

Rapport

Referensarkitektur: Datahantering och utbyte av data mellan myndigheter

ES2024-09





Innehåll

1.	Inledning	5
1.1	Målgrupp	5
1.2	Bakgrund	5
1.3	Syfte.....	5
1.4	Ambition.....	6
1.5	Användning.....	6
1.6	Förklaringsmodell	7
2.	Området datahantering	8
3.	Möjliggöra effektivt datautbyte mellan organisationer	9
3.1	Grundläggande begrepp och termer	10
3.2	Dataprodukt.....	11
3.2.1	Dataproduktsspecifikation.....	12
3.3	Datakatalog	12
3.3.1	Att dela data -att skörda data	13
3.3.2	Identitet och åtkomst	15
3.4	API	15
4.	Förmågor för datahantering	15
4.1	Huvudförmågor.....	16
4.2	Datapipeline	18
4.3	Ramverk och standards	19
4.4	Avgränsningar.....	19
4.4.1	Ledning och styrning.....	19
4.4.2	Tillämpningar	19
4.4.3	Relaterade referensarkitekturer	20
5.	Komponenter.....	21
5.1	Återanvändning av komponenter och standards.....	24
5.2	Nationellt stöd	25
6.	Generella styrande principer	26
7.	Referenser	27
8.	Appendix.....	28
8.1	Införande och skalbarhet	28
8.1.1	Exempel på införande	29
8.2	Vidareförädling med stöd av referensarkitekturen.....	31
8.3	Digg REST API profil.....	31
8.4	eDelivery.....	32



8.5	Data mesh och Data fabric.....	34
9.	Begreppskatalog.....	35
9.1	Begreppsmodeller.....	49
9.1.1	Begreppsmodell för olika typer av data.....	49
9.1.2	Begreppsmodell för data och relationer till dataprodukt mm.....	49



1. Inledning

1.1 Målgrupp

Denna beskrivning riktar sig i första hand till arkitekter inom en organisation som avser att utveckla sin datahantering för att effektivare kunna dokumentera samt dela data inom och utanför den egna organisationen.

1.2 Bakgrund

Ökade krav på tillgång till data och på själva data driver oss myndigheter mot en mer harmoniserad syn på hur vårt organisationsdata ska hanteras och även kunna utbytas mellan organisationerna.

Hantering av data är fragmenterad inom organisationer i Sverige. Ökade krav på tillgång till data driver oss myndigheter mot en mer harmoniserad syn på hur data ska hanteras och även kunna utbytas mellan organisationerna. Vårt data beskrivs ofta som bränslet i processerna men också som en av organisationernas stora om inte viktigaste tillgångar. Organisationer idag har olika behov samt är olika utvecklade avseende datahantering. I ljuset av detta så har eSam tagit fram en referensarkitektur för hantering av data för att stödja en mer gemensam datahantering för organisationssverige.

1.3 Syfte

Syftet med referensarkitekturen är att ge stöd och vägledning avseende hur en hantering av data kan implementeras, efter egna förutsättningar, för att möjliggöra en effektiv och säker tillgång till data inom en organisation samt delning av data mellan organisationer inom Sverige, exempelvis mellan myndigheter.

Rapporten är skriven av Jan Aspenfjäll (Riksarkivet), David Bengtsson (Havs och vattenmyndigheten), Tina Eriksson (Migrationsverket), Peter Hammar (Tullverket), Anders Levin (Skatteverket), Johan Lindeberg (Skolverket) och Dan Stenvall (Tullverket).



1.4 Ambition

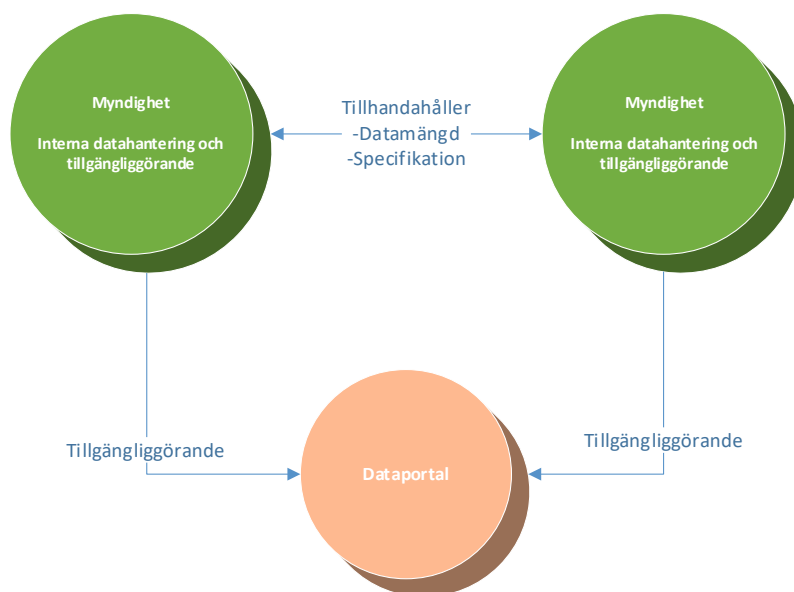
Ambitionen med referensarkitekturen är att ta ett brett och generellt perspektiv, skapa förståelse samt guidning inför ett införande. Detta på grund av att krav på funktionalitet kan skilja sig ganska kraftigt mellan olika organisationer samt tillämpningsområden. Referensarkitekturen syftar till att ge olika insikter hur en realisering kan te sig, viktiga val samt möjlighet att växa och skala upp när behovsbilden förändras.

