



Dataanalyse i offentlig sektor

A Nesta Guide

Eddie Copeland

Tom Symons

Hilary Simpson

Nevena Dragicevic

Oversatt av tolketjenesten
ved Trondheim kommune

nesta



Denne guiden er lisensiert av ©Nesta under Creative Commons BY-NC-SA 4.0 Internasjonal Lisens.

For å se en kopi av lisensen, besøk <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Trondheim kommune har også spesifikt bedt om, og fått, tillatelse til å lage denne oversettelsen.



- Introduksjon

- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Om guiden

Denne guiden er til virksomheter i den offentlige sektoren som er interessert i bruk av dataanalyse for å ta bedre valg og forbedre offentlige tilbud.

Metodene og forslagene er hovedsakelig basert på Nesta sitt program, **Offices of Data Analytics** (ODA). Gjennom uttesting i London, Nordøst-England og Essex, har vi utforsket hvordan byer og regioner kan bruke ODA metoden til å sette sammen, analysere og handle basert på data hentet fra flere avdelinger i den offentlige sektoren for å forbedre offentlige tilbud.

For mer informasjon om ODA, besök:

<https://www.nesta.org.uk/project/offices-data-analytics>

- Introduksjon

- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Opplæringsmodulene

Presentasjonen inneholder disse modulene:

Modul 1: Hvordan dataanalyse kan hjelpe den offentlige sektoren

Modul 2: Utfordringer ved bruk av data i den offentlige sektoren

Modul 3: ODA metoden

Modul 4: De 8 fasene av et ODA prosjekt

Modul 5: Lovlig og etisk bruk av data

Modul 6: Planlegging og bruk av et ODA forprosjekt

Modul 7: Hvordan å gjøre det enklere neste runde

Modul 8: Sette opp et Office of Data Analytics

Modul 1:

Hvordan dataanalyse kan hjelpe den offentlige sektoren

- Introduksjon
- **Modul 1**
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

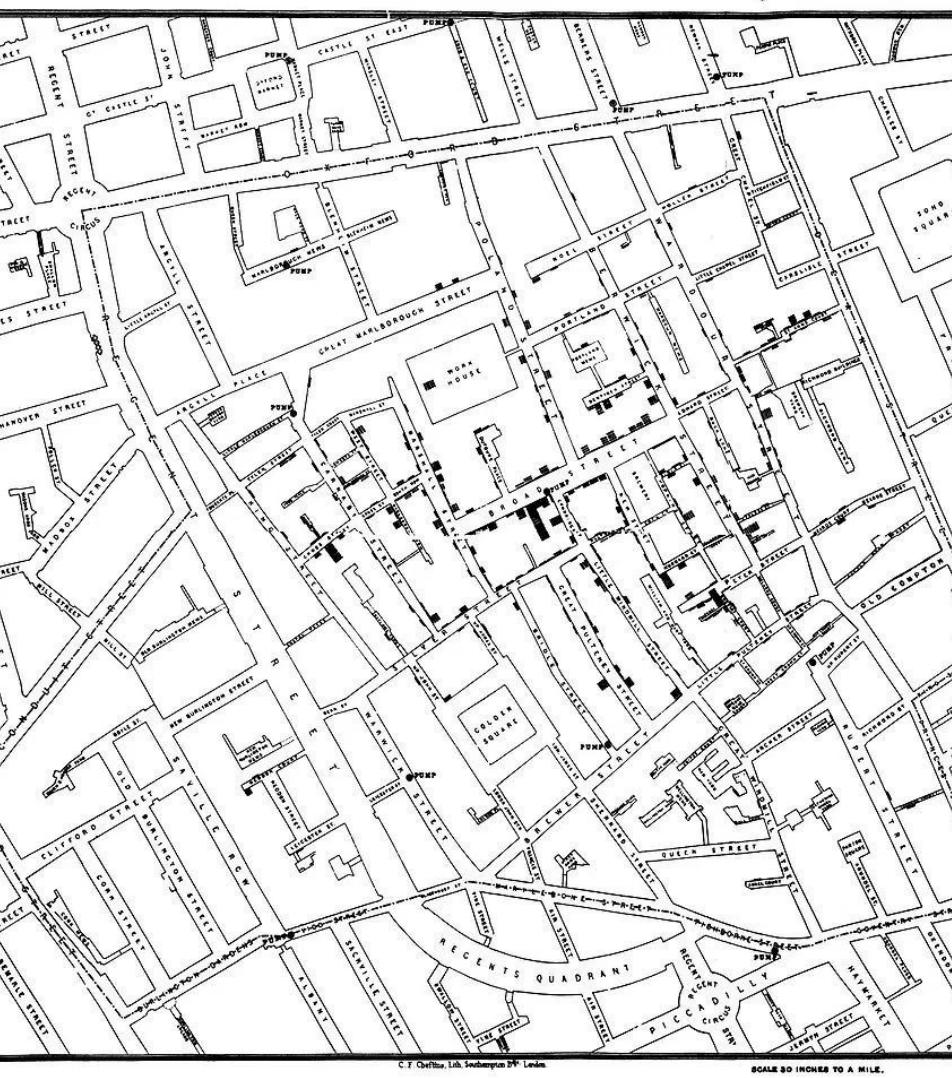
Dataanalyse: Fra innsikt til handling

Dataanalyse er oppdagelsen, tolkningen og kommuniseringen av betydningsfulle mønster i data.

Det kan brukes av team og organisasjoner individuelt til å ta velinformerte beslutninger og utføre bedre handlinger.

Det kan også brukes for å hjelpe flere team til å samarbeide mer effektivt.

Hos Nesta mener vi at dataanalyse har mest verdi når det fører til bedre handlinger. Derfor fokuserer presentasjonen på innhenting av data som er **anvendelige**.



Gamle arbeidsmåter, ny teknologi

Bruk av data for anvendelig informasjon er ikke noe nytt. I 1854 ble John Snow kjent for sin kartlegging av dødsfall i Soho, London, som viste at et kolerautbrudd i området skyldes en lokal vannpumpe som var forurenset. Kartleggingen førte til at håndtaket på pumpen ble fjernet, som reddet mange liv.

Den største forandringen i dag er at datamaskiner lar oss analysere bredere mengder med informasjon på mer sofistikerte måter.

nesta

- Introduksjon
- **Modul 1**
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Hva kan dataanalyse gjøre?

Organisasjoner innenfor den offentlige sektoren jobber mot flere forskjellige mål. Hvor passer dataanalysen inn?

Dataanalyse er spesielt hjelpsomt for*:

- Identifisering av spesifikke tilfeller i en større gruppe
- Prioritering av tilfeller basert på risiko og behov
- Utvikling av raskere varslingsverktøy
- Ta bedre og raskere valg
- Raskere bevilgning av ressurser

La oss se på noen eksempler ...

*Kategorier hentet fra NOLAnalytics: <https://datadriven.nola.gov/nalytics/>



Effektiv boliginspeksjon av HMOs

I London analyserte 12 forskjellige bykommuner samt the Greater London Authority data på kjente risikofaktorer assosiert med ulovlige Boliger med Flerfoldige Husholdninger (HMO - Houses in Multiple Occupation) for å hjelpe boliginspektører til å finne eiendommer som sannsynligvis er ulovlige. Målet var å øke lisensinntekter og beskytte sårbare leietakere.

Les hele studiet

Bilde: AD_Imgaes | Pixabay CC0 Creative Commons



Bekjempe moderne slaveri

Essex Politiet og Essex County Council undersøkte hvordan datadeling og analyse kan gi dem en bedre forståelse for lokale bedriftsinspeksjoner.

Målet var å forbedre samarbeidet mellom de mange forskjellige organisasjonene i den offentlige sektoren som er involvert i evalueringen av sikkerheten og lovlydigheten til bedrifter.

Bilde: Brian A Jackson / Shutterstock.com



Optimalisering for utplassering av utrykningskjøretøy

I New Orleans ble data om karakteristikker, åsteder og tidsberegninger av tidligere nødssituasjoner analysert for å bedre forutse hvor og når fremtidige nødssituasjoner kunne oppstå.

Ved hjelp av analysen kunne de identifisere de optimale parkeringsplassene for ambulanser for å redusere utrykningstiden.

Les hele studiet

Bilde: [Emergency Medical Transport, Inc.](#)



Newcastle NEET¹ Analyse

En analyse av ungdommer innenfor “NEET” utført av Newcastle City Council hjelper de lokale myndighetene med å identifisere hvilke barn som har størst risiko for å falle utenfor utdanning, arbeid eller opplæring.

Les hele studiet (side 26)

1 Oversetters anmerkning: NEET, “Not in Education, Employment, or Training” refereres også til som “unge på kanten” på norsk.

Bilde: 5477687 | Pixabay CC0 Creative Commons



Å forstå motorulykker

En analyse utført av The Behavioural Insights Team i Øst-Sussex over KSI² som resulterte i alvorlig skade eller dødsfall. Analysen motbeviste mange utbredte antagelser om hva som forårsaker ulykker, og hjalp de lokale myndighetene finne hvor tiltak ville ha størst utbytte.

Les hele studiet (side 21)

2 Oversetters anmerkning: KSI står for Killed or Seriously Injured, eller saker med alvorlig personskade.

Bilde: SteelFish | Pixabay CC0 Creative Commons

Kort sagt så er dataanalyse hjelpsomt fordi...

Det gjør at mange av de
velprøvde metodene
fungerer bedre

Men det er ikke alltid like enkelt...

Modul 2:

Utfordringer ved bruk av
data i den offentlige
sektoren

- Introduksjon
- Modul 1
- **Modul 2**
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Uffordringer ved bruk av data i den offentlige sektoren

Virksomheter i den offentlige sektoren møter mange utfordringer når det kommer til databruk.

Det første er et grunnleggende problem rundt **datakvalitet**. Vanlige kvalitetsproblemer er:

- Dokumentene eksisterer kun på papir
- Dokumentene er digitalisert, men i et format som er vanskelig å analysere, for eksempel en PDF
- Informasjonen er ikke dokumentert på en konsistent måte, slik som "Smith gate" og "Smith gt.".
- Dokumenter om en og samme person eller det samme temaet har ikke et gjennomgående kjennemerke
- Dokumentet er ubevisst duplisert

- Introduksjon
- Modul 1
- Modul 2**
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Uffordringer ved bruk av data i den offentlige sektoren

Tilgjengelighet er et annet problem. Virksomheter i den offentlige sektoren har ofte tusenvis av dokumenter som har samlet seg opp over tid og kan være vanskelig å lete igjennom. Dette fører ofte til at ansatte ikke har kjennskap til hvilken nyttig informasjon andre team sitter med.

I et annet tilfelle vet vi at dataene eksisterer, men tenker de er for vanskelig å ta i bruk fordi de er i form av åpne tekstfelter, gamle e-poster, møtereferater, osv. Når IT tjenester er outsourced kan man risikere at avdelingen ikke får tilgang til data som omhandler sine egne tjenester, eller at de må betale ekstra avgifter! (Vårt råd: Dere burde eksplisitt unngå lignende situasjoner i fremtidige kontrakter med eksterne leverandører.)

Nesta har fremstilt flere mulige løsninger i rapporten vår:
[Can Government Stop Losing its Mind?](#)



Puslespillproblemet

Enda en utfordring er at mange dataanalyseprosjekter krever at du finner, analyserer og tar i bruk data fra forskjellige team og virksomheter.

Dette blir gjort vanskelig på grunn av **puslespillproblemet**: hver gruppe har sin del av datapuslespillet, men man kan sjeldent sette sammen alle bitene for å se det større bildet.

