

VÄRDESKAPANDE TJÄNST

Tekniska paradig och externa faktorer – Delrapport AP3

Glenn Erikson, Elias Arnestrand

Mars 2016

PARTNERS



FINANSIÄR



Innehållsförteckning

TEKNISKA PARADIGM	4
Big data	4
Internet of Things	5
Fortsatt läsning.....	5
Länkad data och semantisk webb	6
SPARQL.....	7
Fortsatt läsning.....	8
Länkad data inom kollektivtrafikområdet.....	8
Plattform för länkad öppna data	8
Data kvalitet, standard och format	9
Metadata	10
 SAMVERKAN KRING ÖPPNA DATA	 11
Samverkan mellan dataägare	11
Utbyte mellan dataägare och dataanvändare	11
Utbyte mellan dataägare.....	13
Dataförmedlare	14

SAMMANFATTNING DELRAPPORT

Delrapporten för arbetspaket tre fokuserar på dels tekniska paradigmskiften som eventuellt kan förändra eller påverka hur öppna data tillhandahålls. Dels förklaras Big data kortfattat och vilka skillnader- likheter det har med öppna data. Länkad data och semantisk webb, om hur det kan länka ihop data globalt genom att använda gemensamma beskrivningar och kategorisering vad data innehåller. Några presenteras några pågående eller nyligen avslutade forskningsprojekt som syftar till att förenkla publicering och tillgängliggörande av öppna data.

Andra delen av delrapporten förklara några icke tekniska frågor om hur samarbetet mellan dataägare och mellan tredjepartsutvecklare ser ut idag. Delar av denna redogörelse kommer från E-delegationens arbete som avslutades 2015. Avslutande stycke beskriver en ny aktörsroll som agerar som dataförmedlare och kopplar ihop dataägare med olika målgrupper.

Tekniska paradigmm

Big data

Det finns olika definitioner av Big data och innebörden av begreppen har också förändrats över tid. En källa som ofta används är Wikipedia som definierar begreppet ”dataset som är så stora att traditionella data hantering och behandlingsverktyg är inadekvata”. En tidig definition som myntades av Gartner Group är de tre V:na, volym (*volume*), snabbhet (*velocity*) och variationsrikedom (*variety*). Beroende på hur man ser öppna data, kan det ses som en delmängd av Big data enligt bild 1. I dag består öppna data till stor del av statisk eller aggregerad data som enkelt kan användas och analyseras av traditionella dataverktyg. Exempel på statisk data är tidtabeller, vägnät och där ekonomiska rapporter, scannade beslut anses vara aggregerad data. Motsatsen till aggregerad data är rådata, vilket är data som finns tillgängligt i sin ursprungliga struktur och detaljnivå. Rådata passar in på definitionen av Big data, förutsatt att volymerna är så stora att det inte går använda traditionella data hanteringsverktyg för att göra rapporter och analys.

Däremot finns det andra strömmar av data som börjar göras tillgängliga som både har stora volymer och finns i sitt ursprungliga format. Exempel på dessa strömmar är data från väderstationer, fordonssensorer, vägsensorer och andra mätstationer som skickar kontinuerliga strömmar av rådata.