1.5 Användning

Detta dokument är en av flera produkter inom eSam som berör området data. Referensarkitekturen är en rekommendation som är till för att användas av myndigheter inom eSam för framdrift av området. Referensarkitekturen är tänkt att användas som stöd och som referensarkitektur för återanvändning av mönster för att ta fram en hantering av data resulterande i ett tillgängliggörande av förädlade, paketerade och beskrivna datatillgångar. Se också kapitel 7, referens 1 Styrning och ledning av data (eSam), samt referens 7 och 8.

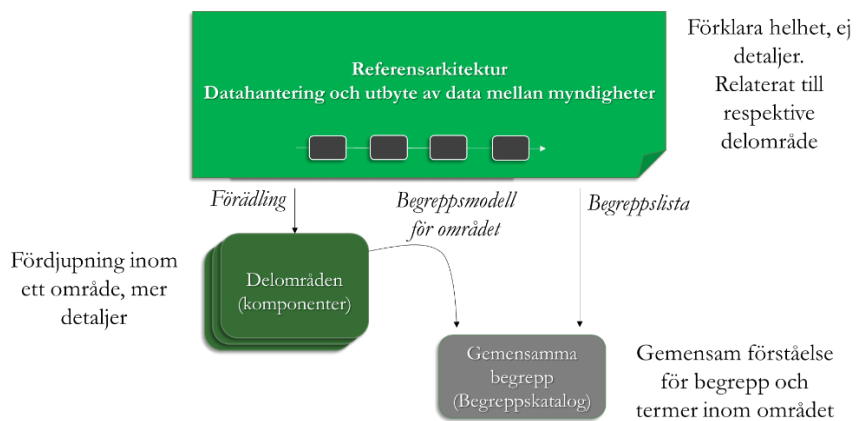
Här visar vi tre huvudscenarios vid hantering och tillgängliggörande av data där referensarkitekturen beskriver de två första:

- Hantering och tillhandahållande av data internt inom en myndighet.
- Hantering och tillhandahållande av data till andra myndigheter.
- Hantering och tillgängliggörande av data till dataportal. Referensarkitekturen beskriver inte tillgängliggörande av data till (Sveriges) dataportal. Så när det gäller tillgängliggörande av data via Sveriges dataportal så ska de riktlinjer som DIGG tagit fram följas, se 7 referens 9.



1.6 Förklaringsmodell

Bilden nedan uttrycker relationen mellan själva referensarkitekturen och delområden. Delområden, inklusive gemensamma begrepp, beskrivs (förädlats) i separata dokument reffererade till i texterna i detta dokument.





2. Området datahantering

Det övergripande området som referensarkitekturen ska stödja brukar bland annat kallas för *Databantering*, på engelska *Data management*. En teknisk realisering för datahantering ska kunna stödja organisationens insamling, bearbetning, dokumentation samt tillgängliggörande av data. Realiseringen av datahantering ska också kunna stödja organisationen i att hantera sin data på ett säkert sätt. Detta inkluderar aspekter som behörighet och åtkomst av data, samt rättssäkra aspekter av datahantering.

Det finns ett flertal koncept att luta sig mot inom området. De koncept som valts här för referensarkitekturen är influerad av koncepten *Data fabric* och *Data mesh*, vilka kompletterar varandras styrkor och "elimineras" deras brister på ett bra sätt. En sammanfattning av koncepten finns att finna i Appendix kapitel 8.5.

I referensarkitekturen har vi även valt att använda begreppet dataproduct som benämning för den datamängd som kan delas och tillgängliggöras. Benämningen dataproduct signalerar en paketering samt ett producent- och konsumentperspektiv.



3. Möjliggöra effektivt datautbyte mellan organisationer

För att möjliggöra ett effektivt datautbyte mellan organisationer så krävs att man nyttjar överenskomna standarder för att utbytet ska kunna ske på ett effektivt sätt. Man ska också tänka in olika nivåer och kontext där datautbyte sker. Exempelvis på myndighets, sektor, nationell samt EU-nivå. När det gäller standards så är det typiskt att man får förhålla sig till överordnade överenskomna standarder samt vara kompatibla med dessa. Sedan medger en del standarder att man gör specifika utökningar inom standarden exempelvis sektorsspecifika tillägg. Man ska också alltid sträva mot att använda sig av öppna standarder och öppen källkod eftersom dessa sänker barriärer för att även inkludera framtida eller okända parter.