- Introduksjon
- Modul 1
- **Modul 2**
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Uffordringer skapt av puslespillproblemet

Puslespillproblemet forhindrer virksomheter i den offentlige sektoren fra å gjøre at mange av de velprøvde metodene fungerer bedre. For eksempel:

1) Delte tjenester: det er vanskelig for virksomheter å se hvor de kan utdele ressurser med samarbeidspartnere om de ikke har oversikt over størrelsen og plasseringen av problemene, etterspørsel og muligheter utenfor sitt eget område.

2) Fokusere på områder som trenger mest hjelp: det er vanskelig å dele ut ressursene på en effektiv måte hvis virksomhetene ikke har til dataene som viser hvor befolkningen og områdene som trenger mest hjelp ligger.

- Introduksjon
- Modul 1
- **Modul 2**
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

Uffordringer skapt av puslespillproblemet

3) Intelligent koordinering av team: det er vanskelig for team å koordinere aktiviteter effektivt innenfor komplekse områder slik som sosialomsorg for voksne (hvor opptil 30 organisasjoner kan måtte samarbeide for å støtte ett individ eller en familie) hvis de ikke har oversikt over hva de andre holder på med.

4) Forutse og Forebygge: det er vanskelig å ta tak i problemene tidlig hvis organisasjonene ikke kan samle og analysere datasett som kollektivt ville vist til tilfellene med høyst fremtidig risiko.

Puslespillproblemet har flere kjente problemer, blant annet:

Tekniske: Organisasjoner og grupper bruker ofte forskjellige IT system, og ikke alle kommuniserer like enkelt med hverandre eller gjør dataene lett tilgjengelig.

Data: Dokumentene kan være lagret i forskjellige formater og etter forskjellige konvensjoner, som kan gjøre det vanskelig å matche og analysere informasjon om den samme personen eller det samme stedet.

Juridiske: Det er enkelte ting loven ikke tillater. Det er enda flere ting som data kan brukes til som ikke skjer fordi de fleste som jobber i det offentlige ikke er klar over hva datasikkerhetsloven tillater eller forbyr.

Organisasjonens: Alle enheter i den offentlige sektoren er satt opp for å fokusere på en spesifikk gruppe, på en spesifikk måte. Det kan ta store forandringer innenfor virksomheten og kulturen for å sette i gang systematisk samarbeid.

Hvorfor det er så vanskelig å dele data (og hvordan det kan gjøres enklere)

Hvorfor det er vanskelig:

Tekniske

- Tilpassede, isolerte IT systemer
- Eldre systemer som gjør det vanskelig å hente ut data
- Outsourcede IT tjenester som krever penger for data tilgang
- Mangel på felles plattform for datadeling

Data

- Data i vanskelige format, slik som PDF
- Data som ikke er oppført på en konsistent måte
- Bruk av forskjellige standarder
- Mangel på gjennomgående kjennetegn
- Mangel på lett tilgjengelig data

Hvordan gjøre det enklere

Tekniske

- Bruk av teknologi som samsvarer med vanlige standarder for gjennomførlighet.
- Insistere på at all IT har åpne programmeringsgrensesnitt (API)
- Pass på at kontrakter gir full adgang til data
- Investere i en delt plattform for datadeling

Data

- Lagre all data i et format som kan leses på maskin
- Oppretthold konsistent oppføring av data
- Bruk felles standarder
- Bruk unik identifikasjon, f.eks. unike stedsreferansenummer (UPRN – Unique Property Reference Number)
- Frigjør ikke-personlig data automatisk

Juridiske

- Ledelse som unngår risiko
- Ansatte som er usikre på datasikkerhetslover
- Den mest senior av dataspesialistene er sjef for Datasikkerhet
- Mangel på maler på delingsavtaler av data

Organisasjon

- Grupper skapt for å fokusere på deres egne ansvarsområde
- Mangel på dedikert tid og ressurser til datasamarbeid
- Ledere mangler forståelse for rollen og virkemidler ved bruk av data

Juridiske

- Alle ansatte bør ha opplæring om personvernkonsekvenser (DPIA)
- Ansatt Chief Data Officer står for ansvarlig datadeling
- Bruk mal for informasjonsdeling protokoll

Organisasjon

- Etabler avdelinger for dataanalyse
- Sett interne ansatte til å jobbe med datavitenskap i stedet for KPI rapportering
- Ledere sørger for å bruke data før vurdering

- Introduksjon
- Modul 1
- **Modul 2**
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8

ODA Metoden

For å komme over noen av disse utfordringene har Nesta eksperimentert med og raffinert en metode for å kjøre analyseprosjekter i den offentlige sektoren. Vi kaller den for **ODA** (Office of Data Analytics) **metoden**

Arbeidet var originalt inspirert av arbeidet til [lenke: Mayor's Office of Data Analytics] (MODA) i New York, etablert i embetstiden til Borgermester Michael Bloomsberg.

Vi vil videre utforske denne metoden og vise hvordan det kan hjelpe deg med å fastlegge om et problem kan løses med dataanalyse.

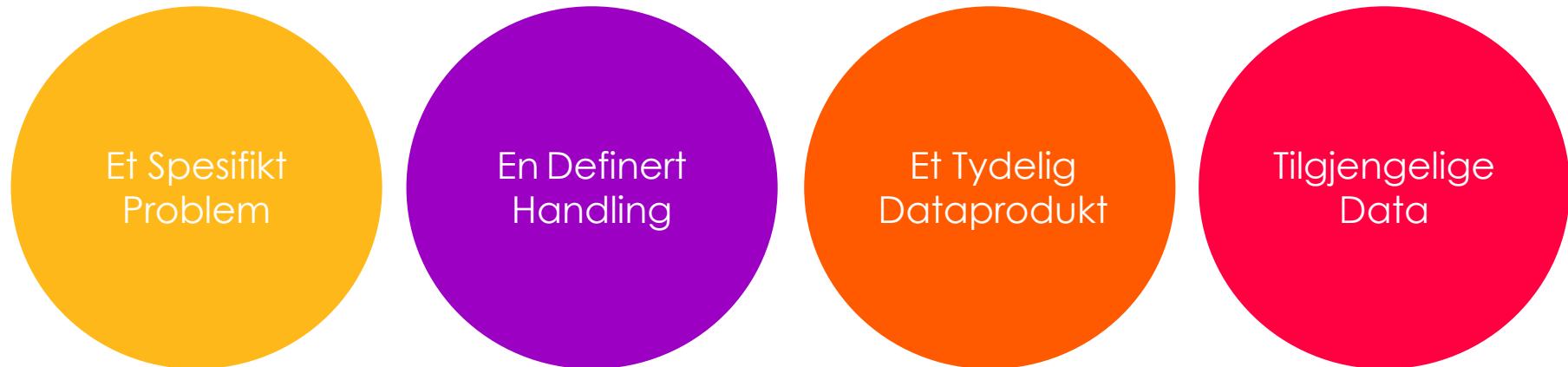
Modul 3:

ODA (Office of Data Analytics) metoden

Hvordan se om et problem kan løses med dataanalyse

Virksomheter i den offentlige sektoren møter mange forskjellige utfordringer. Hvordan ser vi hvilke av dem som kan løses med dataanalyse?

Vi tror at suksessfulle dataanalyseprosjekter består av fire hovedelementer:



La oss se på dem en og en.

Hva er det spesifikke problemet?

HOVEDPOENG: Det er ytterst viktig å bevege seg fra større, makronivå problemer til noe spesifikt og gjennomførbart.

Virksomheter i den offentlige sektoren håndterer mange store problemer, men enkelte av dem er for bredt definert til å ha en umiddelbar løsning. For eksempel så er problemstillingen: "Det finnes tilfeller av moderne slaveri i byen" for vag.

Etter litt betenkning kan problemstillingen omformuleres til noe mer spesifikt, slik som "Vi vet ikke hvilke av de regulerte bedriftene som har større sannsynlighet for å utnytte offer for moderne slaveri."



Et Spesifikt
Problem

Fem spesifikke problemtyper



Et
Spesifikt
Problem

Når man forsøker å gjøre problemstillingen mer spesifikk og gjennomførbar, så er det hjelpsomt å tenke over hvilke problemer dataanalysen er egnet for å håndtere.

Analyseteamet i **New Orleans Avdeling for Resultater og Ansvar** har vært hjelpsomme nok til å legge fram fem spesifikke problemstillinger:

5 Spesifikke Problemtyp*
Det er vanskelig å identifisere spesifikke mål i en større befolkning
Tjenestene kategoriserer ikke tilfeller med høyest prioritet tidlig
Ressurser er overdrevent fokusert på reaktive tjenester
Beslutninger blir tatt gjentatte ganger uten tilgang til all relevant informasjon
Ressurser blir planlagt eller sendt ut uten input fra nyligste tjenestedata

*Hentet fra: <https://datadriven.nola.gov/nolalytics/>

Hva er det spesifikke problemet?

Den spesifikke problemstillingen din burde ikke være et spørsmål, men heller formulert slik:



**Problemet vårt er at ... [sett inn spesifikk
problemstilling].**

Hva vil du muliggjøre?

HOVEDSPØRSMÅL: Hva ville du gjort
annerledes om du hadde all informasjonen
du trengte om ditt spesifikke problem?

For å være tydelig, **dataanalyseprosessen er ikke tiltaket.** Det er viktig å identifisere praktiske løsninger og tiltak som er innenfor din kontroll å forandre. For eksempel så kan ikke virksomheter 'løse' bostedløshet på egenhånd - men du kan kanskje ta for deg et spesifikt aspekt i ditt miljø.

Finn ut nøyaktig **hvem** som skal gjøre noe,
hvor og **når** det skal gjøres.



En Definert
Handling

Hva vil du muliggjøre?

En
Definert
Handling

Det er fem mulighetstyper assosiert med de fem problemtypene vi la fram tidligere.*

Spesifikk Problemtyp	Muligheter
Det er vanskelig å identifisere spesifikke mål i en større befolkning	Identifiser spesifikke tilfeller i en bredere gruppe
Tjenestene kategoriserer ikke tilfeller med høyest prioritet tidlig	Prioriter tilfeller basert på risiko eller behov
Ressurser er overdrevet fokusert på reaktive tjenester	Opprett tidlige varslingsverktøy for å jobbe proaktivt
Beslutninger blir tatt gjentatte ganger uten tilgang til all relevant informasjon	Ta bedre, raskere beslutninger
Ressurser blir planlagt eller sendt ut uten input fra nyligste tjenestedata	Optimaliser allokering av ressurser

*Hentet fra: <https://datadriven.nola.gov/nolalytics/>

Hva vil du muliggjøre?

List opp alle handlingene eller tiltakene du ville iverksatt for å løse det spesifikke problemet **hvis du hadde hatt mer informasjon**:



Problemet vårt er at ... [sett inn spesifikk problemstilling].



For å løse dette vil vi ... [liste over tiltak du ønsker å implementere].

Hvilke dataprodukt trenger du?

HOVEDSPØRSMÅL: Hvilken informasjon må en person ha tilgang til på skjermen for å kunne utføre handlingene som ble bestemt i det forrige steget?

Det er usannsynlig at de som utfører tiltaket (f.eks førstelinjearbeidere eller kundeansvarlige) vil be om et regneark eller rådata. I stedet vil de ofte ha informasjonen lagt fram på en mer forståelig måte som gir **faktisk innsikt**- det er det vi mener med et “dataprodukt”.

Et dataprodukt kan være et kart, varmekart, prioritert liste, varselsystem, et dashboard, en visualisering og så videre.



Et Tydelig
Dataproduct

Hvilke dataprodukt trenger du?

