helse 30 år senere. Han kontrasterer dette med «forskning på effektiviteten til ulike inkontinens-bind» som har rask og viktig impact men som ikke blir sett på som like «excellent». Relasjonen mellom excellence og impact er nok et område hvor det behøves mer kunnskap. Det kan legges til at ASIRPA-metoden er en annen sofistikert metode basert på «foroverskuende case» som beskrives i mer detalj i kapittel fire.

En annen lignende metode er SIAMPI – Social Impact Assessment Methods through (the study of) Productive Interactions – som er basert på et større EU-prosjekt (se Spaapen & van Drooge 2011).⁶ Antakelsen her er at de endelige effektene av forskning er svært vanskelige å evaluere på en meningsfull måte, og metoden går i stedet ut på en systematisk kartlegging av de ulike *samspillene* som forskningen inngår i. Her forsøker man altså å løse problemene som en lineær forståelsesmodell kan skape. Det skilles mellom direkte samspill (møter mellom forskere og brukere), indirekte samspill (gjennom publikasjoner, teknologier, gjenstander, infrastruktur og annet) og finansielt samspill (oppdrag, bidrag osv.). Metoden er tatt i bruk både for teknisk-naturvitenskapelig forskning og innenfor samfunnsvitenskap og humaniora. Tabell 1 viser eksempler på ulike former for produktiv interaksjon.

Tabell 1. Ulike former for produktiv interaksjon i SIAMPI-evalueringsmetoden. Kilde: www.siampi.eu.

Productive Interaction	Social impact	Stakeholder	Assessment tool
Direct, personal	Behavioural change	One-to-one, personal and professional networks	Interviews, focus groups
Indirect, media	Uptake, use	Different audiences	Quantitative data collection
Financial or in kind support	Collaboration	Joint projects	Annual reports, other documents

Det finnes også metoder som tar utgangspunkt i hvilke verdier forskningen retter seg mot (Bozeman & Sarewitz 2011). Utgangspunktet her er at ulike verdier, slik som et trygt samfunn, ren luft og vann og effektive transportsystemer, ofte er en viktig begrunnelse for finansiering av forskning, men at slike verdier ofte blir borte når forskningen og dens effekter skal evalueres. «Public value mapping» søker å synliggjøre og operasjonalisere slike verdier og evaluere forskningen opp mot dem. Metoden er interessant men ikke tatt bredt i bruk, selv om tankegodset har påvirket diskusjonen om evaluering mer generelt.

Alternativet til foroverskuende case er retrospektive undersøkelser som starter med en bestemt impact, oftest en ny teknologi eller et forskningsbasert produkt, og sporer bakover i tid hva slags forskning som var viktig for teknologien eller produktet, og på hvilken måte. Også dette handler om å se på impact primært som en prosess snarere enn et forhold mellom innsats og resultater (input-output). Et prosessperspektiv handler i alle tilfeller om å definere hva som er «nøkkelhendelser», kartlegge disse systematisk over tid og gjennom en vurdering av kontekst og sammenheng gjøre en meningsfull vurdering av kausalitet.

⁶ Det finnes mye informasjon på prosjektets nettside: <http://www.siampi.eu/>

Det kan legges til at impactcasene i det britiske REF-systemet ikke er like systematiske og grundige som de casebaserte evalueringemetodene nevnt i denne delen av rapporten. I REF ble alle fagmiljøer i universitetssektoren bedt om å sende inn ett eller flere eksempler på nytte. Disse impactcasene fulgte en mal på to sider som gir en kort beskrivelse av hva slags nytte det er snakk om, for hvem, samt dokumentasjon på at nytten kunne attribueres til et konkret forskningsresultat/vitenskapelige publikasjoner. Her ble det ikke fokusert på nøkkelhendelser eller andre kjennetegn ved prosessen. REF-tilnærmingen er som nevnt også tatt i bruk i Norge, selv om metoden er kritisert for i praksis bare å inkludere *eksepsjonelle og individuelle*, lett dokumenterbare, positive og instrumentelle eksempler snarere enn de mer *normale og organiserte*, kritiske, indirekte, tilfeldige og langsiktige samfunnseffektene av forskning (Sivertsen & Meijer, 2018). De norske casene bl.a. fra evalueringene av humaniora og samfunnsvitenskap er nok enda mindre systematisk gjennomført enn de britiske. Mange avviker ganske mye fra oppsettet i malen, dokumentasjonen er svært varierende, og det er ikke foretatt en vurdering av dem i etterkant. Det kan skyldes at dette var første gang metoden ble tatt i bruk i Norge, men også at hensikten med de norske fagevalueringene er annerledes enn i REF.

3.5 Målingenes politiske natur

Hindsight og Traces er to av de eldste og mest kjente systematiske evalueringene av impact. Begge er fra USA og fulgte en retrospektiv case-tilnærming, og de illustrerer godt at evalueringer foregår i en praktisk og politisk kontekst som kan ha innflytelse på utfallet.

På 1960-tallet ønsket det amerikanske forsvaret å kartlegge kostnader og effekter av ulike typer forskning, blant annet på bakgrunn av bekymring for sterkt økende FoU-utgifter og et ønske om å identifisere den mest «effektive» forskningen. Forsvaret definerte 20 av sine viktigste og mest avanserte våpenteknologier som utgangspunkt for en omfattende evaluering kalt *Hindsight*. Som i andre prosessmetoder ble det definert et sett med «hendelser» som kunne knyttes til forskning og som hadde bidratt til teknologiene. Resultatene viste at mindre enn én prosent av de rundt 700 identifiserte nøkkelhendelsene kunne klassifiseres som grunnleggende forskning. Drivkraften i prosessene var i de fleste tilfeller et identifisert praktisk behov som ble adressert av anvendt og teknologisk FoU, en type aktivitet som forsvaret ga massiv økonomisk og annen støtte til.

Resultatet hadde forskningspolitisk sprengkraft, og som et direkte svar initierte det amerikanske forskningsrådet National Science Foundation (NSF) sin egen evaluering som fikk navnet *Traces*. NSF valgte som utgangspunkt fem sivile teknologier som de på et eller annet tidspunkt hadde støttet (p-pillen var et av casene).