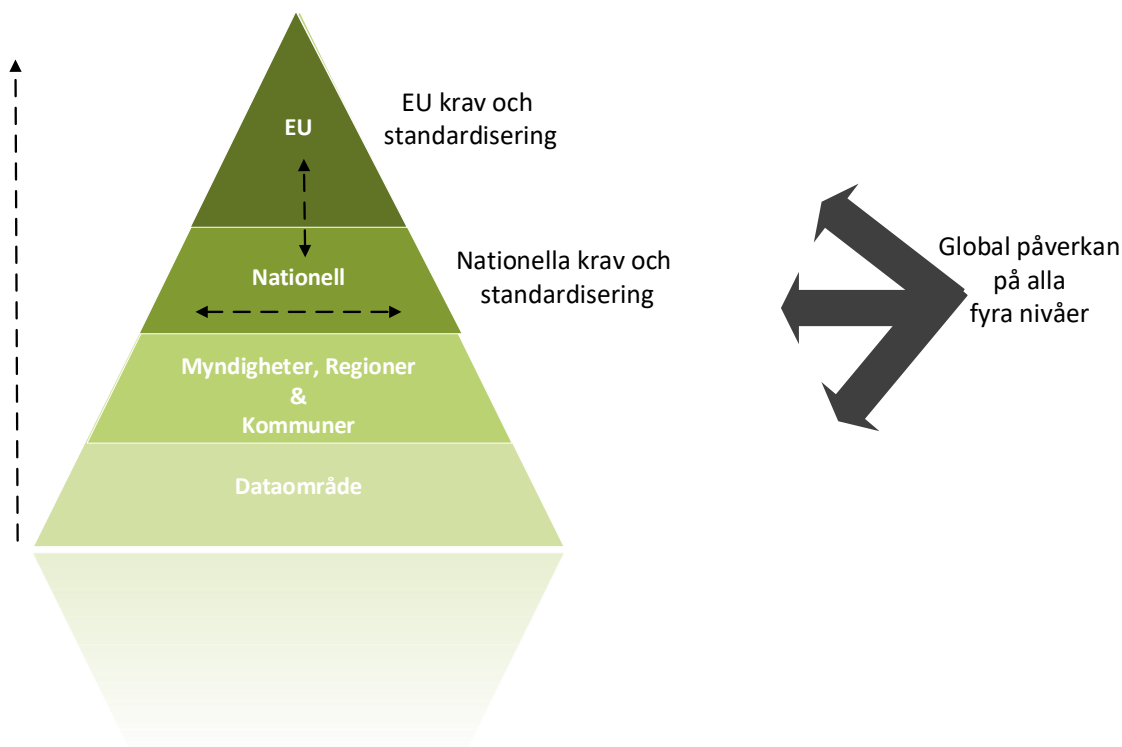
Ett effektivt datautbyte mellan organisationer kräver att man nyttjar överenskomna och öppna standarder för att utbytet ska kunna ske på ett effektivt sätt.

Myndigheterna påverkas av beslut från EU. Vårt datautbyte med EU kommer att öka över tiden, hur datautbytet utvecklas kan svenska myndigheter påverka till viss eller ringa grad. Nationellt sett finns såväl en förväntan och anledning till att organisationerna mer och mer ska utbyta data sinsemellan. På nationell nivå har en etablering av grunddatadomäner påbörjats. Att betrakta den egna organisationen som en dataproducent samt bruka data internt (konsument) är för flertalet organisationer en görlig sak att åstadkomma. Att som organisation utbyta data mellan andra organisationer och agera både dataproducent och datakonsument ställer helt andra krav på organisationerna. Standardisering för utbyte av data ur ett antal perspektiv är nödvändigt. Det innebär att olika typer av beskrivningar måste vara standardiserade och i enighet med involverade organisationer eller andra parter. På en global nivå påverkas myndigheter t.ex. av internationella trender och utvecklingar inom datahantering och teknologi, vilket kan ta sig uttryck i globala standarder och bästa praxis.

Utöver de internationella och nationella kontext som påverkar de flesta eller samtliga myndigheter/organisationer så kan specifika dataområden påverkas. Påverkan sker på



organisationer som har verksamhet relaterad till gemensamma områden. Utvecklingen av dataområden ”data spaces” på europeisk nivå kan vara exempel och avgörande för myndigheter att integrera och samverka med.



3.1 Grundläggande begrepp och termer

Nedan är ett uraval av viktiga begrepp använda i dokumentet. För en mer heltäckande lista se kapitel 7, referens 3.

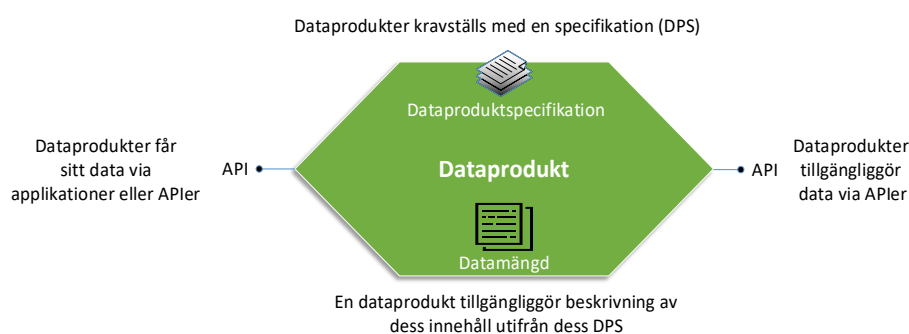
- *API*: Application Programming Interface är specifikation av hur olika applikationsprogram kan använda, kommunicera och/eller överföra data mellan varandra
- *Data*: Data är råa, opolerade fakta och symboler som representerar händelser, objekt eller mätvärden
- *Datakatalog*: En datakatalog är en produkt som lagrar metadata om en datamängd
- *Datakälla*: Ursprunget för data, var datat har skapats
- *Datamängd*: Identifierbar samling data.
- *Dataprodukt*: Datamängd eller serie av datamängder som erbjuds en konsument