Et Tydelig
Dataproduct

Enkelte dataprodukter passer best til spesifikke problemer eller mulighetstyper

Spesifikk Problemtype	Muligheter	Eksempel på dataprodukt
Det er vanskelig å identifisere spesifikke mål i en større befolkning	Identifiser spesifikke tilfeller i en bredere gruppe	En graf som viser avvik eller ville observasjoner
Tjenestene kategoriserer ikke tilfeller med høyst prioritet tidlig	Prioriter tilfeller basert på risiko eller behov	En prioritert liste
Ressurser er overdrevet fokusert på reaktive tjenester	Opprett tidlige varslingsverktøy for å jobbe proaktivt	En advarsel for å informere om problemer når en grense er nådd
Beslutninger blir tatt gjentatte ganger uten tilgang til all relevant informasjon	Ta bedre, raskere beslutninger	En visualisering av data
Ressurser blir planlagt eller sendt ut uten input fra nyligste tjenestedata	Optimaliser allokering av ressurser	Et kart eller varmekart som viser hvor tilfeller oppstår

Dataproduct eksempler

Her er fire eksempler på dataprodukter brukt av virksomheter i offentlig sektoren i Storbritannia og USA:

EN PRIORITERT LISTE: I London baserte bygningsinspektørene i flere distrikt seg på tilfeldige inspeksjoner eller tips for å finne ulovlige boliger med Flerfoldige Husholdninger (HMO - Houses in Multiple Occupation). Et ODA-forprosjekt ville forandre dette ved å utvikle prioriterte inspeksjonslister som kunne lede inspektørene til eiendommer som sannsynligvis var HMOs, basert på karakteristikk fra tidligere ulovlige HMOs.

ET VARSEL. For å redusere at politibetjenter utøvde overdreven bruk av vold, kombinerte byen Charlotte i North Carolina demografier, trening, lønnslister, interne forhold og andre data [Lenke: [for å utvikle et tidligvarselsystem](#)] for når en betjent var sannsynlig for å ha et negativt sammenstøt med offentligheten.

ET KART. I samarbeid med SumAll Foundation, New York City bekjemper rekordhøye antall hjemløse ved å analysere og visualisere mønstrene for utkastelser som fører til hjemløshet for familier. Prosjektet er også med på å forbedre treffsikkerheten av oppsökende tjenester ved å forutsi hvilke nabologene, bygningene og spesifikke adresser der ressursene er mest behov for.

ET Dashboard. Louisville, Kentucky har forbedret behandlingstid for ambulanse ved å bruke data for å identifisere hindringer for raskere respons, som har spart byen 1,4 millioner dollar (USD). Ekspeditører støttes nå med rapporter fra en datamaskinstøttet utsendelsessystem, som oppdager skjulte ineffektiviteter og overvåker sanntidspllassering av ambulanser for å få til raskere responstidene.

Dataprodukt eksempler

Sunderland City Council utviklet **Adult 360**, et prosjekt for å samle informasjon om enkeltpersoner og deres liv fra flere kildesystem, inkludert velferd, CES, TeleCare, mellomliggende omsorg, bysykehus og politiet.

Dette har bidratt til bedre og mer koordinert omsorg og utstyr over 350 helse- og sosialomsorg praktiserende med en helhetlig oversikt over personens interaksjoner- som demonstrert i denne modellen.

John Smith

Data sources: Spoa, Swift, Dfg, Rah, Police, Jontek, Mesals, Chs

Date of birth: 12 Jan 1970 (46 yrs old)

Date of death: N/A

NHS Number: 9999999999

PID: 999999

Address: Address SR9 9SR (as of 10 August 2015)

Main category: Physical Support: Personal Care Support

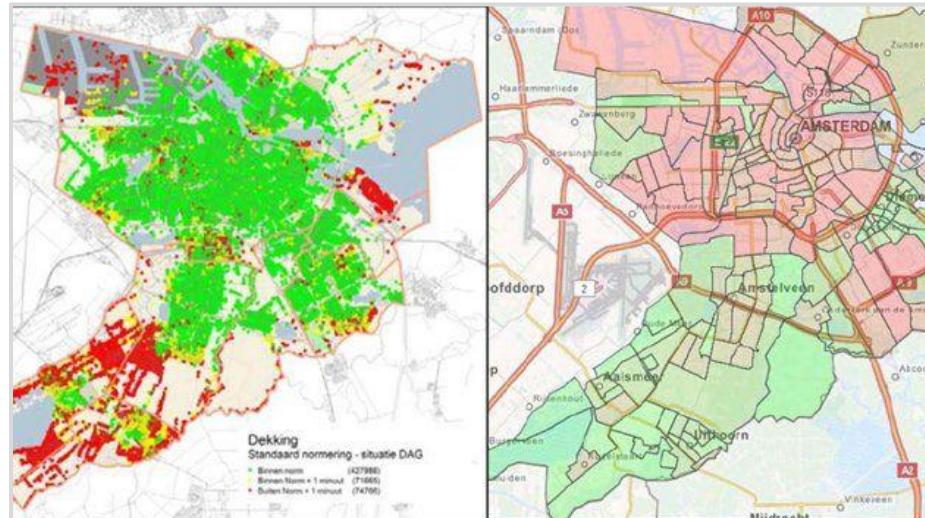
Hazards: Ap - Adult Protection



Dataprodukt eksempler

Brannvesenet i Amsterdam tok data fra forskjellige kilder (informasjon om veier, togspor, bygninger, nabolag osv.) og matchet dem med tidligere hendelser i området.

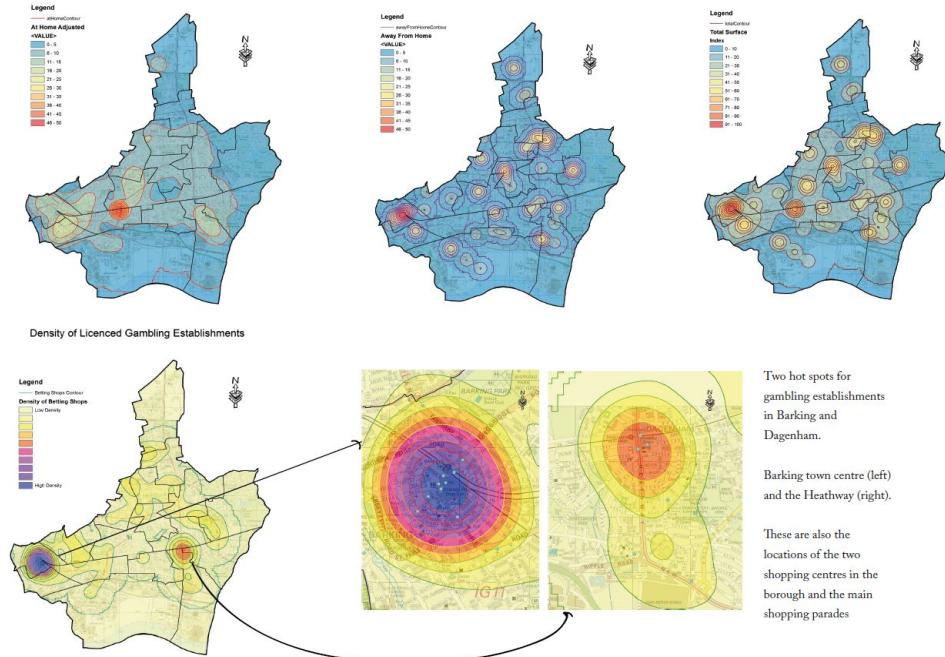
Dataene ble så visualisert gjennom kart som brannvesenet i Amsterdam bruker til å se hvor, når og hvor ofte det oppstår brann



Dataproduct eksempler

Bykommunen Barking & Dagenham i London gjennomførte en analyse for å identifisere områder hvor folk er mer sannsynlig å oppleve gambling-relatert vold.

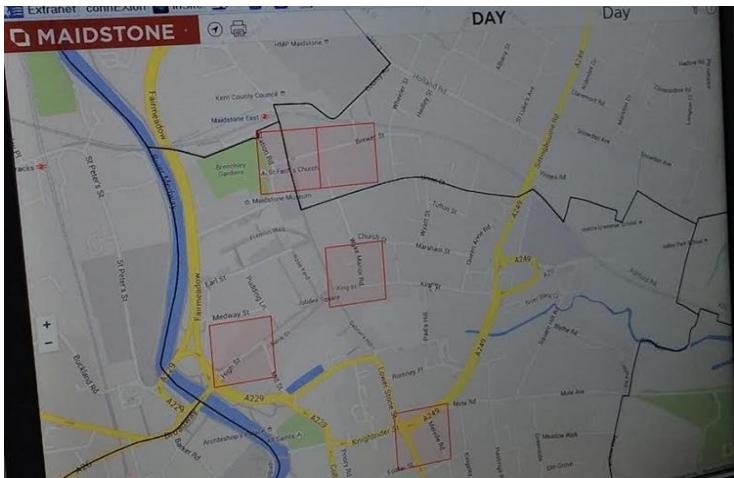
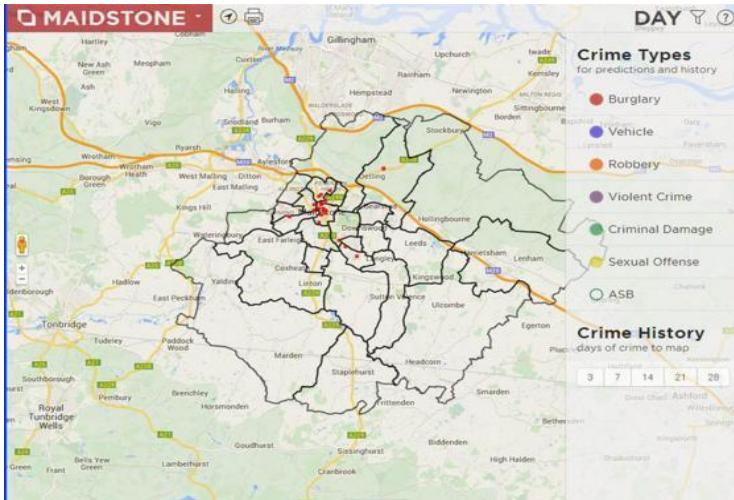
Analysen gav kontekst til revisjonen av de lokale lisensreglene for gambling, og bidro til opprettelsen av en “profil for lokalområdet” som identifiserte viktige klynger og avkrettet antagelsen om at sårbare mennesker og gamblingsteder var spredt rundt i bykommunen.



Dataprodukt eksempler

Politiet i Kent har brukt data om tidligere forbrytelser i nærområdet for å optimalisere allokering av ressurser og forbedre overvåkning.

Dataene inneholder rapporter om lovbrudd i nærområdet fra fem år tilbake i tid. Politibetjentene mottar daglige oppdateringer på 180 "hotspots" i området og kan bruke dem til å bestemme hvilke områder som bør patruljeres mest



Dataprodukt eksempler

Politiet i Durham bruker [HART](#) et dataverktøy som hjelper varetektsbetjenter ta konsistente beslutninger når de evaluerer risiko for fremtidige lovbrudd.

Verktøyet kalkulerer en risikovurdering fra høy til lav, hentet fra data om den mistenktes tidligere forbrytelser sammen med alder, kjønn, postnummer og intelligensvurdering.



Hvilke dataprodukt trenger du?

Du kan nå se om informasjonen fra et av dataproduktene kan muliggjøre en eller flere av handlingene du definerte tidligere:



Problemet vårt er... [sett inn spesifikk problemstilling].



Hvis vi kunne sett/Hvis vi visste... [sett inn hva dataproduktet viser].