Denne evalueringen fant at om lag to av tre nøkkelhendelser kunne klassifiseres som grunnforskning – en aktivitet som NSF selv hadde et hovedansvar for.

Det må understrekes at begge prosjektene er solide fra et evalueringsfaglig ståsted og utført av anerkjente fagmiljøer fra amerikanske universiteter – så hvorfor er resultatene så divergerende? Bortsett fra at selve casene er forskjellige, er nok mye av forklaringen metodologiske valg. Hindsight inkluderte forskningshendelser slik som nye ideer presentert av forskere på møter, mens Traces knyttet hendelser primært til vitenskapelige publikasjoner. Dessuten gikk NSF-studien 50 år tilbake i tid mens forsvaret nøyde seg med 20 år.

Med andre ord forsvinner ikke de metodologiske utfordringene med impactevaluering gjennom bruk av casebaserte metoder. Små og store forskjeller i praktiske valg knyttet til tidshorisont, definisjon av forskning og nøkkelhender, samt valg av case i seg selv, kan ha enorm betydning for resultatene som Hindsight og Traces tydelig viser. Slike valg har selvsagt minst like store konsekvenser i kvantitative undersøkelser. Et felles problem er at slike metodologiske antakelser, valg og definisjoner, samt drøftingen av konsekvensene av dem, ofte vil være gjemt bort i en metodedel i evalueringsrapporter, langt bortenfor sammendrag og hovedkonklusjoner. Dette er nok en viktig forklaring på at det er relativt enkelt å stille spørsmål ved validiteten til enkeltundersøkelser som f.eks. proklamerer et bestemt tall for avkastningen av bestemte typer forskning eller offentlige virkemidler. Og det kan være lett å misbruke dem i en retning som særskilte aktører ønsker.

De to amerikanske evalueringene synliggjør impactevalueringenes politiske natur. Det å demonstrere impact gjennom sterke case og relaterte narrativ eller gjennom imponerende tall synes i økende grad å ha blitt essensielt i argumentasjonen for økonomisk støtte til forskning. EARTO skryter av tallene til sine medlemsinstitutter, og de norske universitetenes nettsider har en rekke maleriske fortellinger om impact basert på case levert inn til de store fagevalueringene av samfunnsvitenskap og humaniora.

En kan kanskje hevde at alle former for impactevaluering har fått et element av lobbyvirksomhet. Ulike aktører kan gjøre mer eller mindre bevisste og synlige metodologiske valg som leder til et resultat som støtter den typen av forskning som de er mest opptatt av, slik NSF og det amerikanske forsvaret gjorde allerede på 1960-tallet og dermed synliggjorde en sentral motsetning mellom forskning drevet av fagenes egne behov og FoU som tar utgangspunkt i praktiske utfordringer og bruker- og samfunnsbehov. Det kan legges til at både Hindsight og Traces-prosjektet er referert til i Områdegjennomgangen av Norges forskningsråd, som riktignok konkluderte med at det må satses mer på grunnforskning uten tematiske begrensninger uten at akkurat denne konklusjonen kan kobles direkte til de to amerikanske evalueringene (Områdegjennomgang 2017).

Impact er ikke et nøytralt begrep, og evalueringer av det initieres eller etterspørres av mektige aktører i forskningssystemet. Begrepet er primært brukt i forskningspolitikk – selv om mange forskere er opptatt av nytte og samfunnsbidrag, er det nok få av dem som går rundt og snakker om impact eller avkastning på forskning i sitt daglige arbeid. Dette betyr at impact kan oppleves som fremmed eller kontroversielt. Da REF-undersøkelsen ble lansert, skrev mer en 17.000 britiske forskere under på et opprop mot den – rundt en tredel av alle universitetsforskerne som ble omfattet av evalueringen. Senere har forskerne protestert mot å gjøre endringer i REF fra ekspertvurderinger til kvantitative indikatorer.

Kritikken handler også om kostnadene ved og nødvendigheten av impact-evalueringer, og de bredt anlagte evalueringssystemene har blitt sammenlignet med Frankensteins monster (Martin 2011). Kostnadene ved siste REF i 2014 er anslått til 246 millioner britiske pund (Stern, 2016), noe som mange mener heller kunne gått til å finansiere FoU. Andre hevder at en slik kostnad – to og en halv norske milliarder hvert sjette år i et stort forskningssystem som det britiske – er relativt beskjeden og helt nødvendig for å sikre forskningens legitimitet og støtte, i det minste hos de bevilgende myndighetene.

3.6 Oppsummering

Tabell 2 oppsummerer hovedtypene av evaluering av impact og hva slags aktører som typisk tar initiativet til dem. Den tradisjonelt vanligste formen for måling er kvantitativ og rettet mot å forsøke å måle hva slags effekter (output) som en bestemt forskningsinnsats (input) har hatt. Dataene kan være hentet fra ulike registre eller samlet inn gjennom særskilte spørreundersøkelser – sistnevnte kan gi mer spesifikke data men kan gi tilleggsutfordringer for reliabilitet og validitet. Ofte er metoden brukt ved evaluering av større satsinger som programmer og virkemidler, og det overordnede siktemålet er beslutning om satsingen skal fortsette eller ikke. Mange slike evalueringer er svært sofistikerte metodisk sett, men det kan likevel ofte være store utfordringer knyttet til kvaliteten på dataene og indikatorene som brukes, i tillegg til de mer generelle problemene med attribusjon, tidsperspektiv og linearitet.

Evalueringemetoder som på ulike måter analyserer eksempler på impact (case) har mange av de samme generelle utfordringene. Denne formen for måling er som regel rettet mot å gi svar på andre typer spørsmål, gjerne det som kalles formativ evaluering hvor hensikten i større grad er læring og organisasjonsutvikling. Også her er det utviklet sofistikerte tilnæringer som kan gi god innsikt i prosessen som leder til impact på en måte som kan hjelpe forskningsmiljøer eller de som finansierer forskningen. Men dersom det skal gjøres grundig nok, er metoden svært ressurskrevende.

Tabell 2. Ulike evalueringsmetoder og deres hensikter og utfordringer.