- *Dataproduktspecifikation*: Specifikation av en dataproduct tillsammans med ytterligare information som möjliggör att densamma kan skapas och distribueras.
- *Metadata*: Metadata används för att beskriva innehållet och/eller strukturen för en viss datasamling. Detta kan göras ur mer än ett perspektiv.
- *Referensarkitektur*: En modell eller ramverk som beskriver en rekommenderad struktur och organisering för ex system inom ett specifikt område. Den fungerar som en vägledning och standardiseringsgrund.
- *Skörda data*: Data harvesting är en process som läser/hämtar datamängder och/eller deras metadata mellan producent och konsument
- *Tillgängliggöra*: Ge tillgång till data. Oavsett om det görs frivilligt eller på grund av en skyldighet i en författning

3.2 Dataproduct

För att på ett effektivt sätt kunna utbyta och nyttja data så ska data paketeras, beskrivas och tillgängliggöras, på så sätt att det underlättar för konsumenten att förstå och att använda data för sina tillämpningar, en så kallad dataproduct. En dataproduct innehåller en eller flera datamängder och en dataproductspecifikation som beskriver dataproducten. En dataproduct tillhör ett specifikt område samt har en ägare, se också kapitel 7, referens 4. Själva datamängden kan vara både i form av ostrukturerad eller strukturerad data medan dataproductspecifikationen alltid är i strukturerad form dvs metadata. Som konsument kan man konsumera dataproducten via ett API som i sin tur kan medge nyttjande av hela eller delar av datamängden. Se också 7, referens 4.



Dataproducten ska lagras, livscykelhanteras samt behörighetsstyras på ett sätt som uppfyller lagstiftning och säkerhetskrav. Vanligt är att dataproducten är säkerhetsinformationsklassificerad. En annan viktig aspekt är att vid tillgänggörande kunna uppfylla överenskomna tillgänglighets samt prestandabehov.

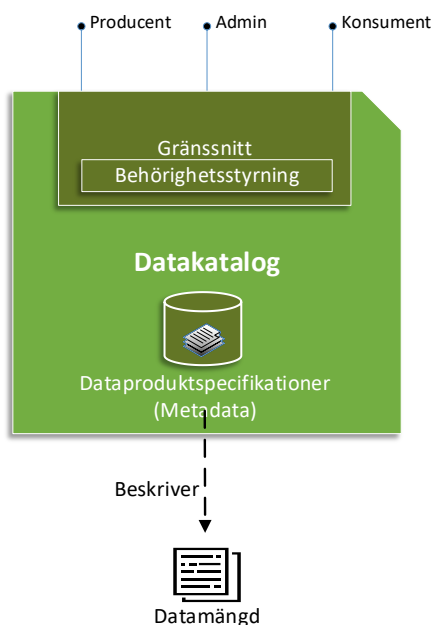


3.2.1 Dataproduktsspecifikation

I den här kontexten för referensarkitekturen så säger vi att en dataproduktsspecifikationen sätter upp ramarna för vilken metadata som ska anges vid beskrivning av en datamängd. Se också kapitel 7, referens 4.

Vad är då metadata? Det är en uppsättning data som används för att definiera andra data, med andra ord data om data. Metadata är ett fundament för en datakatalog men även också för dataprodukt. Se också kapitel 7, referens 2.

3.3 Datakatalog



En datakatalog är en centraliserad förteckning över dataprodukter (ett kartotek) som finns tillgängliga i en organisation eller mellan organisationer. I en datakatalog kan man hitta information om en dataprodukts metadata så som information om dataproduktens namn, beskrivning, källa, tillgänglighet, datakvalitet, informationssäkerhetsklassificering och användning, men också metadata om en mängd olika dataresurser, som databaser och API:er.

Relationen mellan en datakatalog och en dataprodukt är att en datakatalog fungerar som en inventering av alla dataprodukter som finns tillgängliga, Datakatalogen innehåller dataproduktsspecifikationer (eller referens till den) baserad på metadata. Datakatalogen behöver inte nödvändigtvis innehålla dataproduktens datamängd. Genom att ha en uppdaterad och korrekt datakatalog kan man lättare hitta och använda dataprodukter på ett effektivt sätt och man kan säkerställa att användningen av dataprodukterna sker på ett korrekt och säkert sätt. Se också kapitel 7, referens 2.