Ville vi... [sett inn liste over tiltak du ønsker å implementere].

Hvilke data trenger du?

HOVEDSPØRSMÅL: Hvilke data trenger du for å lage dataproduktet, eksisterer de, har du tilgang til dem, og kan de brukes?

Data kan komme fra flere forskjellige kilder, slik som:

- Åpne data (f.eks. data.gov.uk)
- Offentlig sektor
- Bedrifter og tredjesektor (frivillige organisasjoner)
- Allmennheten

Du kan bruke en enkel mal slik som den på neste lysbilde til idémyldring om hvilke datasett som kan hentes fra de forskjellige kildene.



Et Tydelig
Dataproduct

Åpne data

Innbyggerdata

Offentlig sektor data

Potensielle
databaser

Bedrift/Tredjesektordata

Eksisterer dataene du trenger?

Hvis dataene du trenger til å opprette dataproduktet ditt ikke eksisterer, burde du kanskje tenke over:

- 1) Er det andre datasett som kan bidra med en lignende type informasjon, ellers om kan brukes som en stedfortreder?
- 2) Kan dataene samles slik at analysen er mer gjennomførbar i fremtiden? (Dette er fortsatt et nyttig resultat av ODA prosessen.)



Tilgjengelige
data

Eksisterer dataene du trenger?

I Modul 5 vil vi undersøke hvordan du kan sjekke om dataen er brukbar og, om nødvendig, den kan deles på lovlig og etisk vis.

Foreløpig så er det nok å finne ut om dataene du trenger for å lage dataproduktet er prinsipielt tilgjengelige.

Du burde nå sitte igjen med fire steg som følger:



Tilgjengelige
data



Problemet vårt er... [sett inn spesifikk problemstilling].



Hvis vi kunne sett/Hvis vi visste... [sett inn hva dataproduktet viser].



Ved å bruke disse datasettene... [sett inn hva datasettene du planlegger å bruke].



Ville vi... [sett inn liste over tiltak du ønsker å implementere].

Modul 4:

De 8 fasene av et ODA
prosjekt

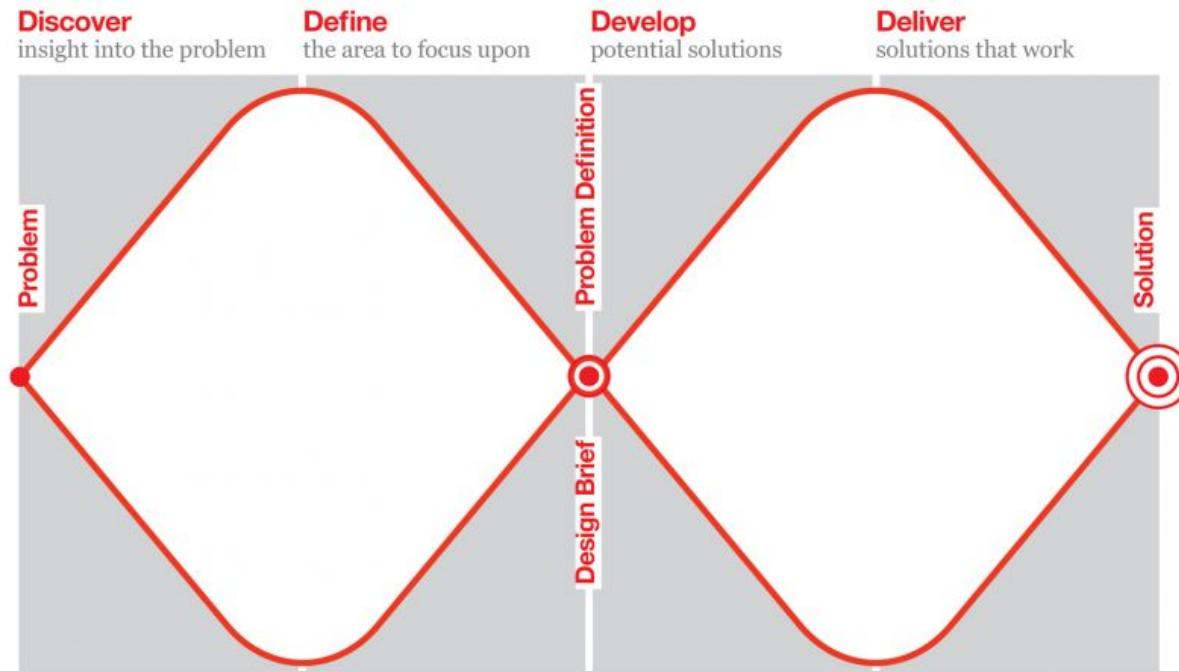
De 8 fasene av et ODA prosjekt

La oss anta at du har brukt den firestegs ODA metoden til å definere en utfordring som du mener kan takles med dataanalyse. For å gjøre prosjektet "live" burde prosjektplanen bestå av åtte kjernefaser:

1. **Forståelse:** vurder prosjektets gjennomførbarhet og raffiner tilnærmingen
2. **Sikre engasjementet til prosjektpartnere:** identifiser hvem som må involveres, deres rolle og ansvarsområder
3. **Informasjonodeling:** opprett avtaler om datadeling
4. **Data innhenting:** innhent de nødvendige dataene
5. **Dataanalyse og prototypbygging:** analyser dataene og bygg den første versjonen av dataproduktet
6. **Testing og evaluering:** test dataproduktet i virkeligheten og mål resultatene
7. **Raffinering:** forbedre dataproduktet basert på tilbakemelding
8. **Skalering:** ta i bruk dataproduktet på en permanent/bredere basis

Stadiene av et ODA prosjekt

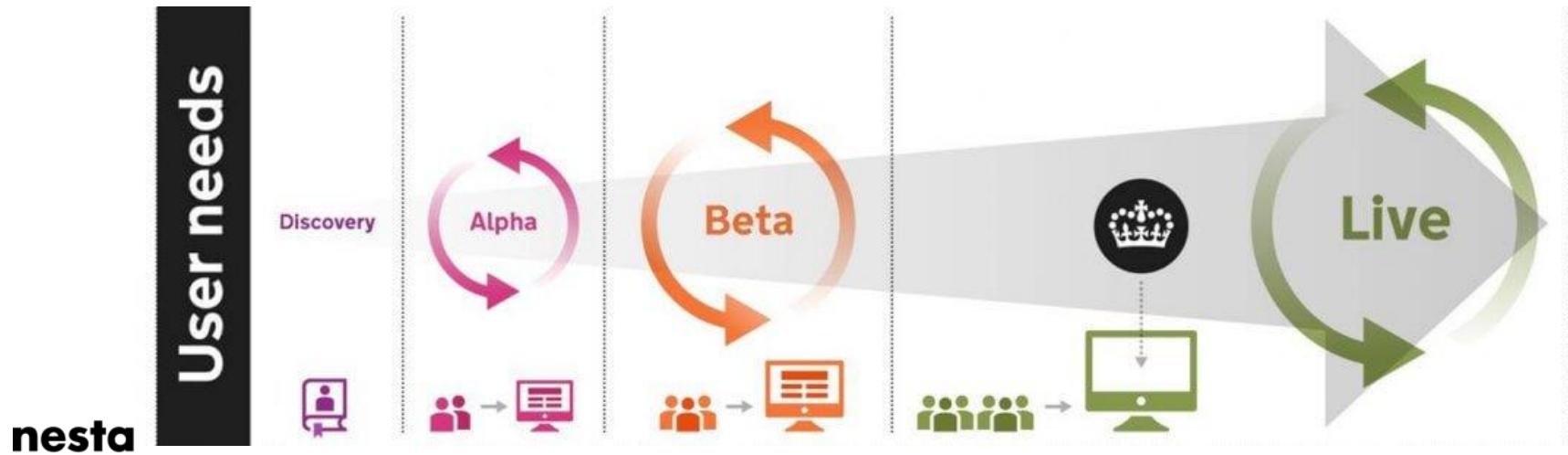
Et ODA prosjekt består av flere stadier som går fra å finne generelle muligheter til å snevre det inn til noe mer spesifikt. Det Britiske Designrådet bruker [[**double diamond diagrammet**](#)] som vist nedenfor.



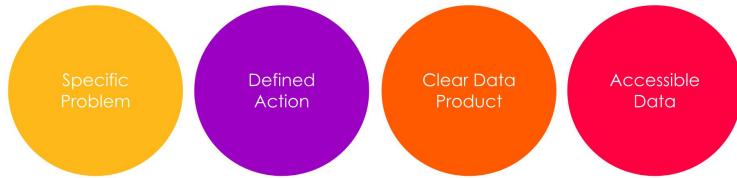
Fase 1: Forståelse

Forståelsesfasen er vanligvis brukt i digitale utviklingsprosjekter. Den sørger for at det rette problemet blir identifisert og verifiserer at den foreslalte løsningen gir mening og er gjennomførbar.

Den digitale tjenesten for den britiske regjeringen har sin egen [[guide til å utføre en forståelsesfase](#)], og bruker diagrammet nedenfor til å vise hvordan den passer med prosjektstadiene.



Fase 1: Forståelse: Fordypning i de fire stegene



I løpet av forståelsesfasen burde du gå grundig igjennom de fire stegene, og forsøke å forbedre formuleringen din.

Noe av dette kan tas via en skrivebordsanalyse, men også inkludere **intervju og workshops** med målgruppen som skal støttes med dataanalyseprosjektet, slik som **kundeansvarlige og førstelinjearbeidere**, og de som vil påvirkes, slik som spesifikke sluttbrukere og grupper. **Samarbeid med disse gruppene i testing og design av potensielle løsninger er utrolig viktig i hvert eneste steg av ODA programmet.**

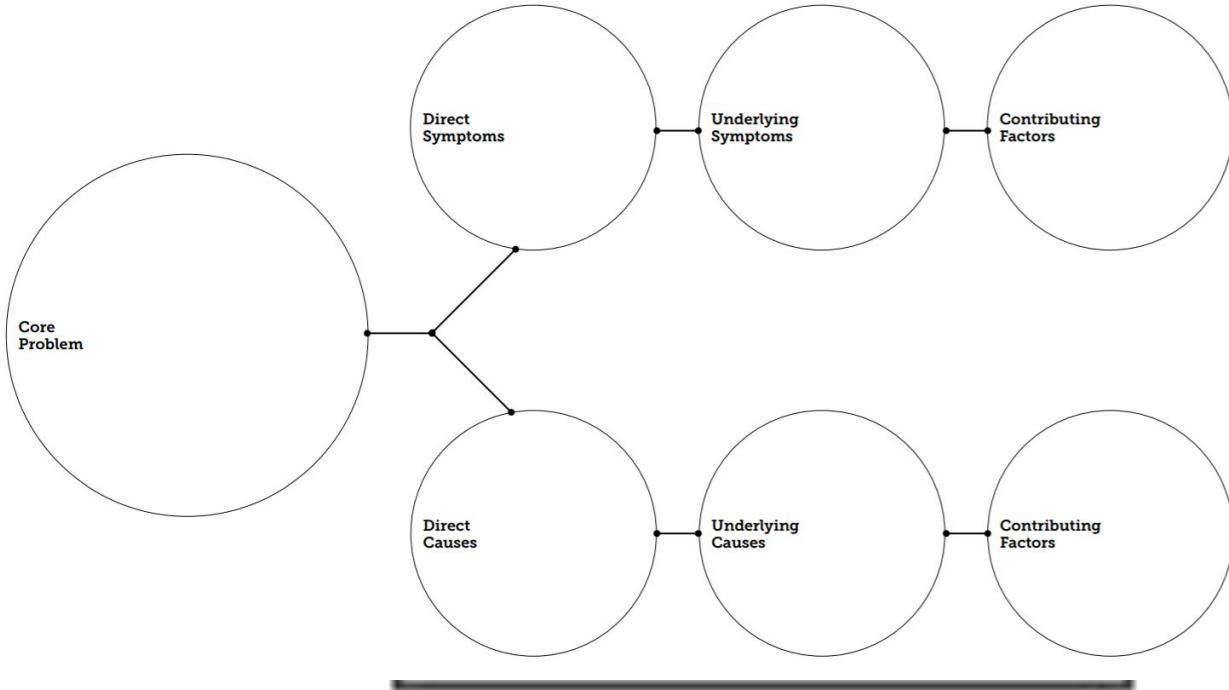
Målet bør være å oppnå en grundig forståelse for:

- 1) Naturen og kompleksiteten av problemet som skal tas
- 2) Bredden på forskjellige tiltak som kan brukes for å løse problemet
- 3) Hvilket dataprodukt ville vært til hjelp for de som skal bruke det
- 4) Om dataene du trenger er tilgjengelig, om de er av god kvalitet, og om det er lovlig og etisk forsvarlig å ta dem i bruk.