Evalueringsmetode	Hensikt	Typisk aktør	Utfordringer
Effektmåling (kost-nytte og lignende)	Summativ evaluering, lobbyvirksomhet	Departement, store forskningsorganisasjoner	Attribusjon, datakvalitet, tidsperspektiv, finne gode effektindikatorer, lineært perspektiv på impact
Altmetrics	Metodeutvikling: finne nye indikatorer for å vurdere effekter av forskning	Forskningsmiljøer, akademiske forlag, forskningsråd	Tidsperspektiv, svak validitet til indikatorene, lette å manipulere
Foroverskuende case	Formativ evaluering, i noen grad lobbyvirksomhet	Forskningsråd, store forskningsorganisasjoner	Attribusjon, tid, ressurskrevende metode, sammenligning, vanskelig å definere nøkkelhendelser
Bakoverskuende case	Formativ evaluering, vurdering av forskningens produktivitet	Forskningsråd, forskningsmiljøer	Svært ressurskrevende metode, vanskelig å definere nøkkelhendelser
Konseptuell	Metodeutvikling: finne nye tilnærminger til evaluering av impact	Forskningsråd, forskningsmiljøer	Implementering, avgrensning, potensielt ressurskrevende

Det generelle budskapet er likevel at alle tilnærminger har betydelige metodiske utfordringer, og at kvalitet i måling av impact handler om å velge den metoden som er best tilpasset det formålet som målingene har (andre forhold som analyseobjekt, tilgjengelige data osv. må også tas med). Med andre ord er designet på evalueringssystemet en viktig faktor. Vi kommer tilbake til dette i det siste kapitlet.

De nyeste litteraturgjennomgangene argumenterer for at de mest pålitelige formene for måling baserer seg på bruk av flere metoder på en gang, og at alle indikatorer krever innsiktsfull fortolkning og vurdering av eksperter (se f.eks. Wilsdon m.fl. 2015). Dette høres ut som en enkel og tilforlataelig anbefaling, men i praksis gjøres dette svært sjelden, og forskningsevalueringsfeltet kan synes fragmentert med ulike «stammer» som i liten grad kommuniserer med hverandre. Det gjør at evalueringer sjelden blir sammenlignbare og at det ikke har utviklet seg noen tydelige standarder. Større datakilder og digitalisering kan åpne opp for nye muligheter i evaluering av impact, både knyttet til forbedring av eksisterende metoder og for utvikling av friske perspektiver og tilnærminger. Det er likevel ingen grunn til å tro at dette i seg selv vil løse de metodologiske utfordringene – som jo også er knyttet til hva slags syn en har på forskningen og dens rolle i samfunnet.

4 Evalueringer av impact i noen europeiske forskningsinstitutter

Her har vi valgt å gå i noe mer detalj inn i tre evalueringssystemer som er direkte koblet til forskningsinstitutter og som representerer tre ulike tilnærminger – tre av hensiktene fra oversikten i 3.1. Materialet er basert på rapporter, presentasjoner fra et EARTO-seminar om impact fra våren 2018 og samtaler med minst en forsker fra hvert av instituttene.

4.1 Fraunhofer

Fraunhofer-Gesellschaft (FhG) er et av verdens største forskningsinstitutter med omfattende aktiviteter innenfor naturvitenskapelige, teknologiske og samfunnsvitenskapelige fag. I praksis er det en paraplyorganisasjon over om lag 60 institutter som har til felles at de arbeider med mål om å bygge broer mellom grunnleggende forskning, teknologiutvikling og praktisk utnyttelse. Totalt arbeider nær 25.000 personer i FhG-systemet og omsetningen er på over to milliarder Euro. Litt over en tredel av omsetningen er oppdrag for næringsliv, som er instituttets viktigste bruker.

Instituttene i FhG er kjent som nære partnere til tysk industri, og eksempler som utviklingen av MP3-teknologien er blitt brukt i diskusjoner om impact. De siste årene har Fraunhofer tatt initiativ til mer systematiske undersøkelser av sin egen effekt på den tyske økonomien. Undersøkelsene er mer akademiske enn evalueringer og er gjennomført av forskere fra FhG i samarbeid med kolleger ved universiteter i andre land.⁷

Undersøkelsene har et mål som kan beskrives som en kombinasjon av 3 (lobbyvirksomhet) og 4 (innsikt) fra diskusjonen i 3.1. Først og fremst representerer

⁷ Mer informasjon om tre undersøkelser – bidrag i det tyske innovasjonssystemet, effekter på små- og mellomstore bedrifter, og effekter av samarbeid med FhG – finnes her: <https://www.fraunhofer.de/en/research/range-of-services/impact-of-fraunhofer-research.html>.

de et ønske om å måle det makro- og mikroøkonomiske bidraget til FhG, men resultatene brukes også for å markedsføre instituttets viktighet for Tyskland gjennom pressemeldinger og sammendrag på papir og nett.

De tre undersøkelsene er rent kvantitative og inneholder ingen konkrete «impact case» eller lignende type eksempler, men er i stedet opptatt av å være noe annet enn slike «anekdotiske» tilnærminger (Comin m.fl. 2018). De to største undersøkelsene bruker statistiske data. Det dreier seg om tidsserier om tyske bedrifter, samlet inn som del av den europeiske innovasjonsundersøkelsen og med noe annen statistikk i tillegg, kombinert med konfidensiell informasjon om hvem som har hatt oppdrag/samarbeid med Fraunhofer i perioden 1997 til 2014. Den tredje undersøkelsen baserer seg på en spørreundersøkelse blant små- og mellomstore-bedrifter blant FhGs kunder, og representerer en enklere tilnærming som viser at brukerne stort sett er fornøyde.

Analysene for øvrig søker å finne Fraunhofers bidrag til bedriftenes omsetning og innovasjonsaktiviteter, samt Fraunhofers makroøkonomiske bidrag målt i skatteinntekter og endringer i BNP. Tallene som presenteres er på mange måter imponerende – bedriftene får betydelig økning i omsetning etter samarbeid med FhG, Tyskland får tre ganger så mye tilbake som de bruker på FhG og bedriftene som samarbeider får 18 ganger innsatsen tilbake. Disse tallene kan kritiseres ut fra metodeutfordringene diskutert i forrige kapittel, og de forsknings- og innovasjonspolitiske implikasjonene er uklare (kanskje annet enn at Fraunhofer er viktig og fortjener generøs støtte, slik de selv presenterer det).