En datakatalog är också en programvara som måste kunna administreras och utvecklas efter organisationens behov. En datakatalog består vanligtvis av följande huvudbeståndsdelar:

- Gränssnitt; Producentgränssnitt, Konsumentgränssnitt och Administrativt gränssnitt
- Behörighetsstyrning
- Dataproduktspecifikationer

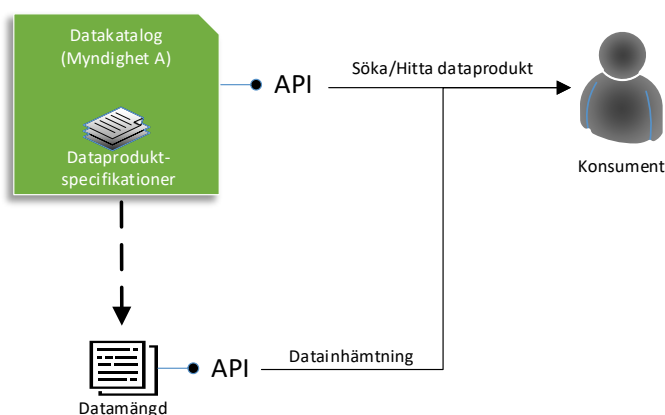
3.3.1 Att dela data -att skörda data

En datakatalog kan finnas tillgänglig:

- internt inom en organisation
- som av en organisation tillgängliggjord resurs som gör det möjligt för andra organisationer att hitta och konsumera delat data, eller som en gemensam resurs mellan organisationer.¹

Internt inom en organisation

Datakatalogen som en internt tillgängliggjord resurs gör det möjligt för den egna organisationer att hitta och konsumera delat data. Bilden nedan illustrerar hur data inhämtas av en datakonsument från en datakatalog. Det är två olika APIer som används för att komma åt data. Ett API tillhörande datakatalogen för att en konsument ska kunna söka efter och hitta dataprodukter samt ett annat API för att hämta datamängden från datakällan. APIerna kan användas direkt av en konsument alternativt att man bygger ett användargränssnitt ovan på dessa, som konsumenten sedan kan nyttja.



¹ Notera att referensarkitekturen inte beskriver tillgängliggörande av data till (Sveriges) dataportal utan hänvisar till de instruktioner och beskrivningar som tillhandahålls av Digg

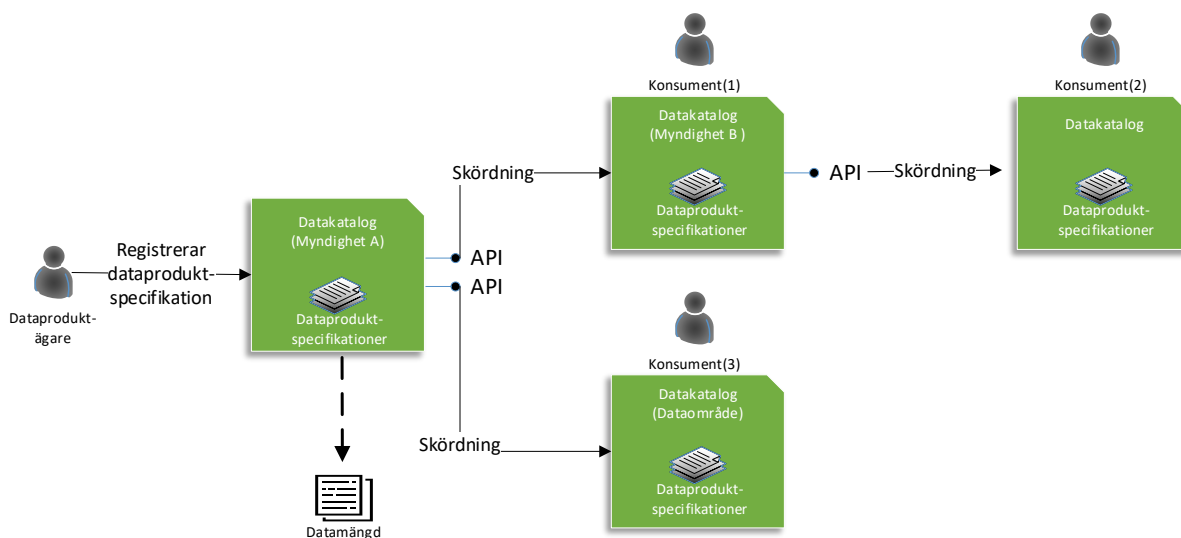


Externt dela data till andra organisationer

Datakatalogen som en externt tillgängliggjord resurs gör det möjligt för andra organisationer att hitta och konsumera delat data eller datakatalogen som en gemensam resurs mellan organisationer

Datakatalogen är också viktig för att en effektiv datadelning kan ske mellan organisationer. Att insamla data från andra organisationer kan ske genom att först hämta specifikation (skördar metadata) från den organisation som tillhandahåller datakatalogen. När metadata väl är skördat så hämtar man sedan själva datamängden, genom distribution från den dataägande organisationen i fråga. Att effektivt kunna skörda specifikationer och hämta data mellan organisationer förutsätter att alla involverade organisationer följer gemensamma standards och ramverk rörande metadata, APIer och säkerhet.