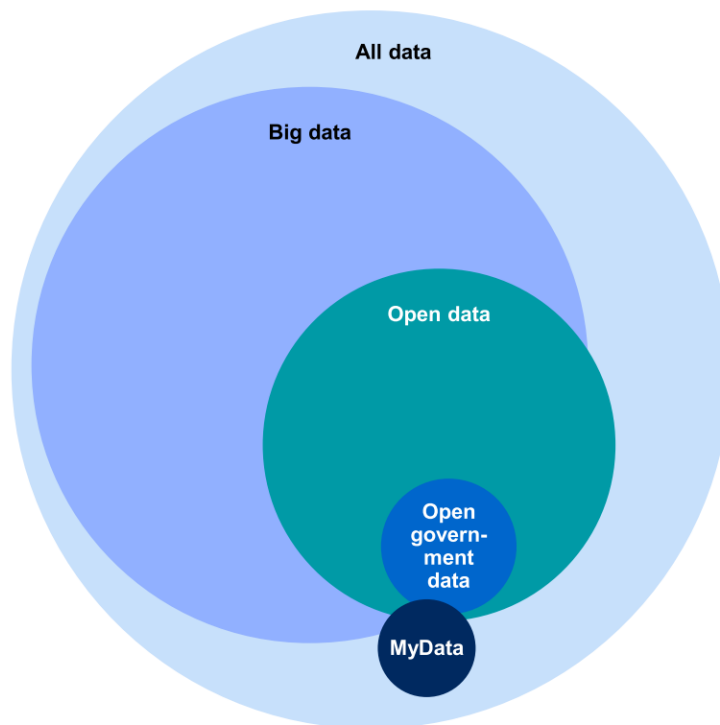


Bild 1: Källa McKinsey Global Institute

En användbar tumregel för att klassificera om data kan benämnas Big data är om datamängden är tillräcklig stor (se tidigare definition) och om den har låg informationstäthet, det vill säga data som i sin obehandlade form inte avslöjar kausala samband, beroenden eller skapar en djupare insikt. Dataset som avslöjar samband kan vara resultat av statistisk analys eller redan aggregerad data i form av rapporter och har oftast hög informationstäthet. Enligt ovanstående tumregel är exempelvis öppna data som innehåller färdiga rapporter, tidtabeller eller trafikhändelser enkelt att tillgodogöra sig och har en hög informationstäthet. Ytligare en skillnad är att öppna data som finns tillgängligt i dag oftast tagits fram med ett syfte att användas interna inom verksamheten eller erbjuda data inom specifika domänområden.

Om arbetet med att tillhandahålla öppna data styrs enbart av avgöranden angående vilken data som är användbart eller inte, kan detta ha negativ inverkan på innovationsmöjligheten och potentialen för tjänsteutvecklare att utvinna värde från datakällor för att skapa unika tjänsteerbjudande som redan inte finns på marknaden. En för tydlig styrning av tillgänglighetsförändring av öppna data utifrån ett nyttoperspektiv skulle göra öppna data mindre användbart i innovationssyfte. När öppna data erbjuds i form av informationsprodukter efterliknar de mer vanliga websidor eller nyhetsflöden än fenomenet Big data. Tidigare studier¹ med tredjepartsutvecklare som använder Trafikverkets öppna datatjänster tyder på att tillgång till rådataströmmar från flera olika aktörer skapar möjligheter för tjänsteutvecklare att erbjuda unika kunderbjudande som möjliggör bättre förutsättningar för livskraftiga tjänster.

Internet of Things

En eventuellt stor källa till data och information inom när framtid är *Internet of Things* (IoT) där allt ifrån bilar, kylskåp, bankomater, patientvård på sjukhus, kommunala soptunnor, är bara några av möjligheterna där IoT kan appliceras. Genom att kombinera IoT med RFID (*radio-frequency identification*) finns möjligheterna att skapa sensorer med unik identifikation vilket själva kan skicka operativ och status relaterad data. Dessa tekniska lösningar är öppna standarder och relativt billig teknik som kan vara en källa till kontinuerliga dataströmmar som kan publiceras via öppen data och bidra till innovation potentialen.

Fortsatt läsning

Blogg om IoT och RFID;
<http://blog.atlasrfidstore.com/internet-of-things-and-rfid>

Forbes artikel om IoT och smarta städer;

¹ Värde av öppna data, ur ett tredjepartsutvecklarperspektiv & Öppna data, detaljer av värdeskapande för tredjepartsutvecklare

<http://www.forbes.com/sites/bernardmarr/2015/05/19/how-big-data-and-the-internet-of-things-create-smarter-cities/>

Länkad data och semantisk webb

Länkad data och semantisk webb möjliggör sammanlänkning av data från olika system, vilket möjliggörs genom att beskrivning av data, så kallad metadata, ingår som en naturlig del av data. Data saknar oftast dessa beskrivande egenskaper eftersom det tar tid att komplettera med betydelsefulla beskrivningar och att ägaren har kunskap om dess betydelse. Problemet uppstår när andra som inte varit med att ta fram data eller känner till verksamheten, skall tyda innebörden. Exempelvis kan det vara avsaknaden av beskrivning av vilket måttenhet som används eller kryptiska namn på kolumner. För att uppskatta vilken nivå av förklarande egenskaper data har finns det ett klassificering² system för öppna data som är framtaget av grundaren till internet Tim Berners-Lee. Klassificerings system har följande fem nivåer.