Det er imidlertid en rekke andre interessante aspekter som kommer fram i analysene som ikke direkte går på kvantifisering av selve impacten. To aspekter er spesielt interessante også for institutter i andre land og deres eiere og bidragsytere. For det første viser analysene at prosjektstørrelse er en viktig variabel: store prosjekter har mye større impact enn små, mye mer enn hva bare forskjellen i størrelse skulle tilsi. For det andre blir det vesentlig mer impact av gjentatt og hyppig samarbeid enn samarbeid som er mer sporadisk. Dette uttrykker sannsynligvis at fordi impact tar så lang tid, er det nødvendig med relasjoner som strekker seg over mange år for å få til et fruktbart samspill mellom forskning og utnyttelse.

Lignende undersøkelser er teknisk mulige å gjøre i Norge. Fordelen med Fraunhofer og Tyskland er at instituttet er så stort at det er mange bedrifter som har samarbeidet med det – men også mange avanserte bedrifter som ikke har gjort det. Koblingen mellom prosjektdata og offentlig statistikk er ressurskrevende og fordrer en prosjektgruppe med gode rutiner for å sikre konfidensialitet og data-sikkerhet. Analysene som refereres her er resultat av flerårige og store forskningsprosjekter mer enn enkeltstående evalueringer.

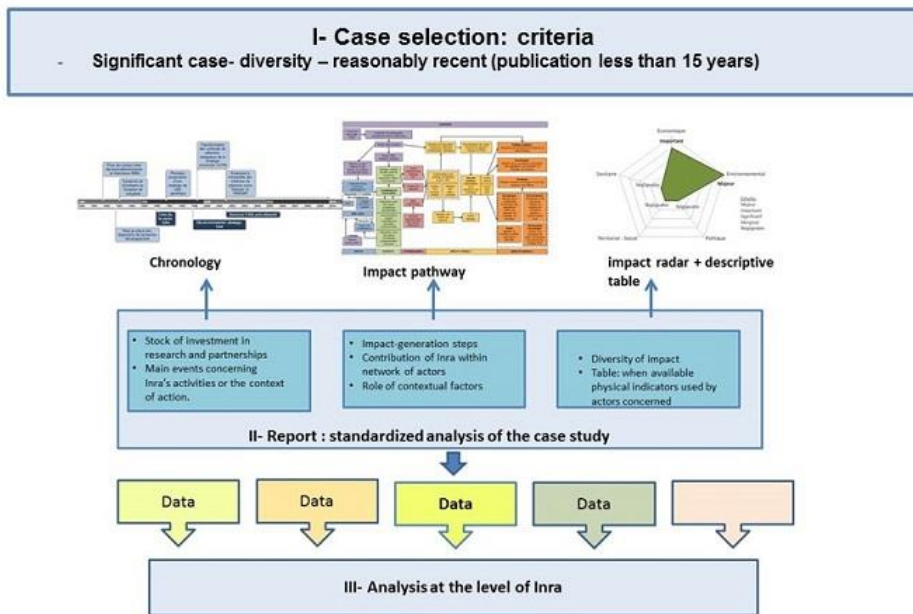
4.2 INRA

INRA er et fransk institutt for landbruksforskning (agronomisk forskning) som er det største i sitt slag i Europa og nest størst i verden med over 8000 ansatte og rundt 900 millioner Euro i omsetning. Instituttet startet i 2011 et prosjekt for å utvikle en forskningsbasert metode for å måle sine egne samfunnseffekter, kalt ASIRPA (akronym basert på målet med prosjektet). ASIRPA har vært faglig svært ambisiøst og har resultert i flere vitenskapelige artikler i ledende fagtidsskrifter. Mer enn mange andre som arbeider med impact, har ASIRPA-prosjektteamet søkt å forankre metoden i et omfattende og solid faglig rammeverk. Syv personer har vært koblet til arbeidet med å utvikle metoden, som ifølge prosjektlederen har kostet et sted mellom en og to millioner Euro.⁸

ASIRPA er casebasert og har tre hovedmål: kartlegge mekanismene som leder til impact, få fram de ulike dimensjonene i impact, samt utvikle en standardisert metode som gjør det mulig å sammenligne case og gjøre kvantitative analyser på tvers. Utgangspunktet for casene er bestemte prosjekter hos INRA slik som nye varianter av tomater og nye metoder for vinproduksjon. For hvert case går forskerne inn og ser etter aktører, nøkkelhendelser, «produktiv interaksjon» mellom forskere og brukere og annet. Ut fra dette skapes en tidslinje/sti (pathway) som viser hva de ulike aktørene har bidratt med og på hvilken måte. Det skilles mellom en rekke effekter sortert i fem kategorier: økonomi, miljø, regional utvikling, politikk og helse.

Analysen består av tre deler. Først etableres en kronologi hvor hendelser og aktører plasseres inn i et tidsdiagram. Deretter søker forskerne å forstå interaksjon, overføring og andre aktiviteter slik at en kan si noe meningsfylt om kausalitet og systemeffekter. Til slutt lages en «impact-radar» som viser hva slags dimensjoner av impact som det bestemte caset scorer høyt eller lavt på. Dette er vist i Figur 5. ASIRPA-prosjektet analyserte totalt 40 case på denne måten (jf. Gaunand m.fl. 2016) Nå er prosjektet inne i en ny fase hvor en søker å rutinisere målingene av impact og å lære opp flere personer til å gjøre det krevende analysearbeidet.

⁸ Mer informasjon på engelsk finnes her: https://www6.inra.fr/asirpa_eng/ASIRPA-project.



Figur 5. ASIRPA-tilnærmingen til impact.