Fase 1: Forståelse - verktøy og teknikker

Det er flere verktøy og teknikker du kan ta i bruk for å fordype deg i problemstillingen du har valgt i løpet av forståelsesfasen.

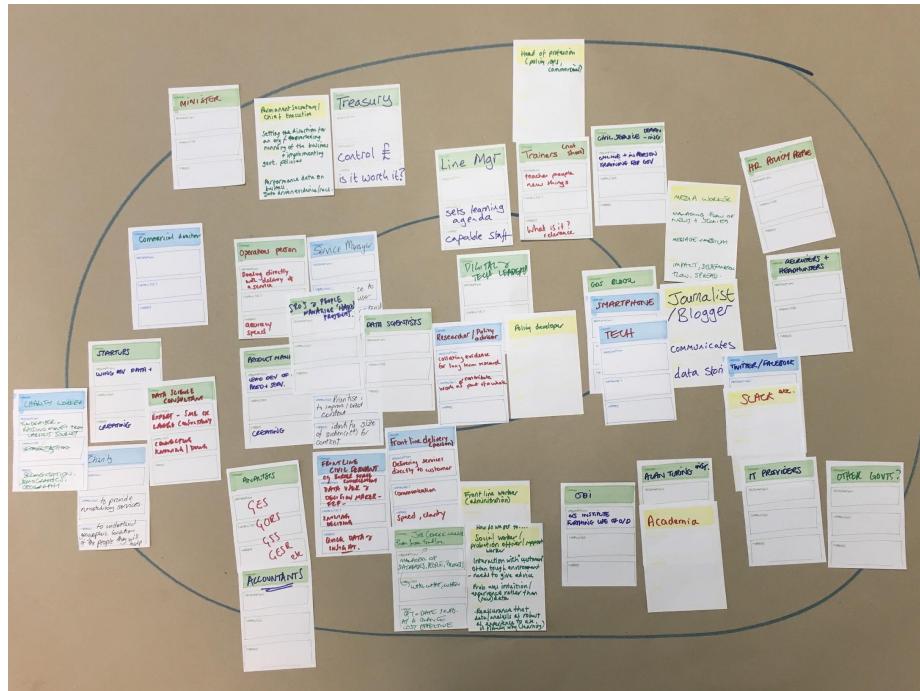
For eksempel, Fem hvorfor teknikken, et Fiskebein diagram og et Årsakssdiagram (se bilde) kan hjelpe deg med å identifisere de medvirkende faktorene av problemstillingen du ønsker å løse.



Fase 1: Forståelse - verktøy og teknikker

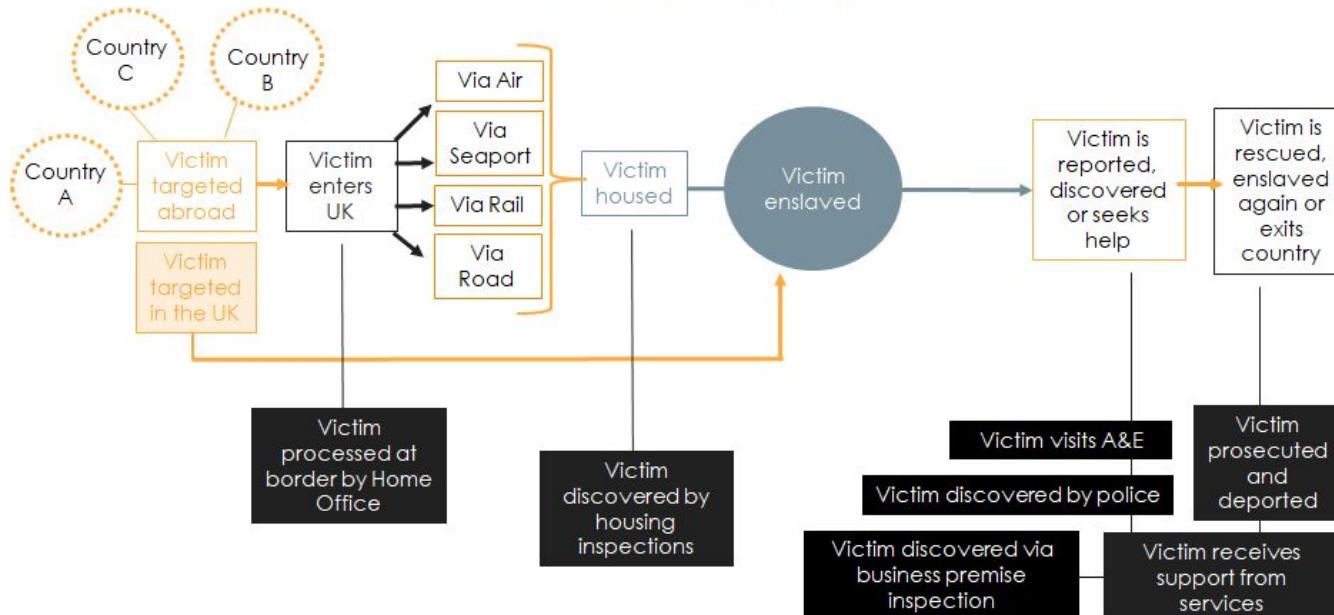
En kartlegging av **aktører** kan hjelpe deg finne virksomheter som er relevant til problemet du ønsker å løse.

Disse virksomhetene kan potensielt være kilder for ekspertråd, flere datasett, eller så kan de til og med være potensielle samarbeidspartnere i dataanalyseprosjektet.



Fase 1: Forståelse - verktøy og teknikker

Opprett et kart over **brukerreisen** som kan hjelpe deg finne punktene hvor den offentlige sektoren kommer i kontakt med et problem. Dette kan hjelpe deg forstå hvilke data som samles og hvor det kunne potensielt blitt designet bedre tiltak. Diagrammet under er et hypotetisk og forenklet eksempel på berøringspunkter innenfor den offentlige sektoren med et offer for moderne slaveri.



Modul 5:

Lovlig og etisk bruk og deling av data

I løpet av forståelsesfasen så er det veldig viktig å sjekke at dataene du trenger kan brukes, og om nødvendig deles, på lovlig og etisk vis.

De fleste lover angående deling av offentlig eide data omhandler **personlig data**. [UK Data Protection Act](#) definerer personlig data som “data som omhandler ett levende individ som kan identifiseres basert på dataene eller dataene i kombinasjon med annen informasjon.”

Loven definerer også **sensitiv personlig data**, som består av personlig informasjon om rase, etnisitet, politisk ståsted, religion, medlemskap i fagforeninger, fysisk og psykisk helse, seksualliv og kriminell bakgrunn. Strengere krav må møtes for å dele sensitive personlige data.

ICO har en [nyttig guide](#) for å anslå hva som telles som personlig data.

Forståelsesfase: Bruk og deling av data og lovlig deling

I mai 2018 introduserte den Europeiske Unionen Personvernfordringen (GDPR, [\[The General Data Protection Regulation\]](#)), som satte høyere krav til bedrifter som samler og bruker personlig data.

Det engelske Datatilsynet (UK Information Commissioner's Office) har en nyttig [Guide til Personvernfordringen \(GDPR\)](#), som inkluderer et [Verktøy for egenvurdering av Datavern](#).

A large red circle is positioned on the right side of the slide. Inside the circle, the text "Tilgjengelig Data" is written in white, sans-serif font, centered horizontally and vertically.

Tilgjengelig
Data

Forståelsesfase: Bruk og deling av data og lovlig deling

Når du har identifisert datasettene du vil bruke, så er det best å utføre en **Personvernkonsekvensutredning (DPIA)**. En personvernkonsekvensutredning er en standard prosess med spørsmål som hjelper brukere å identifisere potensielle risiko eller fordeler ved deling av personlige data.

Samtidig oppfordrer Personvernkonsekvensutredningen brukere å utvikle begrensningsstrategier for å forhindre potensielle baksider ved deling av informasjon.

Denne [[redigerbare personvernkonsekvensutredningen](#)] er basert på informasjon fra Det engelske Datatilsynet.



Tilgjengelig
Data

Forståelsesfase: Bruk og deling av data og lovlig deling

Hvis du må bruke personlig data så er et viktig steg å identifisere de **lovlige inngangene** som gir virksomheten din tillatelse eller autoritet til å følge opp enkelte formål, som kan støttes med deling av personlige data.

For eksempel, under en test for **ODA i London** som ville identifisere ulovlige HMOs identifiserte de også to lover - The Housing Act 2004 og the Crime and Disorder Act 1998 - som viste at de lokale myndighetene var pliktig til å forbedre boligkvaliteten og forhindre kriminalitet og uorden.



Tilgjengelig
Data

Forståelsesfase: Bruk og deling av data og lovlig deling

HOVEDPOENG: Deling av ikke-personlig data har færre vilkår.

Når det er mulig så er det bedre å bruke ikke-personlig identifiserende data. Hvis kildedataene er personlig så kan det være mulig å fjerne navn (og andre identifiserende kjennetegn) og aggregere dataene i en testgruppe som er stor nok til at den ikke lenger er personlig.

Mer informasjon om data anonymisering og pseudonymisering finner du i [Håndbok for forskningsetikk](#).

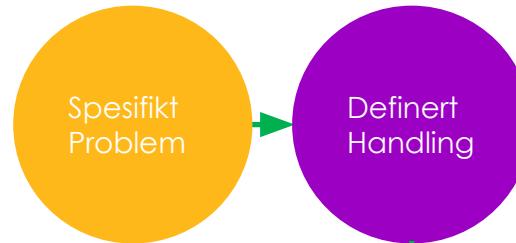


Tilgjengelig
Data

**Dette diagrammet viser
hvor dataproduktet burde
tilpasses basert på om
spesifikke datasett kan
brukes.**

Finnes det en
lovlig inngang til
datadeling, eller
kan dataene
anonymiseres?

Yes



Hvilke data trenger du for å
lage dataproduktet?

- Tenk på:
- Åpen data
 - Offentlig sektor
 - Bedrifter og
tredjesektor
 - Allmennheten

Kan de brukes eller
deles?

No

No

Er dataen personlig?