★ - Gör din information tillgänglig på webben under en öppen licens

(även svårbearbetade format som skannade dokument är ok)

★★ - Gör informationen tillgänglig som strukturerad data

(t. ex., Excel-format istället för en bild av en tabell)

★★★ använd icke-proprietära format

(t. ex., CSV istället för Excel)

★★★★ - Använd URI:er för att identifiera ting,

och RDF för att uttrycka påståenden om dem

★★★★★ - Länka data till andras data, vilket ger sammanhang

Om öppna data, offentlig data och andra källor kompletteras med dessa egenskaper möjliggör det ett nytt paradigmskifte för hur vi använder webben, där vi går från att publicera information på webbsidor till att publicera data i maskinläsbart format. Vad det innebär i praktisk är att vi går ifrån att tillhandahålla information anpassade för människor till att göra den läsbar för maskiner. Som en del i detta utnyttjas format som är bättre lämpade för maskiner, exempelvis XML och JSON. För att förklara egenskaper hos data används metadata som beskriver dess beskaffenhet genom att använda gemensamt vokabulär för termer som nyttjas inom ett domänområde, detta benämns ontologisk grund. Länkad data bygger på länkning till en ontologi som kan tillämpas inom ett eller domänområden. Exempelvis för att tillföra metadata till data om olika regioners befolkningsantal, kan länkning till en redan existerande ontologi för regioner. Det som styr vilken ontologi som används är

² [Fem stjärnigt utvecklings schema för öppen data](#)