Hentet fra https://www6.inra.fr/asirpa_eng/Method-and-Cases

Metoden er interessant, og enda mer enn Fraunhofer-analysene får den fram betydningen av kumulative relasjoner og den kapasiteten og kompetansen som bygges opp gjennom langvarig og jevn interaksjon. Som nevnt viser undersøkelsen at gjennomsnittlig tid fra forskning til impact var 19 år. Den viser også noen av motsetningene innenfor impact: noen av casene har hatt betydelige og positive effekter på miljø men ikke på økonomiske indikatorer, eller motsatt. Impact-radarene får fram dette også på en visuelt slående måte som muliggjør konkrete diskusjoner om hva slags verdier forskningen bidrar til.

Uavhengig av impact-typer finner forskerne fire hovedtyper av prosesser til impact, basert på i hvor stor grad det er samhandling mellom forskere og brukere, og i hvor stor grad forskningen leder til endringer andre steder. Den første typen er forskning med moderat eller lav grad av samarbeid som likevel leder til omfattende endringer andre steder. Mange av disse casene handler om sikkerhet og miljø, slik som håndtering av vannforurensning fra jordbruk. Den andre typen kalles «intensiv transformasjon basert på eksisterende nettverk», hvor mange av prosjektene har vært svært anvendte og rettet mot områder som dyrevelferd og reduksjon i bruk av plantevernmidler. Tredje type kalles «marked for teknologier» og refererer i hovedsak til eksempler der forskning på INRA har ledet til en kom-

mersialiseringsspross for et nytt produkt gjennom et patent eller en nyetablering. Siste type beskriver langvarige samarbeid hvor brukerne i lang tid har benyttet seg av forskningsbasert kunnskap.

Institusjonaliseringen av målingene på instituttet etter pilotundersøkelsen betyr også at INRA søker å bruke resultatene ikke bare i markedsføringen av seg selv, men også i måten instituttet organiserer arbeidet sitt på (formativ evaluering kombinert med innsikt, type 2 og 4 fra 3.1). Metoden er ny og en interessant kandidat for utprøving andre steder. Problemet med tilnærmingen, foruten at den er svært ressurskrevende, er nok at den fortsatt i hovedsak fokuserer på selve forskningen og velger case ut fra dette (rent fremoverskuende i stedet for begge veier).

4.3 VTT

Det finske forskningsinstituttet VTT har litt over 2000 ansatte og en omsetning på om lag 150 millioner Euro. Deres tilnærming til måling og forståelse av impact har vært å lage et system som kan brukes primært til å utvikle organisasjonen og praksis ved instituttet. På mange måter er dette en formativ evaluering men utviklet av instituttet selv snarere enn departementer eller andre aktører.⁹

VTT har definert sine mål i to kategorier: excellence og impact, begge basert også på hva som er de sentrale forskningspolitiske begrepene i vår tid. Sistnevnte er delt inn i to underkategorier: det å skape fortrinn for VTTs kunder/brukere, og det å skape effekter for det finske samfunnet. For hver av disse to kategoriene er det utviklet tre kvantitative indikatorer. For effekter for brukerne måles kompetanse (capability development), hvordan brukerne gjør det sammenlignet med en kontrollgruppe, og at brukerne får tilgang til internasjonale nettverk.

Det samles inn data, både fra tredjeparter og gjennom brukerundersøkelser, hvor hensikten primært er å følge hvordan VTT gjør det sammenlignet med sine egne tidligere resultater. Tallene i seg selv er ikke nødvendigvis veldig meningsfulle, men VTT har f.eks. gjort en «customer impact survey» hvert år på samme måte siden 2004, noe som gir en god pekepinn på utviklingstrekk og et godt verktøy for å styre instituttets aktiviteter.

Sammenlignet med de to foregående metodene, er tilnærmingen for det første mer praktisk. Måling av impact er direkte koblet til mål og strategiplaner for instituttet, og deler av målesystemet har vært i bruk i såpass mange år at det kan sies å være institusjonalisert. For det andre er tilnærmingen pragmatisk. Her brukes det meste av tilgjengelige data, både fra instituttet selv og offentlig tilgjengelige databaser, uten at det er gjort grundige vurderinger av underliggende teoretiske perspektiver og antakelser (slik både Fraunhofer og INRA har gjort). Terskelen for å introdusere et målesystem som hos VTT er nok lavere enn for de to foregående

⁹ VTT presenterer sin egen impact her: <https://www.vttresearch.com/impact>.

eksemplene. Men også dette krever en infrastruktur (systemer for strategisk ledelse, data) og vil være relativt ressurskrevende.

4.4 Sammenlignende oppsummering

De tre ulike tilnærmingene til instituttenes egne målinger av impact finnes i Tabell 3. Som tabellen viser, har tilnærmingene litt forskjellige utfordringer siden de underliggende metodene varierer. Et fellestrekk er at alle metodene er diskutert utenfor instituttene – på akademiske konferanser eller i nettverksmøter mellom forskningsinstitutter slik som gjennom interesseorganisasjonen EARTO.

Tabell 3. Tre europeiske institutters tilnærming til måling av impact.

Institutt	Hensikt	Organisering	Utfordringer
Fraunhofer	Finne Fraunhofers bidrag til den tyske økonomien Innsikt, lobbyvirksomhet	Prosjektteam med eksterne deltakere, kombinerer eksterne og interne data, teoretisk forankring (økonometriske studier)	Datasikkerhet, størrelse på datamaterialet, attribusjon, tidsperspektiv
INRA	Forstå hvordan INRAs forskning gir ulike typer av impact Innsikt, verktøy for ledelse	Prosjektteam med eksterne deltakere, interne data som utgangspunkt for casestudier, teoretisk forankring (STS, innovasjonsstudier)	Identifisering av nøkkelhendelser, kvantifisering av størrelse på effekter
VTT	Hjelpe VTT til å ha solid impact i Finland Verktøy for ledelse	Prosjektteam med interne deltakere, indikatorer basert på VTTs mål og tilgjengelige interne og eksterne data	Datakvalitet, uavhengighet, indikatorenes validitet og reliabilitet