I bilden nedan hittar ni tre stycken (data-)konsumenter som kan med hjälp av varje organisations datakatalog skörda en specifikation på metadata och sedan inhämta en datamängd från den dataägande organisationen. I varje steg i kedjan så är det varje organisations datakatalog som är central för att möjliggöra att nästa organisation hittar rätt dataprodukt och får tillgång till dess specifikation och metadata. Bilden beskriver att ägare av data på myndighetsnivå registrerar en dataproduktspecifikation. Konsument(2) av data kan skörda en specifikations metadata tillhörande en dataprodukt på myndighet (B). Detta via en datakatalog vars konsument(1) har skördat specifikationens metadata från myndighet(A). När väl en konsument skördat specifikationen så kan konsumenten hämta data från ursprungskällan, i detta fall myndighet(A).





3.3.2 Identitet och åtkomst

För att säkerställa rätt åtkomst till datakatalog, dess specifikationer samt datamängder så behöver en behörighetslösning appliceras för att validera åtkomsträttigheter mot gränssnitt och API:er. Beroende på behov, exempelvis intern användning så kan, i de flesta fall, organisationens befintliga lösningar användas. Ska man dela data mellan flera organisationer så är federerade lösningar (tillitsnätverk) där respektive ansluten organisation ansvarar för att styra de interna behörigheterna förmodligen mer effektiva.

När det gäller behörigheter och behörighetsstyrning så kan man dra nytta av datakatalogens innehåll då dataprodukterna är informationssäkerhetsklassificerade och kanske kategoriserade. Som exempel så skulle man kunna utforma behörighetsprofiler som ger åtkomst till vyer av dataprodukter och dess innehåll exempelvis inom en viss kategori och upp till en viss informationsskyddsklass. På så sätt får man en mer dynamisk behörighetsstyrning då dataprodukter kan tillkomma samt avvecklas.

3.4 API

När det gäller att tillgängliggöra datamängder så kan ett API (application programming interface) etableras som används till att överföra datamängden mellan producent och konsument på ett formaliserat sätt. Idag används ett flertal olika metoder för att göra just datautbyte, exempel kan vara att man använder sig av ftp (file transfer protokoll) eller kanske webgränssnitt där man kan ladda upp eller hämta data. I referensarkitekturen väljer vi att peka på mer standardiserade alternativ eftersom det på sikt leder till ett mer effektivt datautbyte och interoperabilitet organisationer emellan. Området finns utvecklat i Appendix kapitel 0.

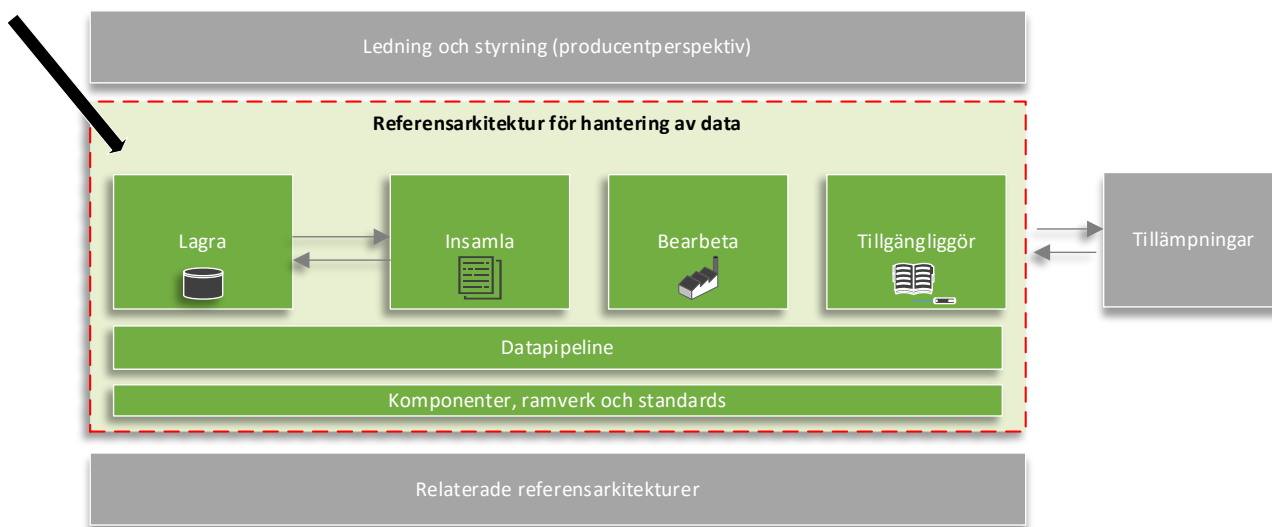
4. Förmågor för datahantering

Området i (kart-)bilden nedan, markerat med *streckad röd linje*, visar på omfattningen av referensarkitekturen för hantering av data med dess fyra huvudförmågor:

- Insamla
- Bearbeta
- Tillgängliggöra
- Lagra



Bilden visar också på stödförmågor som ”Datapipeline”, ”Komponenter, ramverk och standards”. De tre gråa rutorna i bilden är till för att visa omgivningningar till referensarkitekturen som belyses kort. De grå rutorna ”Ledning och Styrning”, ”Relaterade referensarkitekturer”, ”Tillämpningar” är dock inte en del av referensarkitekturen. Referensarkitekturen behandlar heller inte legala perspektiv av datahantering.