en den unik länk som finns med i RDF beskrivningen. Definition av termer för länkad data är användarstyrkt, vilket möjliggör att det kan kompletteras eller skapas nya om det saknas. Standarden som möjliggör globalt sammanlänkning av data kallas *Resource Description Framework* (RDF). Ramverket RDF byggs upp av tre element som benämns som RDF trippel, subjekt, objekt och predikat. Bild 2 visar ett exempel av en RDF uttryck beskrivning av en region, för att mer ingående beskrivning av hur länkad data och semantisk webb fungerar, se länkarna i slutet av stycket.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:region="http://www.country-regions.fake/">
  <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Oxford">
    <dc:title>Oxford</dc:title>
    <dc:coverage>Oxfordshire</dc:coverage>
    <dc:publisher>Wikipedia</dc:publisher>
    <region:population>10000</region:population>
    <region:principaltown rdf:resource="http://www.country-regions.fake/oxford"/>
  </rdf:Description>
</rdf:RDF>
```

Bild 2: Exempel på RDF metadata

För att hantera och lagra sammanlänkad data används grafbaserade databaser vilket erbjuder flexibilitet för att dynamisk definiera relationer för att koppla ihop objekt, subjekt och predikat, vilket är en förutsättning för att skapa en semantisk modell. För att tillföra semantisk betydelse till länkad data finns det standardiserade dokumentstrukturer, kallat scheman, vilket möjliggör beskrivning av klassificering och relationer för att definiera ontologier. Dels finns RDFS vilket är ett dokumentschema som tillhör RDF för att definiera enklare vokabulär, och för mer uttryckfulla ontologier finns scheman som *Web Ontology Language* (OWL) vilket också baseras på RDF syntax. Genom att skapa en semantisk klassificering beskrivs hur objekt, subjekt och predikat är kopplade till varandra och utgör en ontologi.

SPARQL

För att ställa frågor på publicerad data som är definierad med RDF används SPARQL som liknar vad SQL är för relationsdatabaser. Det som skiljer däremot är att syntax även fungerar som ett protokoll, då data vanligtvis paketeras tillsammans med frågeställningar. SPARKQL möjliggör frågeställningar på data som är definierade enligt RDF standarden och koppla ihop olika datakällor som använder samma ontologi. Grunden i syntaxen bygger på samma element som RDF, subjekt, predikat och objekt. Exempelvis på frågor som kan ställas via länkade datakällor kan ge svaret på alla skolor i Sverige, Norge och Finland som exempelvis har 1800 elever eller mer från stadsdelar där den årliga medelinkomsten är lägre än 180 000kr. Det som skiljer frågan från vanlig så kallad SQL join, är att

frågan kan ställas globalt mot vilken data källor som helst förutsatt att data om skolor och inkomstfördelning är publicerad på kända URI adresser och att samma ontologi används.

Fortsatt läsning

Guide till grunder om länkad data och semantisk web;
<http://www.linkeddatatools.com/semantic-web-basics>

Introduktion till länkad data;

<http://lankadedata.se/vitbok/introduktion.html>

W3.org beskrivning av semantiska webben;
<http://www.w3.org/standards/semanticweb/>

Länkad data inom kollektivtrafikområdet

Många av de datamängder inom trafikområdet som finns tillgängliga idag som öppna data skulle tjäna på att publiceras som länkad data. Data tillgängliggörs från en rad olika aktörer idag som skulle dra nytta av att använda en gemensam ontologisk grund. Ett exempel på detta är hur hållplatser och stationer beskrivs av olika aktörer idag, där det finns olika identifiering och beskrivningar för saker kopplat till hållplatser. Dessutom finns flera olika namn för samma hållplats/station. Bakgrunden till denna situation är att olika organisationer sitter med olika tekniska system som utvecklats separat genom åren utan koordinering av benämning och kategorisering av saker kopplade till en hållplats. När nu informationen tillgängliggörs från dessa system visar det sig att det uppstår svårigheter med att koppla ihop dessa datakällor på grund utav att alla aktörer har skapat egna kategoriseringar. Ett exempel på hur man framgångsrikt hanterat denna problematik är Englands departement för transport som etablerat en nationell standard för hållplatser som idag används av alla aktörer i ekosystemet. Däremot har man valt använda en annan standard än RDF för att definiera egenskaper av data.

Fortsatt läsning

Dokumentation för publika nationell stationer och hållplatser i England;

<https://www.gov.uk/government/publications/national-public-transport-access-node-schema>

Plattform för länkad öppna data