Tre andre aspekter kan trekkes fram. For det første kan det diskuteres om slike målinger mest er noe for de store instituttene, både fordi målingene er svært ressurskrevende i seg selv og fordi disse er store nok til at en i det hele tatt klarer å finne samfunnseffekter i større datasett. Når en vet at INRAs dybdestudier av 40 case kostet opp mot 2 millioner Euro (med franske timepriser), er det klart at dette ikke er noe for alle aktører. For det andre er det åpenbart at det er forskjell mellom «innvortes» og «utvortes» bruk. Dersom målingene også skal brukes til lobbyvirksomhet eller andre former for ekstern synliggjøring av instituttenes bidrag, er legitimiteten viktig. INRA og Fraunhofer løser dette ved å organisere impact-målingene som forskningsprosjekter med eksterne partnere og med vanlig vitenskapelig publisering som en sentral del av prosjektene. For det tredje er det en implisitt problematikk knyttet til landenes størrelse. Fraunhofers målinger baserer seg til dels på en matching-metode hvor bedrifter som samarbeider med instituttet sammenlignes med mest mulig like bedrifter uten slikt samarbeid. Sannsynligvis ville dette vært vanskeligere (kvaliteten på matchingen ville vært dårligere og det ville vært vanskeligere å si noe om kausalitet) f.eks. for VTT og Finland. Slike målinger kan i Norge tenkes som en del av evalueringssystemet til Norges forskningsråd,

eller det kan tenkes som et spleiselag mellom ulike aktører hvor målet kan være å se på en gruppe med institutter i stedet for bare ett stort.

5 Oppsummering og prinsipper for gode vurderinger av impact

Denne rapporten har hatt som mål å diskutere impactbegrepet med særlig blikk på anvendt forskning og forskningsinstitutter med dette som sin mest sentrale oppgave. Impact er et sentralt forskningspolitisk begrep som uttrykker en forventning om at forskningen skal ha samfunnseffekter, og som også uttrykker en ambisjon om å måle og synliggjøre disse effektene på en systematisk måte. Ordet brukes også på norsk, blant annet i Norges forskningsråds evalueringer av forskningsfelt og -miljøer.

I kapittel 2 viste vi at impact er definert til å handle om nytte av forskning i svært vid forstand – det handler om effekter knyttet til helse, økonomi, politikkutforming, miljø, kultur eller livskvalitet, utenfor forskningssystemet. Det er slik det toneangivende REF-systemet i Storbritannia definerer begrepet, en definisjon som har spredd seg til mange land. Impact representerer enda større ambisjoner for evalueringer av effekter sammenlignet med tidligere begreper som relevans, nytteverdi og innovasjon. I tillegg retter impact seg mot alle fagområder, og det uttrykker høye forventninger til hva forskningen skal bidra til. For anvendt forskning bemerket vi at begrepet ser ut til å ha forsvunnet fra den forskningspolitiske debatten, men at det representerer den største aktiviteten i det offentlige forskningssystemet. I utgangspunktet vil det være enklere å måle impact for anvendt forskning enn grunnforskning, men det er uklart hva en større vektlegging av impact i vurderinger av forskningsprosjekter vil bety.

Kapittel 3 diskuterte ulike målemetoder og deres assosierte problemer. Det finnes et hovedskille mellom kvantitative målinger av forholdet mellom innsats og resultater og kvalitative målinger som gjerne retter seg mot bestemte eksempler eller case. For hver av hovedtypene finnes det en rekke underkategorier. Ulike metoder brukes gjerne til ulike formål. Et sentralt skille her er mellom summativ og formativ evaluering, hvor førstnevnte har mer preg av å gjøre opp et regnskap for en bestemt innsats, mens sistnevnte er mer rettet mot læring og organisasjonsutvikling. Måling av impact har dessuten ofte en tydelig forskningspolitisk side knyttet til synliggjøring av effekter på en slik måte at forskningen blir sikret fremtidige

ressurser. Generelle problemer for alle målinger er attribusjon (hvor mye av effekten skyldes forskningen og hvor mye skyldes andre forhold), tidsperspektiv (impact tar svært lang tid), kvalitet på data og indikatorer, og perspektiv på impactprosessen (samfunnssiden er i liten grad studert).

I kapittel 4 så vi på tre europeiske forskningsinstitutters systemer for måling av impact. Tyske Fraunhofer har lagt ned et stort arbeid for å måle sin egen samfunnsøkonomiske effekt, både for bedrifter som bruker instituttet og for den tyske økonomien totalt sett. Franske INRA har laget et omfattende verktøy for å studere hva slags effekter deres prosjekter har hatt og hvordan disse effektene har fremkommet. Finske VTT har et system som er tett koblet til instituttets mål og strategier, og som primært brukes til interne strategiske diskusjoner. Alle tilnærmingene har styrker og svakheter, men de har også noen interessante fellestrekk i resultatene, f.eks. betydningen av gjentatt kontakt over lang tid mellom forskere og brukere, og betydningen av at forskningsprosjekter har en viss størrelse.

Helt til slutt vil vi peke ut noen prinsipper vi mener kan være aktuelle å forholde seg til for å skape gode vurderinger av impact i instituttsektoren. Utgangspunktet vårt er at vurderinger av impact er en mulighet til bedre å forstå relasjonen mellom forskning/forskningsmiljøer og samfunn. Dessuten representerer feltet en mulighet for institutter og for organisasjoner som finansierer forskning til å utvikle bedre strategiske verktøy.

Som diskutert tidligere er det en rekke problemer knyttet til slike vurderinger. De kan ha et for stort preg av revisjon eller av egenreklame, og de bruker mange ganger indikatorer og perspektiver som representerer smale eller lineære definisjoner av impact. Det er ikke noe galt i seg selv å gjøre rede for effekter av offentlige midler til FoU, og det er heller ikke noe galt i å skape systemer som bidrar til sterk finansiering av forskning og forskningsmiljøer. Men hvis dette blir de eneste hensiktene med vurderinger av impact, er det noen viktige muligheter som går tapt. I tillegg er det trolig at metodeproblemene blir mer påtakelige fordi målingene blir en rent politisk øvelse hvor det potensielt står mye på spill for alle organisasjonene som er involverte.