4.1 Huvudförmågor

- **Insamla data.** Förmågan syftar till att styra hur data samlas in samt att tillgängliggöra data för vidare bearbetning eller distribution via andra ingående delar i referensarkitekturen. Att starta en insamling kan innebära att man exempelvis samlar in data från externa källor sett från referensarkitekturens perspektiv. Externa källor kan vara annan organisation eller källor inom den egna organisationen. Det är i denna förmåga som man identifierar och tillhandahåller data, för att i ett senare steg kunna bearbetas för den egna organisationens fortsatta behov. Att starta en ny insamling av data dvs innan data är paketerat, informationssäkerhetsklassificerat och dokumenterat till en dataprodukt kan vara en tidskonsumerande aktivitet. När väl produktifieringen av data är genomförd så finns tydliga likheter mellan insamling och förmågan tillgängliggörande. I insamlingsförmågan kan referensarkitekturen ge stöd i form av gemensamma tjänster och standards som underlättar arbetet med att samla in och tillgängliggöra data för vidare hantering. I en idealvärld kan det vara en dataprodukt man tar emot vilket borde minska arbetet med insamlingsförmågan väsentligt.



- **Bearbeta data.** Förmågan syftar till att bearbeta data för flera ändamål. Det kan vara att organisera, dokumentera och paketera data som dataprodukt, informationssäkerhetsklassificera data men även livscykelhantera data, paketera data för arkivering, innan tillgängliggörande alternativt återlagring av data till datalagring efter bearbetning. Som stöd vid bearbetning av data finns tillgång till tjänster som i sin tur kan orkestreras med hjälp av ett processtöd i det fall flera olika moment och delbearbetningar ska utföras och koordineras. I den här typen kan man med fördel använda AI som teknikstöd för bearbetning av data.
- **Tillgängliggör data.** Förmågan syftar till att ge tillgång och kunna hitta dataprodukt. Som konsument kan man söka bland registrerade dataprodukt och på så sätt hitta dokumentation om, samt hur man kan konsumera dataprodukt. För att få tillgång till en dataprodukt så måste användaren ansöka om behörighet till dataprodukt. Villkoren för behörigheter bestäms av organisationen som tillgängliggör dataprodukten. Det kan vara juridiska krav men även andra krav beroende på dataproduktens beskaffenhet. Data kan i vissa fall finnas öppet utan att ansöka om behörighet och det finns inte alltid behov av särskilda villkor för användandet. Tillgängliggöra data kräver någon typ av katalog av tillgängliga data, dataprodukt, detta för att kunna hitta data, läsa beskrivningar om data, erhålla information om ex informationsskyddsklass om data samt referenser till APIer. Det är också i denna förmåga som Identitets och åtkomsthantering har sin plats för att garantera att rätt personer har tillgång till rätt data. Oavsett hur data insamlats och bearbetats så är det viktigt att dataprodukt följer (gemensamma) ramverk när och om det ska tillgängliggöras för den egna organisationer och/eller externa organisationer.
- **Lagra data.** Förmågan att lagra data är en del av referensarkitekturen, på vilket sätt man väljer att lagra är ett val som varje enskild organisation själva tar ställning till. Dataproduktägare har ett ansvar för lagring av dataprodukten. Ansvaret omfattar att säkerställa kapacitet, säkerhet samt korrekthet i förhållande till dess datakälla och eventuell bearbetning. Dataprodukt kan vara lagrade centraliserat eller decentraliserat beroende på sammanhang, behov och kanske även sett ur ett organisationsperspektiv. Olika tekniker för behov och ändamål kan nyttjas för att fysiskt lagra data, exempel på sådana tekniker kan exempelvis vara SQL-databas, graf-databas, In-memory, Streamingplattform, etc. Lagringsteknik bör anpassas efter dataformatet samt de behov som finns ur ett tillämpningsperspektiv. Beroende på förmågan hos de källsystem som lagrar datamängden, ex äldre system så kan det i många fall vara bra att etablera en gemensam lagringsyta som klarar krav som prestanda m.m. på ett bättre sätt. Källsystemen kan då leverera sin datamängd till denna lagringsyta som i sin tur exponerar datamängden till konsumenterna.