Begreppet *Data and Platform as a service (DaPaas)* är en kombination av plattform och data som service (... *as service*) som är ett forskningsprojekt finansierat av EU, för att förenkla återanvändandet och ökad användningen av publicerad länkad data som bland annat använder RDF standaren. Projektet syftar till att minska barriärer för mindre och medelstora offentliga organisationer att bidra med öppen länkad data.

Fortsatt läsning

DaPass projektets hemsida
<http://project.dapaas.eu/>

DaPass vertyg för att publicera och anpassa länkade data
<https://datagraft.net/>

W3C rapport om DaPaas
http://www.w3.org/2014/03/lgd/papers/lgd14_submission_21

Data kvalitet, standard och format

Datakvalitet har olika betydelse baserat på syfte och målsättning. Att försöka definiera datakvalité utan att betrakta avsikten och användningsområde för data är nästintill en omöjlig uppgift. Detta hänger delvis ihop med problemet att skapa ett dataformat som fungerar för alla syften och alla tänkbara användare. Datakvalitet är kopplat till frågan om standard och format, eftersom det beror på krav och behov för en tilltänkt användargrupp. Därför är det viktigt att differentiera datakvalité och att det kan betyda olika saker för olika användare och deras tilltänkta målsättningar. Exempelvis är ett kalkylblad ett användbart format för finansiell data och bilder av scannade beslut ett användbart format för publicering av offentliga handlingar. En kvalitéspekt som är generellt oavsett användningsområde, är att data är aktuell och publiceras så nära i tid som möjligt från det att det skapades. Dock är implikationer av tidsenlig data olika beroende sitt syfte, exempelvis är nyttan av aktuell väderdata större än offentliga handlingar.

Ett sätt att förhålla sig datakvalitet är att anpassa behovet av exakthet och precision till sitt ändamål. Exakthet är hur väl data återger verkligheten. Exakthet i öppna data kontext är hur väll ett format kan återge det ursprungliga värdet eller betydelsen. Exempelvis kan en dåligt scannad bild vara tillräckligt för en människa att läsa, men vara omöjlig för maskiner att återskapa. Därav är val av ett lämpligt format viktigt för att uppnå en högre nivå av exakthet. Däremot är det en avgångning mellan nivån av exakthet och kostnaden för att återskapa verkligheten i ett maskinläsbart format. Precision är sambandet av detaljer som återges. En rapport med aggregerade resultat gör det svårt för användaren att avgöra precisionen av data eftersom detaljer saknas så användaren själv kan genomföra operationen för att skaffa sig en uppskattning. Generell praxis är att data skall erhållas i sitt ursprungliga format, om data transformeras till andra enheter och standarder kan det vara svårt för användare att härleda och använda data. Däremot kan data behöva harmoniseras för att passa eller användas tillsammans med andra datakällor. Data med hög precision har nära samband eller är data som kan korreleras med hjälp av standarder och mätenheter. Hög precision erhålls oftast i dataformat som erbjuder hieratisk indelning av data som till exempel XML dokument.

Metadata

Inga data är komplett om inte tillhörande metaddata finns tillgängligt. Metaddata beskriver när den skapades, vem som skapade det, ursprungskällan, när det senast ändrades med mera. Vilket gör det lättare för användare att avgöra var data kommer ifrån, vem som skapade, vilka enheter som används för numeriska värden med mera. Metaddata hjälper till med att öka spridningen och användningen av öppna data då det förklarar vem som är upphovsman och vilka licensvillkor som gäller för vidareutnyttjning. W3C har publicerat ett utkast till ett antal principer för metaddata på webben för att förespråka en bättre klassificering och ökad återvändning av data.

DCAT

DCAT är en RDF baserad vokabulär för att indexera publicerad data och möjliggör användning av distribuerade datakällor. DCAT beskriver egenskaper av data, vad det innehåller, när det publicerades och av vem, geografiskt område, adress till datasetet med mera. Men DCAT specificerar inte i vilket format som data görs tillgängligt i och inte heller vilken licenser som gäller för innehållet, däremot innehåller det licens för själva kataloginformationen. DCAT-AP och DCAR-AP är en version av DCAT för en gemensam europeisk datakatalog, som delvis drivs av Vinnova, *Open Knowledge Foundation* och Meta Solutions.

Fortsatt läsning

En definition av datakvalitet;
<https://opengovdata.io/2014/data-quality/>

Best practice för datapublicering från W3C;

<http://www.w3.org/TR/2015/WD-dwbp-20150224/>
<http://www.w3.org/TR/2015/WD-dwbp-20150224/#bp-template>