Våre råd kan oppsummeres i følgende punkter:

1. **Legg et bredt samfunnsperspektiv til grunn for forståelsen av «impact».** Nedslagsfeltet for samfunnseffekter av forskning er ikke forutsigelig og bør ikke kun knyttes til økonomisk effektivitet. Denne tankegangen er bygget inn i nyere evalueringer som de til Norges forskningsråd av humanistisk og samfunnsvitenskapelig forskning. Men det kan gjøres på en enda mer systematisk måte som også synliggjør mulige spenninger mellom ulike effekter. ASIRPA-systemet fra kapittel 4 kan være en rettesnor for et slikt arbeid.

2. **Se ikke bare etter positive eksempler på virkninger av forskning.** Vi vet altså at forskning i noen tilfeller har hatt problematiske eller sammensatte effekter. Selv om definisjonene av impact tar høyde for dette, fungerer i praksis impactmålinger som en synliggjøring av de «positive» eksemplene som en håper kan gi fortsatt støtte til forskningen. Særlig de korte to-siders casene fra utenlandske og norske evalueringer gir et svært begrenset innblikk i forskningens samlede effekter. Her vil det være viktig å supplere slike case med andre metoder.
3. **Vurder hvordan det er tilrettelagt for samspillet mellom forskning og samfunn.** Vi vet at om forskning tas i bruk eller ikke avhenger av både direkte og indirekte former for samspill mellom forskere og brukere. Noe av dette samspillet vil dessuten være utenfor forskernes kontroll. Dette taler for at evalueringsmetoder i større grad må fokusere på impact som prosess og i større grad trekke relevante brukermiljøer med i evalueringene. SIAMPI-metoden beskrevet i kapittel 3 er relevant for å se mer på samspill, men også denne er mer fokusert på forskerne enn på brukerne.
4. **Ta hensyn til målene.** Forskning utføres i organisasjoner med bestemte formål. Det betyr at en ikke bare bør vurdere enkeltforskeres eksepsjonelle bidrag (de aller beste publikasjonene o.l.), men også hvordan organisasjonen normalt yter ut fra sitt formål. Utnytt at hver enkelt organisasjon i instituttsektoren har spesifikke samfunnsoppdrag eller oppgaver som de kan evalueres i forhold til. Ved å identifisere en bestemt del av samfunnet hvor instituttet kan forventes å yte, kan det også være enklere å finne relevante metoder for å måle eller dokumentere impact. Eksempelet VTT fra kapittel 4 har forsøkt å utvikle et slikt system.
5. **Følg god praksis ved utvikling og bruk av indikatorer.** En bør først og fremst bruke tid på å velge de mest robuste og de som egner seg best for det aktuelle fagområdet og samfunnsområdet. I tillegg vil det være viktig å kombinere kvalitative og kvantitative vurderinger, la både datainnsamling og analyse være transparente, og diskutere hva slags signaler og incentiver som ligger i indikatorene. Alle indikatorer har betydelige utfordringer og begrensninger, disse må ikke bli helt borte i «executive summary» eller nettsaken.

Vårt perspektiv er at anvendt forskning ikke krever noe eget vurderingssystem i seg selv. Med dagens definisjoner av impact og ikke-lineære forståelse av relasjonen mellom forskning og samfunn, er det unødvendig å lage målesystemer spesielt for ulike typer forskning. På den annen side kan det godt hende at forskningsinstituttene representerer en særskilt arena for forskning, og at denne arenaen kan trenge noen spesielt tilpassede evalueringssystemer for impact. Det spesielle

med institutter er at forskningen deres som regel har mye mer konkrete og spesifikke mål enn den forskningen en finner ved universiteter og høyskoler (selv om denne også ofte er definert som anvendt). Som eksemplene fra kapittel 4 viser, kan instituttene selv bidra til å utvikle målesystemer som fanger opp det de selv synes er viktige bidrag. Slike systemer kan utvikles til ikke bare å være rettet mot gode rammebetingelser men også (ideelt sett) til noe som kan fungere som en ressurs i strategiske veivalg.

Det ligger altså noen farer i hvordan en tenker om impact i våre dager med at det handler om impact av forskning i seg selv, litt frikoblet de målene en har for forskningen. Her kan instituttene potensielt spille en positiv rolle i debatten og synliggjøre hvilke verdier i samfunnet som deres aktiviteter har en innvirkning på. Vi tror også at evaluering av impact generelt og fra grunnforskning spesielt kan ha noe å lære av at det gjøres solid arbeid for anvendt forskning, hvor impact-vurderinger muligens er noe enklere på grunn av tids- og attribusjonsforholdene.

En sentral utfordring for instituttene i denne situasjonen er en større vektlegging av å dokumentere sine egne resultater og aktiviteter. Dette gjelder kvantitative indikatorer men også det vi kan kalle institusjonell hukommelse. Vi vet at samfunnseffekter av forskning ofte oppstår etter lang tid, også fra anvendt forskning. Ofte kan en effekt i dag være et resultat av langvarig og kumulativt arbeid basert på ideer og forsøk fra flere tiår tilbake. Noen må kunne synliggjøre disse sammenhengene på en godt begrunnet måte, noe som krever innsikt i og refleksjon over organisasjonens historie og kontekst.