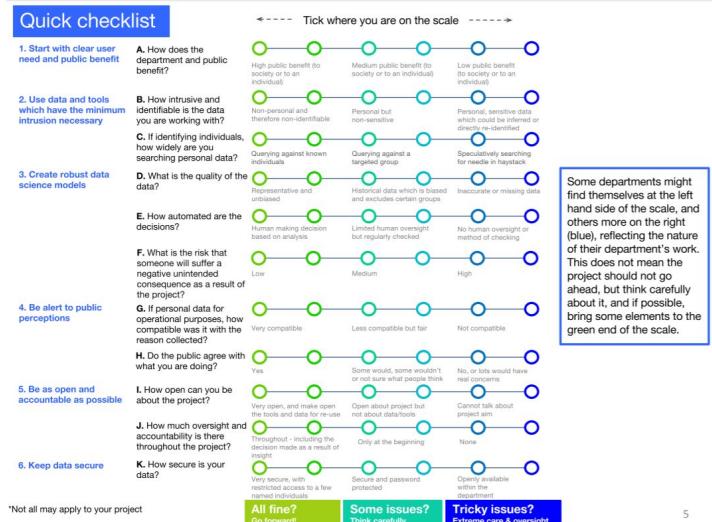
Yes

Hvilke data er
tilgjengelig?

Forståelsesfase: Etisk bruk av data

Uansett om det er lovlig å bruke enkelte datasett så må du også være sikker på at dataanalysen er **etisk forsvarlig**. Etiske vurderinger handler ikke bare om hvilke data som brukes og hvordan de analyseres, men også *handlingene* som dataene muliggjør.

Det finnes flere utmerkede verktøy som kan hjelpe deg tenke over disse spørsmålene, inkludert **The Open Data Institute's Data Ethics Canvas** og the Cabinet Office's Data Science Ethical Framework.



Forståelsesfase: Etisk bruk av data

Samtidig så er Nesta i gang med utviklingen av et verktøy laget spesifikt for egenvurdering hvor dataanalyse blir brukt til å muliggjøre **valg basert på algoritmer**.

[Se utkastet av verktøyet.](#)

Modul 6:

Designe og driftet en ODA
test

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 2-8 av en ODA test

Etter forståelsesfasen, la oss se tilbake på de åtte fasene av et typisk ODA prosjekt:

1. **Forståelse:** vurder prosjekts gjennomførbarhet og raffiner tilnærmingen
2. **Sikre engasjementet til prosjektpartnere:** identifiser hvem som må involveres, deres rolle og ansvarsområder
3. **Informasjondeling:** opprett avtaler om datadeling
4. **Datainnhenting:** innhent de nødvendige dataene
5. **Dataanalyse og prototypbygging:** analyser dataene og bygg den første versjonen av dataproduktet
6. **Testing og evaluering:** test dataproduktet i virkeligheten og mål resultatene
7. **Raffinering:** forbedre dataproduktet basert på tilbakemelding
8. **Skalering:** ta i bruk dataproduktet på en permanent/bredere basis

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 2: Sikre engasjementet til prosjektpartnere

Mot slutten av forståelsesfasen burde du ha god nok oversikt til å se hvilke team og virksomheter som burde involveres, og kan sikre deg samarbeid med dem.

Det kreves flere forskjellige roller:

1. Prosjektleder
2. Brukerundersøkelser
3. Dataforskning
4. Teknisk (DAPI og MVP oppbygging)
5. Lov og informasjonsstyring
6. Implementering
7. Dataleverandører

Grunnfunksjonene til de forskjellige rollene er som følger:

- 1. Prosjektleder** - ser til prosjektdesignet og koordineringen av alle partnerne og deres aktiviteter.
- 2. Brukerundersøkelse** - utreding av de reelle behovene til de som dataprosjektet skal hjelpe
- 3. Dataforskning** - samling, pussing, matching og analysering av data for å gi insikt.

4. Teknisk (DAPI og MVP oppbygging) - sette i stand nødvendige verktøy for å laste opp og analysere dataene; utvikle prototypen for dataproduktet.

5. Lov og informasjonsstyring - forsikre at dataene blir brukt på lovlig og etisk vis.

6. Implementering - virksomhetene eller teamene bruker dataene til sine tiltak

7. Dataleverandører - virksomheter leverer data for å lage dataproduktet

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 2: Sikre engasjementet til prosjektpartnere

Et Omforent Memorandum (**MoU - Memorandum of Understanding**) bør signeres av alle parter for å stadfeste engasjementet de har til dataanalyseprosjektet.

Et MoU fastslår hvilke roller hver enhet påtar seg, deres ansvar, berøringspunkt, og ressursene de vil fordele og hvordan de vil delta og kommunisere i prosjektet.

Når dette er gjort kan du gå videre til neste fase:
Informasjonsdeling.

Fase 3: Informasjonsdeling

Når man deler data mellom partnere, uansett om det angår et fåtall med team innenfor en lokale myndigheter eller flere virksomheter innenfor den offentlige sektoren som jobber sammen, må det opprettes regler og vilkår gjennom en Dataoverføringsavtale, også kjent som en **“Information Sharing Protocol - ISP”**.

En ISP er nødvendig når personlige data blir delt, men er like fullt anbefalt når det deles ikke-personlig identifiserende data.

De essensielle elementene som dekkes av en ISP er:

- Hensikten bak delingen
- Potensielle mottakere og under hvilke omstendigheter de vil ha tilgang
- Nøyaktig hvilke data som deles
- Kvaliteten på dataene - nøyaktighetsgrad, relevans, brukbarhet, osv.
- Datasikkerhet
- Bevaring av felles data
- Individets rettigheter - prosedyrer for etterspørsel av tilgang, spørsmål og klager
- Evaluering av effektivitet/opphevelse av delingsavtalen
- Tiltak hvis individuelle ansatte skulle unnlate eller bryte avtalen

Maler er tilgjengelig via ICO

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 4: Datainnhenting

Fase 4 handler om å faktisk få tilgang til dataene du trenger for å lage dataproduktet ditt. Tilgang til datasett fra flere parter kan ta lang tid, siden virksomheter ofte har forskjellige prosedyrer for uthenting.

I løpet av forståelsesfasen skal du ideelt allerede ha hentet eksempler av dataene for å vurdere kvaliteten og for å forstå karaktertrekkene.

“Data innhenting” kan bety å laste opp et datasett (f.eks. et regneark) eller koble sammen datavitenskapplattformen din med IT systemene til de forskjellige leverandørene for å få tilgang til dataene kontinuerlig gjennom et programmeringsgrensesnitt, (**API - Application Programming Interface**).

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 5: Dataanalyse og prototypbygging

I denne fasen analyserer dataforskerne dataene og jobber med tekniske utviklere for å bygge den første versjonen av dataproduktet.

Dataforskerne kan velge fra et stort utvalg av teknikker til å analysere dataene basert på typen data som brukes og hvilket resultat man er ute etter.

Noen ganger vil de bruke en prosess kjent som **maskinlæring**, hvor datamaskinen kan “lære” naturen til et problem basert på opplæringsdata, for å gjenkjenne mønster og sammenhenger eller forutsi flere slike hendelser.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 5: Dataanalyse og prototypbygging - Maskinlæring

Som et eksempel, et maskinlæreremiddel kunne analysert tidligere finansielle transaksjoner for å se etter svindel. I denne typen prosess vil du typisk sett ha bruk for tre datagrupper:

- 1) Data basert på tidligere kjente saker av det du prøver å finne (f.eks. tidligere oppdaget svindel)
- 2) Data relatert til karaktertrekk av det du er ute etter
- 3) Data relatert til karaktertrekk av det du ikke prøver å finne

Tenk på det som en sporthund som må trenes i hvilke lukter som må meldes fra om, og hvilke lukter som kan ignoreres.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 5: Dataanalyse og prototypbygging

Basert på typen problem som identifiseres kan dataforskerne anvende en passende dataanalyse teknikk:

Hvordan
budsjetteres
dette i år basert
på flere
konkurrerende
faktorer?

Hva er de mest
indikative
tegnene på
svindel?

Hvilke barn har
større risiko for å
falle utenfor
utdanning, arbeid
eller opplæring?

Hvilke lokale
autoriteter er
demografisk lik?

Numerisk
prognose

Hvor markant
er tegnet

Kategoriprog
nose

Gruppering

Nevralt
nettverk

Logistisk
regresjon

“Random
Forests”

K-means

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 5: Dataanalyse og prototypbygging

Når det kommer til analysering av data og bygging av nyttige produkter så er det flere ressurser som kan hjelpe med god fremgangsmåte.

The [Aqua book](#) inneholder nyttig veiledning om kvalitetssikring av dataanalyse.

Matt Upson har skrevet artikler om bygging av opplagte og reproducerbare dataprodukter:

<https://dataingovernment.blog.gov.uk/2017/11/27/transforming-the-process-of-producing-official-statistics/>

Og hvordan de kan gjøres bærekraftig:

<https://software.ac.uk/blog/2018-05-21-why-government-needs-sustainable-software-too>

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 6: Testing og evaluering

Formålet med testing- og evalueringfasen er å prøve ut dataprodukt og tiltak i den ekte verden og måle resultatene.

For eksempel, i løpet av testingen for et ODA i London ble førstelinjen av bygningsinspektører gitt en liste med eiendommer som dataanalysen indikerte kunne være en ulovlig HMO.

Formålet med denne prosessen er både å teste hvor brukbart dataproduktet er, og samle inn resultater om hvorvidt det har hjulpet tiltaket.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 6: Evalueringens viktighet

Evaluering blir ofte oversett eller gjort i all hast på slutten av prosjektet. Dette er en feiltakelse.

Det er kjempeviktig å utføre en evaluering fordi du kan se hva som faktisk virker, som øker sjansen for at det er de beste tiltakene som raffineres og skaleres.

Det tilbys håndbøker som går inn i dybden på gode evalueringer via Nesta ([Retningslinjer for forskningsbevis](#)), sentralmyndighetene ([Magenta Book](#), [Green Book](#)), og brukervennlige nettsider slike som [betterevaluation.org](#).

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 6: Evalueringens viktighet

Ifølge [Magenta Book](#), den essensielle håndboken om evaluatingsdesign i den offentlige sektoren fra myndighetene i Storbritannia, så finnes det tre brede evaluatingskategorier: **prosess, innvirkning og økonomi.** Hvilken kategori som er mest egnet kommer an på hva du ønsker å lære om prosjektet ditt, samt ressursene og ekspertisen som du kan dedikere til evalueringen.

Generelt sett vil **evalueringsprosessen** hjelpe deg med å forstå hvordan og hvorfor tiltaket har en innvirkning, mens **innvirkningsevalueringen** vil måle forandringen som har skjedd og om det forandret seg på grunn av tiltaket ditt. Etter en grundig innvirkningsevaluering så vil en **økonomisk evaluering** være mulig for å tjene på resultatene som ble observert.

Fase 6: Testing og evaluering

Når du begynner å tenke på evaluering (som burde skje veldig tidlig i prosjektet ditt) så er det verdt å tenke over de forskjellige **Bevisstandardene**.

Bevisstandardene er veiledere som hjelper deg bestemme hvor trygg du kan være på resultatene av evalueringen.

Desto høyere standarden, desto grundigere bør metodologien og desto sikrere burde du være på at det er det datainformerte tiltaket og ingen andre faktorer som ligger til grunn for den observerte forandringen.

Level 5

You have manuals, systems and procedures to ensure consistent replication and positive impact

Level 4

You have one + independent replication evaluations that confirms these conclusions

Level 3

You can demonstrate causality using a control or comparison group

Level 2

You capture data that shows positive change, but you cannot confirm you caused this

Level 1

You can describe what you do and why it matters, logically, coherently and convincingly

Fase 6: Anvende bevisstandardene

Basert på hvilket nivå du velger så er det forskjellige krav. La oss se på de første tre nivåene av bevisstandardene.