Teknisk beskrivning av DCAT från W3C, samt DCAT-AP för europeiska dataportaler från EU kommissionen;

<http://www.w3.org/TR/vocab-dcat/>
https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Samverkan kring öppna data

I dag finns det problem i samverkan mellan myndigheter som hanterar öppna data. Dels beror detta på avsaknaden av strukturer för samverkan om vad och hur data skall tillgängliggöras. Det finns också stora skillnader mellan dataägare och tredjepartsutvecklare vilket försvårar samarbete och förståelsen av varandras problem. Frågeställningar kommer delvis från e-delegationens³ arbete som avslutades sommaren 2015 kring hanteringen av PSI-direktivet. Stycket avslutas med en utblick på en typ av aktörroll som opererar mellan dataägare och tredjepartsutvecklare som benämns dataförmedlare eller datajägare.

Samverkan mellan dataägare

Från både dataägare och tredjepartsutvecklare lyfts ofta behovet av en tätare dialog för att hitta former strukturerad samverka mellan parterna. Från en dataägars perspektiv handlar detta dels om en ökad samverkan med andra dataägare som brottas med liknande utmaningar, dels om ett utökat utbyte mellan dataägare och befintliga eller potentiella användare av den information som finns tillgänglig eller kan tillgängliggöras.

Utbyte mellan dataägare och dataanvändare

Dataägare och utvecklare är på många sätt väldigt olika typer av aktörer. Dialogen mellan aktörerna präglas av dessa olikheter och utmaningen att få till ett välfungerande samspel kräver sannolikt en ökad förståelse för varandras arbetssätt, strukturer och drivkrafter. Bild 3 beskriver översiktligt (och förenklat) de sammanhang aktörerna kommer från. I de flesta fall råder ett ojämlikt förhållande mellan dataägare och tredjepartsutvecklare, där det är väldigt svårt för utvecklare att få till stånd en förändring hos dataägare. Det finns också en skillnad vad gäller takt i utvecklingen hos aktörgrupperna. För en dataägare kan det till exempel vara en process på uppemot 2 år att få till stånd en ny datakälla, då arbete behöver planeras i verksamhetsplaner och koordineras mellan förvaltningsobjekt. För en tredjepartsutvecklare kan tidsfönstret för att arbeta med en viss datakälla vara ett par veckor, och ofta finns inte utvecklaren kvar som ”kund” till dataägaren under hela processen med att ta fram nya datakällor.

³ <http://www.edelegationen.se/>

Utvecklare	Dataägare
<ul style="list-style-type: none"> • Agil utveckling • Stort tekniskt kunnande • (Ofta) små organisationer eller privatpersoner • (Ofta) försöks- eller pilotverksamhet • Snabb utvecklingstakt 	<ul style="list-style-type: none"> • Stora organisationer • Komplexa beslutsvägar • Komplex IT-infrastruktur • Krav kring IT-säkerhet, robusthet • Krav att som myndighet uppträda korrekt och fatta välgrundade beslut

Bild 3: Förenklad hur verksamheten skiljer sig mellan offentliga dataägare och tredjepartsutvecklare

De organisationer som tillhör kategorin dataägare har ofta en tradition av att arbeta mot tydliga målgrupper eller kundgrupper. I arbetet med öppna data blir detta arbetssätt ofta problematiskt eftersom det upplevs som svårt att förstå behoven hos ”kundgruppen” tredjepartsutvecklare som representerar en bred grupp med olika motiv och drivkrafter. Samtidigt finns det ibland en ovilja hos dataägare att djupare analysera behoven eftersom man anser att offentliga aktörer inte ska lägga sig i hur och vad öppna datakällor används till. De arenor som idag finns för utbyte mellan dataägare och utvecklare bygger i stor utsträckning på engagemang från eldsjälar hos dataägarna. Ofta blir enskilda personer i de offentliga organisationerna kända ingångar för utvecklare som kan få hjälp att lotsa sina behov vidare inom organisationen. Aktiviteter som stödjer detta utbyte är till exempel hackatons och tävlingar. Inom vissa områden finns även ett mer systematiskt arbete för att stimulera utbytet och dialogen mellan dataägare och tredjepartsutvecklare, ett exempel är arena är Trafiklab för kollektivtrafikdata. Inom ramen för detta initiativ arrangeras löpande så kallade *meet-ups* för att knyta kontakter mellan dataägare och tredjepartsutvecklare. Motsvarande arbete sker inom vissa regioner där det genomförs aktiviteter för att stimulera mötet mellan dataägare och tredjepartsutvecklare. I ett inledande skede, innan en tjänst finns etablerad, krävs information och kontaktytor för att informera om tillgängliga datakällor, förstå möjligheter- begränsningar, samt affärsmässiga konsekvenser av till exempel licensvillkor. I ett skede när en tjänst är etablerad finns dessutom ett behov av en relation kring frågor som rör exempel driftsfrågor, inkrementella förbättringar med mera.