Referanser

- Alston, J.M., M.A. Andersen, J.S. James & P.G. Pardey (2010). *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending*. Springer Science + Business Media, Natural Resource Management and Policy no. 34.
- Benedictow, A. m.fl. (2018). *Evaluation of SkatteFUNN*. Oslo: Samfunnsøkonomisk analyse AS, Rapport 18-2018.
- Bilsen, V., I. De Voldere, M. Van Hoed & K. Zeqo (2018). *Economic footprint of 9 European RTOs in 2015-2016. Final report*. Brussel: Idea Consult. Tilgjengelig her:
[http://www.earto.eu/fileadmin/content/03_Publications/2018/EARTO Economic Footprint Study - Impact of 9 RTOs in 2015-2016 - Final Report.pdf](http://www.earto.eu/fileadmin/content/03_Publications/2018/EARTO_Economic_Footprint_Study_-_Impact_of_9_RTOs_in_2015-2016_-_Final_Report.pdf)
- Bornmann, L. (2013). What Is Societal Impact of Research and How Can It Be Assessed? A Literature Review, *Journal of the American Society for Information Science and Technology*, 64(2): 217–33.
- Bozeman, B. & D. Sarewitz (2011). Public Value Mapping and Science Policy Evaluation. *Minerva*, 49(1): 1-23.
- Comin, D., G. Licht, M. Pellens & T. Schubert (2018). Do companies benefit from public research organizations? The impact of Fraunhofer. Manuskript til vitenskapelig artikkel, tilgjengelig her:
<https://www.fraunhofer.de/content/dam/zv/de/leistungsangebot/wirkung-von-forschung/Do%20Companies%20Benefit%20from%20Public%20Research%20Organizations%20-%20The%20Impact%20of%20Fraunhofer.pdf>
- Damvad (2018). *Evaluation of the Scheme for Research-based Innovation (SFI)*. København: Damvad Analytics.
- Donovan, C. (2011). State of the art in assessing research impact: introduction to a special issue. *Research Evaluation*, 20:175-179, doi:10.3152/095820211X13118583635918.
- Donovan, C. & S. Hanney (2011). The 'Payback Framework' Explained. *Research Evaluation*, 20(3): 181–83.

- Fjærli, E., M. Rybalka & Ådne Cappelen (2016). Innovasjons- og verdiskapingseffekter av utvalgte næringspolitiske virkemidler. Oslo: SSB, Rapport 2016/12.
- Gaunand, A., A. Hocdé, S. Lemarié, M. Matt, E. de Turckheim (2012). How does public agricultural research impact society? Towards a characterization of various pathways. *Research Policy* 44, 849-861. doi:10.1016/j.respol.2015.01.009.
- Gibbons M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott & M. Trow (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Gulbrandsen, M. (2000). *Research Quality and Organisational Factors: an Investigation of the Relationship*. NTNU: Dr.ing.-avhandling 2000:90.
- Gulbrandsen, M. (2017). Hvordan kan impact måles? Oslo: Norges forskningsråd, *Det norske forsknings- og innovasjonssystemet (Indikatorrapporten) 2017*, s. 155-156. Tilgjengelig på: [https://www.forskningsradet.no/prognett-indikatorrapporten/Artikkel/Hvordan kan impact males/1254036676158](https://www.forskningsradet.no/prognett-indikatorrapporten/Artikkel/Hvordan%20kan%20impact%20males/1254036676158)
- Gulbrandsen, M. & L. Langfeldt (2004). In Search of 'Mode 2': the Nature of Knowledge Production in Norway. *Minerva*, 42:237-250. <https://doi.org/10.1023/B:MINE.0000038305.35138.31>
- Harland, K. & O'Connor, I. (2015). *Broadening the Scope of Impact. Defining, assessing and measuring impact of major public research programmes, with lessons from 6 small advanced economies*. Dublin, Irland: Small Advanced Economies Initiative. Tilgjengelig på: [https://www.sfi.ie/resources/SAEI Impact-Framework Feb 2015 Issue2.pdf](https://www.sfi.ie/resources/SAEI%20Impact-Framework%20Feb%202015%20Issue2.pdf).
- Joly, P.-B., A. Gaunand, L. Colinet, P. Larédo, S. Lemarié & M. Matt (2015). ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. *Research Evaluation*, 24(4), 440-453. doi: 10.1093/reseval/rvv015.
- Langfeldt, L. & L. Scordato (2016). *Efficiency and flexibility in research funding. A comparative study of funding instruments and review criteria*. Oslo: NIFU, NIFU-rapport 9/2016.
- Martin, B.R. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Research Evaluation*, 20:247-254, doi:10.3152/095820211X13118583635693.
- Mazzucato, M. (2018). *Mission-Oriented Research & Innovation in the European Union*. Brussel: EU-kommisjonen. Doi:10.2777/360325.
- Områdegjennomgang (2017). *Områdegjennomgang av Norges forskningsråd. Rapport fra ekspertgruppen*. Rapport levert til Kunnskapsdepartementet og Finansdepartementet 7. februar 2017. Tilgjengelig her: [https://www.forskningsradet.no/no/Nyheter/Ekspertgruppens rapport Kval](https://www.forskningsradet.no/no/Nyheter/Ekspertgruppens_rapport_Kval)

[itet i forskningen er et hovedanliggende for Forskningsradet/1254024823068/p1174467583739](https://www.forskningsradet.no/1254024823068/p1174467583739)

- Penfield, T., M.J. Baker, R. Scoble & M.C. Wykes (2014). Assessment, evaluations, and definitions of research impact: A review, *Research Evaluation*, 23:21-32, doi:10.1093/reseval/rvt021.
- Sarewitz, D. (2016). Saving Science. *The New Atlantis*, 49:4-40. Tilgjengelig her: <https://www.thenewatlantis.com/publications/saving-science>
- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3: 17078, doi:10.1057/palcomms.2017.78.
- Sivertsen, G. & I. Meijer (2018). Normal versus extraordinary impact. How to understand, evaluate and improve research activities in their relations to society. *Research Evaluation*, submitted.
- Smith, R. (2001). Measuring the social impact of research. *British Medical Journal*, 323(7312), 528. doi:10.1136/bmj.323.7312.528.
- Spaapen, J. & L. van Drooge (2011). Introducing 'Productive Interactions' in Social Impact Assessment. *Research Evaluation*, 20(3): 211–18.
- Stern, N. (2016). *Research Excellence Framework (REF) review: Building on success and learning from experience*. Ref: IND/16/9. Department for Business, Energy & Industrial Strategy.
- Stokes, D.E. (1997). *Pasteur's Quadrant: basic science and technological innovation*. Washington, D.C.: Brookings Institution Press.
- Wilsdon, J. m.fl. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. London: HEFCE. DOI: 10.13140/RG.2.1.4929.1363
- Weißhuhn, P., K. Helming & J. Ferretti (2018). Research impact assessment in agriculture—A review of approaches and impact areas. *Research Evaluation*, 27:36-42, doi:10.1093/reseval/rvx034

Nordisk institutt for studier av
innovasjon, forskning og utdanning

Nordic institute for Studies in
Innovation, Research and Education

www.nifu.no