Level	Forventning	Hvordan kan beviset produseres?
1	Du kan forklare hvorfor tiltaket ditt kan ha en innvirkning og hvorfor dette vil være en forbedring	Samle brukertilbakemelding eller inspeksjon etter tiltaket; se over den eksisterende forskningen.
2	Du kan vise at en forandring har skjedd, men ikke at tiltaket ditt var årsaken	Samle data før og etter tiltaket gjennom brukerundersøkelser eller dyptgående intervju.
3	Du kan demonstrere at tiltaket ditt er årsaken bak forandringen.	Bruk av grundige eksperimentelle metoder med kontroll- eller sammenlikningsgrupper, som isolerer innvirkningen.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 6: Evalueringens viktighet

Bruk av **Eksperimentelt Design / Randomiserte Kliniske Forsøk** (RCTs) er den beste måten å vise at tiltaket ditt var årsaken til en innvirkning eller forandring. Deltakere deles tilfeldig inn i to grupper hvor en opplever tiltaket (behandlingsgruppen) og den andre ikke (kontrollgruppen).

Når deltakere ikke kan (randomiseres) kan et **quasi-eksperimentelt design** være et alternativ. Det har den samme strukturen som en RCT, men istedet for tilfeldig allokering så kan du velge å sammenligne to eller flere like grupper.

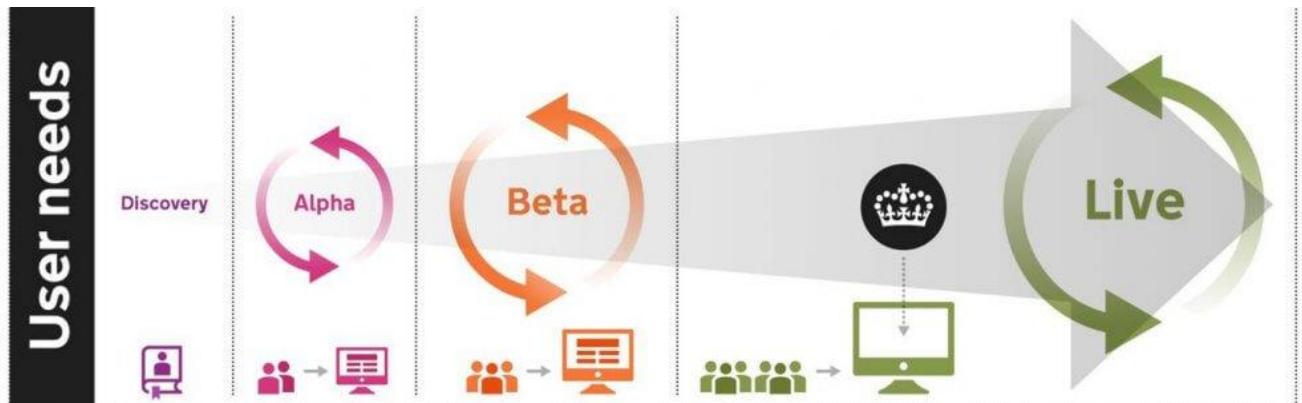
Nesta sin **Innovation Growth Lab** tilbyr et [Experimenteringsverktøy](#) som forklarer eksperimenteringsprosessen på en tydelig måte. [Hypoteseverktøy](#) fra **the Experimentation Hub** er enda et verktøy som lar brukere raskt formulere en hypotese, estimere varigheten på et eksperiment, og bestemme et passende stikkprøveomfang.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6**
- Modul 7
- Modul 8

Fase 7 & 8: Raffinere og skalere

Et ODA prosjekt vil ideelt sett tatt en rask fremgangsmåte: lage prototyper i korte innspurter, for å så teste og forbedre basert på tilbakemeldinger.

Syklusen burde repeteres til dataproduktet og tiltaket er raffinert nok til at det kan skaleres opp eller legges inn i hverdagen.



Modul 7:

Hvordan å gjøre det
enklere neste gang

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- **Modul 7**
- Modul 8

Modul 7: Hvordan å gjøre det enklere neste gang

I løpet av Nesta sine ODA prosjekter så har vi lært mye om hva som fungerer og hva som ikke fungerer i dataanalyseprosjekter innenfor den offentlige sektoren.

Videre vil vi se på noen av de viktigste tingene vi har lært. Hvis du følger denne seksjonen burde det gjøre det lettere for deg å drifte dataanalyseprosjekter i fremtiden.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7**
- Modul 8

Vurdere hvor klar virksomheten din er for et dataprosjekt

Erfaringsnivå med data varierer innenfor forskjellige virksomheter. For å hjelpe deg forstå hvor virksomheten din sitter på spektrumet for øyeblikket så har Nesta utviklet et **datanivå rammeverk**.

Før du starter et stort dataprosjekt, investerer i ny teknologi, eller utvikler en strategi for å få mer ut av dataene dine, så er det verdt å undersøke hvilke styrker og hvilke svakheter virksomheten din har.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- **Modul 7**
- Modul 8

Nesta sitt Datanivå Rammeverk

Datanivå rammeverk ser på 5 områder:

- 1. Databehandling:** Hvordan blir dataene samlet, organisert og hvordan får man tilgang? Hvor nøyaktig og gammel er dataene?
- 2. Data Ledelse og Tilgjengelighet:** Hvor lett er det å dele dataene? Er datadeling oppfordret gjennom klare protokoller og lederskap? Hvordan behandler autoritetene åpne data?
- 3. Databruk:** Hvordan blir dataene brukt i beslutningsprosessen? Blir de regelmessig brukt til å optimalisere service og følge opp prestasjon?
- 4. Dataegenskaper og Dyktighet:** Hvilket nivå av dataferdigheter eksisterer i virksomheten? Hvilke deler av virksomheten sitter det meste av evne og kapabilitet?
- 5. Virksomhetens Datakultur:** Til hvilken grad forstår og verdsetter personene i virksomheten data-informerte beslutninger

Phase 6: Testing and evaluating

Du kan prøve en beta-versjon av vår nettbaserte datanivå egenvurderingsverktøy, utviklet i samarbeid med LGA.

The screenshot shows a web-based assessment tool titled "Data Maturity". At the top, there's a navigation bar with links to "Local Government Association", "LG Inform", "LG Inform Plus", "Data", and "Standards". Below the navigation, a purple header bar contains the text "Start" and "Questions > Your results". A vertical sidebar on the left lists categories: "Use", "Culture", "Management", "Skills", "Governance", and "Finish". The main content area features a large title "Data Maturity" and the subtitle "Rate your organisation's data management skills". A descriptive paragraph explains the tool's purpose: "This tool is designed to help you make an honest assessment of how advanced your organisation is at dealing with data. You can compare your assessment with others from your own organisation, and from around the country." Another paragraph states: "It is intended to stimulate discussion, give ideas as to how you might improve and provide pointers to case studies and resources that will help. It will not be used to rank people or organisations." A third paragraph notes: "The tool was launched in April 2018 as a prototype and your responses, along with general feedback, will be used to improve the tool in future years." At the bottom, a section titled "How your data will be used" includes a note: "You need to register and sign in to agree to the data usage policy, to save your responses and to see resultant improvement advice."

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- **Modul 7**
- Modul 8

Tips for å forbedre datakvaliteten

HOVEDPOENG: Det er først når dataene tas i bruk at det vil vise seg om kvaliteten er bra nok og om det er verdt å forbedre dem. Ikke la perfekt være fienden til god!

Her er to tips for å forbedre datakvaliteten:

Sett i stand prosedyrer for å sikre at ren data samles fra begynnelsen av og tren ansatte deretter. Når mulig, samle data elektronisk og fra rullegardinlister heller enn åpne tekstfelter.

Rett opp feilaktige felt i kildedataen for å unngå å måtte rette opp de samme feilene igjen og igjen når dataene overføres til andre systemer. Husk å dokumentere hvor rádataen ble overført fra og hvordan de ble rettet, så andre kan gjenta prosessen når det trengs.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- **Modul 7**
- Modul 8

Tips for å forbedre datakvaliteten

HOVEDPOENG: Virksomheter som ikke kan tilpasse seg og sette sammen datasett fra deres egne IT systemer får det vanskelig med å samarbeide med andre på dataprosjekter.

Matching og sammensetting av data er prosessen av å finne og fastslå sammenhenger mellom individuelle dokumenter på tvers av ulike datasett. Når de blir satt sammen lagres de integrerte dataene i en index eller et datavarehus.

For eksempel så linker **Innbygger Indexen** til bykommunen Camden i London personer og stedsdata fra forskjellige kilder innenfor de lokale myndighetene for å utvikle en 'gyllen' oversikt over innbyggere. [Lenke: Dette har forbedret tverrfaglig handling og gjort det enklere å finne tegn på svindel].

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7**
- Modul 8

Matching, sammensetting og integrering av data

Tabellen under demonstrerer hvordan data fra fem forskjellige kilder kan samles sammen til en presis 'gyllen' oversikt.

FIELDS	Source A	B	C	D	E	GOLDEN
Full Name	Fred Smith	Freddie Smith	Frederick Smith	Fred Smith	Fried Smith	Fred Smith
DOB	10-1-1995	01-01-1900	10-1-1995		10-1-1995	10-1-1995
Gender	M	M	M		M	M
Address	10 The Grange Camden NW1 0AA	Flat A 10 The Grange Camden NW1 0AA	10 The Grange Camden NW1	Flat A 10 The Grange Camden NW1 0AA	10 Grange Camden NW1 0AA	Flat A 10 The Grange Camden NW1 0AA
Ref	1007689		1007689	1007699	1007689	1007689
NINO	AB123456B					AB123456B
Home Tel		0231-987-7866	0205-937-996			0205-937-996
Mobile	07988674707	077134563				07988674707
Email		fsmith@v.com	smithy@bt.com			fsmith@v.com
Car Reg	FG12KJK		LA56YBB			FG12KJK

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7**
- Modul 8

Matching, sammensetting og integrering av data

Den enkleste måten å gjøre det enklere å matche data på tvers av forskjellige IT systemer er å konsistent bruke gjennomgående kjennemerker.

For stedsbasert data burde den offentlige sektoren i Storbritannia bruke unike stedsreferansenummer (UPRN – Unique Property Reference Number). Ved hjelp av UPRN:

- **Barnsley** har frembrakt en ekstra skattbar verdi på totalt £170 000 for kommersielle eiendommer ved å bruke datamatching muliggjort av UPRN.
- LLG teamet i **Nottingham** har spart £40 000 gjennom et nærmere samarbeid med Bedriftskatt teamet og Verdivurderingskontoret i Nottingham.
- Gjennom rasjonaliseringen av valglokaler har **Northumberland** funnet en måte å spare £39 000 per valg, og identifisert 300 adresser som ikke ble skrevet opp i velgerregisteret.

Kilde: <https://www.geoplace.co.uk/addresses/uprn>.

Se videre informasjon om fordelene med UPRN i [denne bloggen](#) fra Ordnance Survey.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- **Modul 7**
- Modul 8

Matching, sammensetting og integrering av data

Persondata (eller 'kunde' eller 'klient') er mer komplekst å integrere enn stedsdata fordi folk flytter inn og ut av områder, og kan være kjent med forskjellige navn av forskjellige personer.

Forestill deg en situasjon hvor John Nicholson (et barn) er en kjenning av den lokale sosialomsorgen og har hatt sammenstøt med kriminalomsorgen for ungdom og skolevelferd teamet. John Nicolson (stavet forskjellig) er en kjenning av politi- og sykehustjenester. Hvis styret ikke har en effektiv løsning for identitetsmatching så vil ikke senere henvisninger til sosialomsorgen ha hele bildet, fordi det finnes separate og uferdige saksmapper. Det vil kanskje tas feil beslutninger for å sikre tryggheten til John.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7**
- Modul 8

Matching, sammensetting og integrering av data

Enheter i den offentlige sektoren kan prøve å manuelt matche individer på tvers av forskjellige systemer basert på visse regler (**deterministisk matching**) men dette kan være veldig tidskrevende.