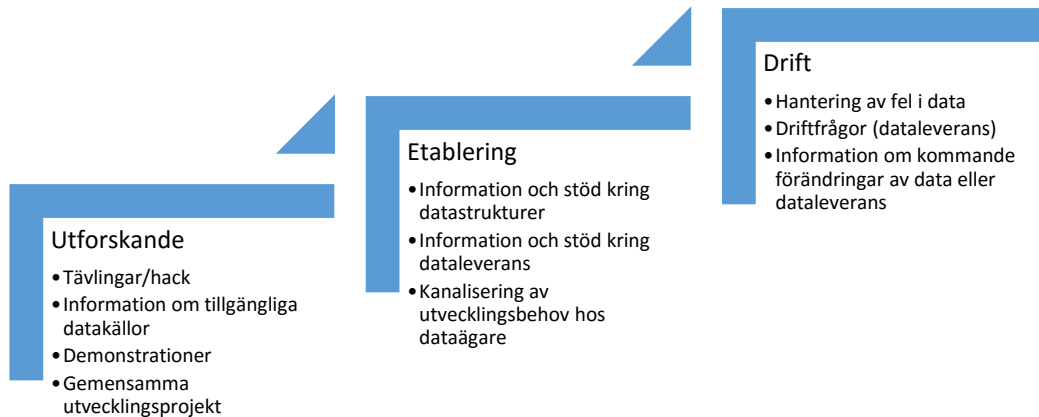


Bild 4: Exempel på olika behov i relationen mellan dataägare och tredjepartsutvecklare baserat på olika faser

På motsvarande sätt kan en dataägare eller specifika datakällor befinna sig i olika tillstånd vad gäller utveckling. I ett tidigt skede kan exempelvis en dataägare behöva ökad kunskap och förståelse för behoven kring en viss datakälla, medan andra datakällor befinner sig i en drift och förvaltningsfas, enligt bild 4.

Utbyte mellan dataägare

Arbetet med öppna data inom offentliga organisationer i Sverige präglas på många sätt av en utforskande verksamhet. Det har i flera olika sammanhang efterfrågats en ökad samordning mellan organisationerna för att inte gå åt olika håll i utvecklingen. Det finns även ett starkt behov av att lära av varandra och utbyta erfarenheter kring vad som varit framgångsrikt respektive mindre framgångsrikt. Det finns olika nivåer av samverkan mellan offentliga organisationer som lyfts fram i olika sammanhang. Inom vissa tematiska områden har man kommit långt i arbetet med samordning medan det i princip helt saknas samordning inom andra områden. Ofta har dessa olikheter sin grund i tidigare arbeten och standardisering inom området, inte som en direkt konsekvens av arbete med ”öppna data”. Ett exempel är samverkan inom geodataområdet.

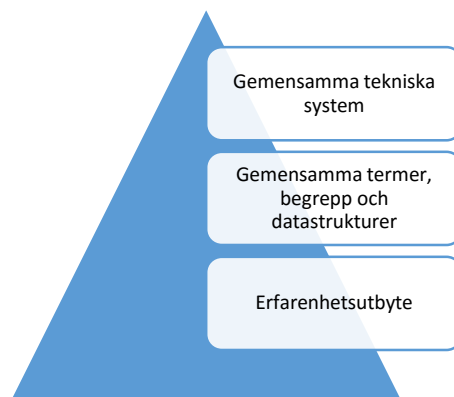


Bild 5: Olika nivåer av samverkan mellan offentliga dataägare inom geo-data

Inom transportområdet sker en viss samordning kring kommunikation och tredjepartsrelationer genom Trafiklab. Det finns dock ingen aktör med ett tydligt ansvar eller mandat kring informationsmodell. Det saknas även ett utpekat ansvar vad gäller helhetsansvar för leveransen inom domänen, då ansvaret idag sträcker sig över en rad olika aktörer med olika former av roller och styrning. Exempelvis kan nämnas Trafikverket, Regionala kollektivtrafikmyndigheter, Samtrafiken, Trafikanalys och Transporstyrelsen.

Dataförmedlare

Internationellt börjar det dyka upp i diskussioner ett behov av en roll i öppna data värdekedjan som kan förmedla (*data broker* på eng.) och koppla ihop medborgare, tredjepartutvecklare med myndigheter som ansvarar och tillhandahåller data. Exempel på detta är när så kallade medborgare (*civic hack* på eng.) grupper som organiserar sig för att lösa samhällsutmaningar. Sådant engagemang skapades under snöstormen som passerade amerikanska östkusten vintern 2016, som förlamade flygtrafiken och lede till avstängning av vägar. Där medborgare organiserade sig för att hjälpa myndigheterna att röja snö på ålderdomshem och skolor när stadens resurser inte var tillräckliga. Detta kunde göras eftersom några städer hade gjort data tillgängligt om vilka vägar och genomfartsleder som prioriterades för snöröjning för tillfället. Vilket lede till att medborgare kunde fokusera på områden som var mindre prioriterade i sitt närområde. Exempelvis gjorde staden Chicago⁴ tillgänglig data om var snöplogning var i realtid på en karta, där medborgarna kunde se i vilken riktning bilarna var på väg.

Dataförmedlare kan också fungera som påtryckare för att myndigheter skall skynda på att tillgängliggöra öppna data inom områden där det finns en förfrågan. Dessa förmedlare kan vara personer eller organisationer som förespråkar en viss användargrupp eller verksamhetsområde.

⁴ [Snow plower tracker](#) - Chicago

Exempelvis kan förmedlare förespråka tredjepartutvecklare så att data görs tillgänglig i marskingläsbart format i större utsträckning, så data kan användas för att skapa nya tjänster. En annan funktion förmedlare kan bistå med är att minska gapet mellan medborgare som inte är tekniskt intresserade att nyttjar dessa datakällor genom att skapa informations produkter. Det finns en risk att det bara är en liten grupp av programmerar, forskare och datakunniga person som använder dessa öppna data resurser som finansiera av skattebetalare. Dataförmedlare är en intermediär aktör som kan bidra till att minska det digitala gapet mellan dataleverantörer, utvecklare och medborgare. Dataförmedlare skulle också kunna samarbeta med dataleverantören genom att anpassa datakällor för enklare användning som passar vissa målgrupper. Det gör att dataleverantören kan koncentrera sina resurser på att bygga datakällor med hög exakthet och detaljrikedom och överlåta till dataförmedlare och andra aktörer i värdekedjan för att sammanställa information för att visa på användningsområde av öppna datakällor.