En annen metode kan være å bruke **verktøy for sannsynlighetsmatching**, som kan hjelpe deg spare tid og forbedre kvaliteten på matchingen.

Deterministiske systemer passer godt med matching av mindre mengder data med færre kompleksiteter, mens verktøy for sannsynlighetsmatching er kraftigere og gir bedre innsikt inn i større datasett.

Hva er forskjellen mellom deterministiske og sannsynlighetsbaserte matchingsystemer?

	Deterministiske matchingsystemer	Sannsynlighetsmatching systemer
Hva er det?	Programmert til matche to eller flere dokumenter basert på et sett med forhåndsdefinerte regler. For eksempel så kan en regen instruere systemet til å sammenligne to dokumenter med det samme NHS nummeret, selv om adressene ikke stemmer overens.	Bruker statistisk teori og dataanalyse for å fastslå sannsynligheten for at to dokumenter er relatert. Skaper en prosentandel, f.eks. 80% for å indikere sannsynligheten for en match.
Anvendelse	Fungerer best med mindre datasett hvor relativt få andre dokumenter (omtrent 20 millioner) skal matches.	Passer bedre i tilfeller hvor det må matches dokumenter med større mengder data og flere databaser.
Finansiering og evner	Kan implementeres til en lavere pris når mengden datasett, datakjennetegn og matchingregler er mindre. Krever avansert Excel-kompetanse.	Kan kreve investering i nye programvarer, men er potensielt mer kostnadseffektiv på lang sikt for større databaser i vekst. Krever erfaring med programmeringsspråk som R og Python, samt en forutsigende analyse og maskin læring.
Skalering	Kan være arbeidskrevende og kostbart akkommodere en voksende database. Funksjonen svekkes etter hvert som det legges til mer tilpasning og flere regler.	Kan lettere absorbere, lære og tilpasse seg til større mengder med data, som forbedrer funksjonen generelt.

Ledere i den offentlige sektoren må lage plass og kultur for data for å gjøre en forskjell.

nesta

Ledere i den offentlige sektoren må sette forventingen at det er uakseptabelt å ta viktige beslutninger eller reformere en tjeneste uten å være klar over hva dataene viser.

De må også anerkjenne at innenfor virksomhetene deres så kan de allerede ha dataanalysere som er frustrerte dataforskere. De må frigjøres fra månedlig arbeid på dashboard og KPI rapportering og må i stedet gis tid til å jobbe med kundeansvarlige og førstelinjearbeidere.

Modul 8:

Sette opp et Office of Data Analytics

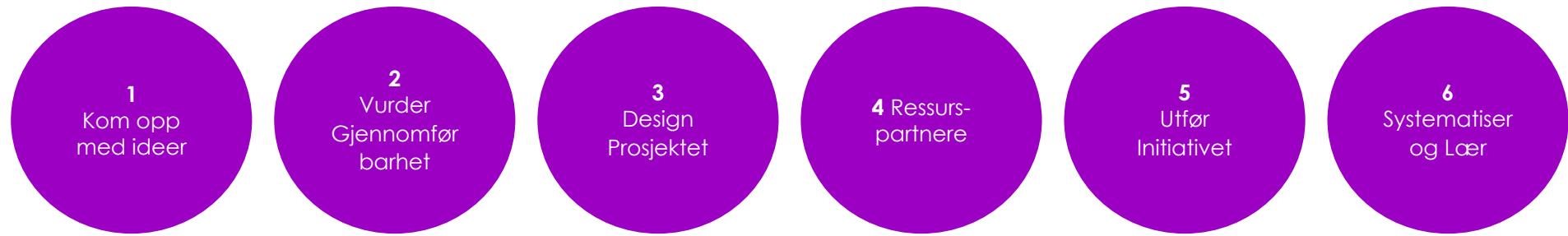
- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8**

Hva er den ideelle modellen for et Office of Data Analytics

Om du har nådd punktet hvor du ønsker å lage en permanent funksjon som lar virksomheten din samarbeide med andre på dataanalyseprosjekter, hvordan ville den funksjonen sett ut?

Office of Data Analytics Modellen

Slik vi ser det så er ikke et ODA prosjekt best sett på som et team med dataforskere, men som en funksjon som ser til og styrer en prosess ende-til-ende som består av seks steg:



Denne metoden forsterker ODA funksjonens egen kapasitet ved å bruke talentene til den bredere offentlige sektoren i en by eller region, eller digitale og tekniske sektorer.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- **Modul 8**

1 – Kom opp med ideer



Ideer for nye dataanalyseprosjekter kan gjøres via nettdugnad fra flere forskjellige steder, for eksempel:

- 1) **Lokale politikere** – f.eks. tatt fra programerklæringer
- 2) **Lokale myndigheter/ansatte i den offentlige sektoren** – via workshops
- 3) **Allmenheten** – potensielt hentet fra e-demokrati plattformer slik som [Prioriteringene dine](#).

ODA metoden kan tilby flere vinduer gjennom året hvor ideer kan hentes inn.

En åpen prosess slik som denne demonstrerer hvordan ODA er designet for å være nyttig for alle i regionen, og ikke for å erstatte arbeidet til individuelle virksomheter i den offentlige sektoren.

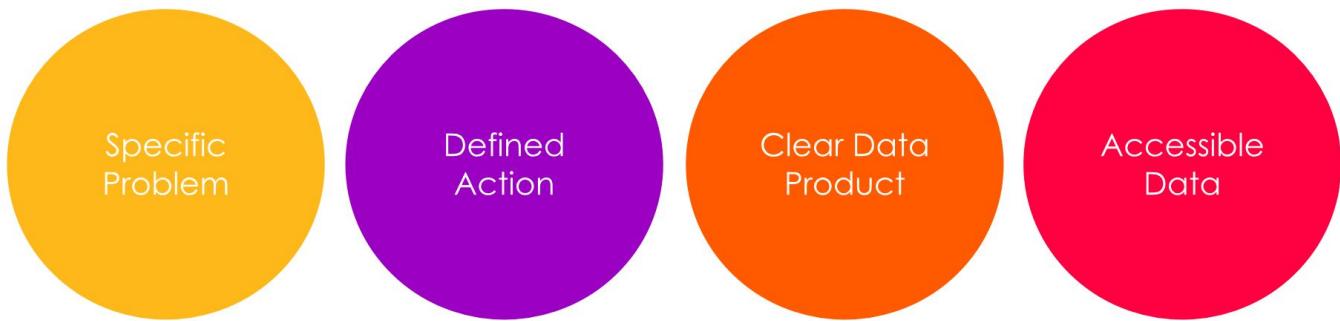
- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- **Modul 8**

2 – Vurder Gjennomførbarheten



Gjennomførbarheten til ideene må avklares slik at de kan godkjennes som *tiltak muliggjort av dataanalyse*.

For å gjøre dette, ODA funksjonen kan bruke de fem problemtypene og fire-stegs metoden som ble forklart i Modul 3.



(Spesifikt Problem) (Definert Handling) (Tydelig Dataproduct) (Tilgjengelige Data)

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- **Modul 8**

3 - Design Prosjektet



Når problemet og det ønskede resultatet (inkludert kriterier for suksess) har blitt tydelig definert, kan en av flere prosjektmetodologier velges, inkludert:

- 1) **Eksperimenter:** ODA funksjonen designer og administrerer alle deler av prosjektet (slik som datasamling, analyse, evaluering, osv.) i samarbeid med partnere i en test.
- 2) **Utfordningspremie:** ODA funksjonen tilbyr en premie til virksomheten eller personen som kommer med den beste løsningen muliggjort av data til et definert problem.
- 3) **Hackathon:** ODA funksjonen organiserer et ‘hackathon’ for problemer på en mindre skala hvor det er mye data, eller som en første fase av Eksperimenter eller Utfordningspremie designmetodene.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8**

4 - Ressurspartnere



ODA funksjonen evaluerer **teamets interne kapasitet** og identifiserer hull som kan dekkes med datavitenskap, teknologi, prosjektadministrasjon og juridisk ekspertise som trengs til hvert prosjekt.

Ved å bruke nettverket sitt med det bredere innovasjonsøkosystemet **søker ODA funksjonen partnere til å fylle hullene**. Partnere kan inkludere lokale myndigheter, lokale teknologibedrifter, små- og mellomstore datavitenskap bedrifter, akademiske institusjoner, ekspertgrupper, borgerlige dataekspertgrupper, osv.

Partnerne kan bli med som **frivillige, eller ansettes** via et simplifisert rammeverk for ansettelse, satt opp av ODA funksjonen.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- Modul 8**

5 – Utfør Initiativet



Dataprosjektet kan dermed utføres. For en typisk ODA test slik den blir beskrevet i guiden så vil det vanligvis involvere åtte prosjektfaser slik de ble beskrevet i Modul 6:

- 1. Forståelse:** vurder prosjektets gjennomførbarhet og raffiner tilnærmingen
- 2. Sikre engasjementet til prosjektpartnere:** identifiser hvem som må involveres, deres rolle og ansvarsområder
- 3. Informasjonodeling:** opprett avtaler om datadeling
- 4. Datainnhenting:** innhent de nødvendige dataene
- 5. Dataanalyse og prototypbygging:** analyser dataene og bygg den første versjonen av dataproduktet
- 6. Testing og evaluering:** test dataproduktet i virkeligheten og må resultatene
- 7. Raffinering:** forbedre dataproduktet basert på tilbakemelding
- 8. Skalering:** ta i bruk dataproduktet på en permanent/bredere basis

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- **Modul 8**

6 – Systematiser og Lær



Siste fasen er for at ODA funksjonen skal forstå innvirkningen av hvert prosjekt og dele erfaringene sine. Dette inkluderer sannsynligvis:

- **Evaluering av innvirkning** mot suksesskriterier utviklet i steg 3. Dette kan potensielt inkludere grundige evalueringsmetoder som Randomiserte Kontroll Tester.
- **Dele erfaringer** i regelmessige blogger gjennom prosjektet, i en siste prosjektrapport og gjennom seminar og arrangement.
- **Gi ut koder, maler, verktøy** og andre materialer som kan hjelpe andre i å bruke det samme arbeidet eller å designe noe nytt.

- Introduksjon
- Modul 1
- Modul 2
- Modul 3
- Modul 4
- Modul 5
- Modul 6
- Modul 7
- **Modul 8**

Hvordan ODA fungerer

Godt oppsummert så har ODA funksjonen to essensielle oppgaver:

1. **Å handle:** gjennomføre 2-3 prosjekter i året basert på den seks-stegs prosessen forklart i denne guiden.
2. **Lage en mal:** lag gjenbrukbare regler for etikk, datastandarder, juridiske dokumenter, prosesshåndbøker og åpne kildeverktøy.

Om ressursene er tilgjengelige så vil man helst legge til to ytterlige funksjoner:

1. **Samling og Koordinering:** fungere som et samlepunkt for regionens datavitenskap og policy miljø; støtte og ernære økosystemet.
2. **Oplæring:** gjennomføre workshops spesifikt rettet mot behovene til ledere, datavitenskap praktiserende, om juridiske, data og metoder.



Dataanalyse i den Offentlige Sektoren

A Nesta Guide

Eddie Copeland
Tom Symons
Hilary Simpson
Nevena Dragicevic

nesta