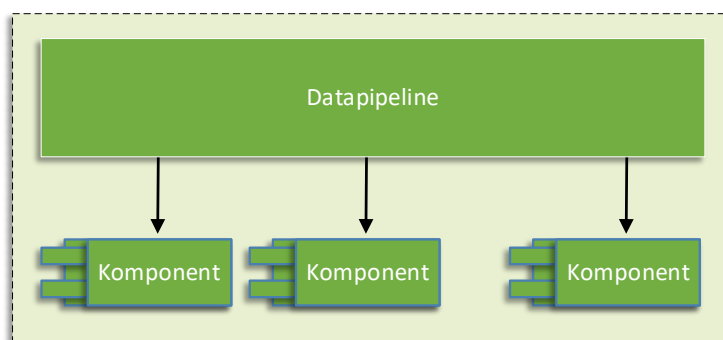
4.2 Datapipeline

Datapipeline är det övergripande termen för alla processer där data flyttas (orkestrering) och består av tre väsentliga delar: källa, bearbetning och destination. I referensarkitekturen innebär datapipeline möjligheten att bygga och koordinera databearbetningsprocesser för att uppnå en hög grad av automation i hela dataflödet, detta från att samla in data till att tillgängliggöra bearbetat data som en levererbar dataprodukt. Det som påverkar pipeline's beteende och implementation är organisationens behov av hur data ska te sig vid ett tillgängliggörande. Datapipeline's påverkas också av olika ramverk, vilka kan tillhandahålla regler, riktlinjer och rekommendationer samt tillhörande anvisningar som de olika datapipeline's utförande baseras på och hur de ska lösa sina uppgifter och interagera.

Exempel på ramverk:

- Loggramverk; Ramverk för spårning och registrering av händelser, fel och prestanda under datainsamlingen för att underlätta övervakning och felsökning.
- Säkerhetsramverk
- Identitet och åtkomstramverk
- Principer för utbyte av metadata och dataprodukter
- Beskrivning och paketering av data
- Exponering av händelser kopplat till en datamängd

Datapipeline triggas av olika händelser eller kan vara schemastyrd. Datapipeline orkestrerar en kedja av moment som data ska genomlöpa exempelvis validering, normalisering, informationssäkerhetsklassificering, dokumentering och paketering samt publicering i datakatalog där terminering då sker. En datapipeline kan orkestrera komponenter genom alla de tre förmågorna (*Lagra, Insamla, Bearbeta, Tillgängliggöra*) men kan också orkestrera komponenter endast i en enskild förmåga. Datapipeline ska vara så automatiserad som möjlig, detta framtalar dock inte eventuella behov av manuell





hantering i vissa lägen. Ett exempel på detta kan vara ett uthopp från den automatiserad kedjan av händelser för att hantera (registrera) en dataprodukt i datakatalog.

Denna referensarkitektur omfattar inte att beskriva implementering av pipelines, utan det är upp till varje organisation att ta fram de ramverk och pipelines som de anser är lämpliga.

4.3 Ramverk och standards

Ramverk och standards är en del av referensarkitekturen men beskrivs inte vidare under detta kapitel. Detta på grund av att ramverk och standards omnämns i ett flertal andra kapitel exempelvis kapitel 0, kapitel 5.1 och kapitel 8.1.

4.4 Avgränsningar

4.4.1 Ledning och styrning

Ledning och styrning är inte en del av referensarkitekturen för att hantera data, se kapitel 7, referens 1 Styrning och ledning av data (eSam). Dock bör man se ledning och styrning i den gemensam kartbilden för att få en mer heltäckande förståelse för hantering av data. För att leda och styra en organisation så har hantering av data en central roll och är en grund för många områden och måste där med hanteras effektivt. Exempel på omfattning: Budgetstyrning; Riskhantering, Lagstiftning och policy, etc.

4.4.2 Tillämpningar

Huvudförmågan "området" Tillämpningar är inte en del av referensarkitekturen för att hantera data. Dock bör man se området i den gemensam kartbilden för att få en mer heltäckande förståelse. förmåga Tillämpningar är "något" som via referensarkitekturen nyttjar tillgängliggjorda dataprodukter för att skapa verksamhetsvärde. Tillämpning kan vara av olika typer exempelvis:

- Analysplattformar (AI baserade)
- Arkivering
- Applikationer/bearbetning av data
- Människa
- Andra (externa) organisationer

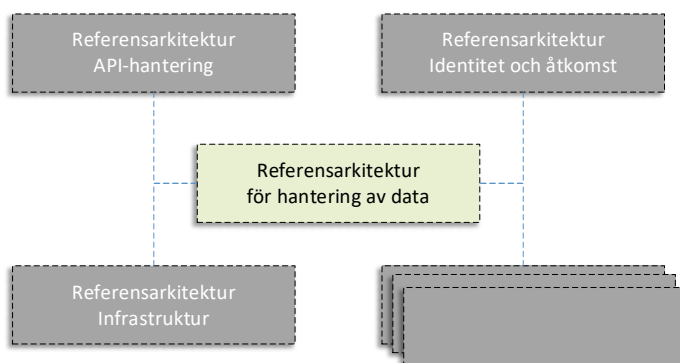
Ur ett konsumentperspektiv så kan dataprodukter bearbetas vidare för att exempelvis användas vid analyser. Efterbearbetningen kan innebära att man vill skapa nya



dataprodukter. Lagring av dessa nya produkter kan efter organisationens behov ske utifrån vilken lagringsprincip organisationen valt.

4.4.3 Relaterade referensarkitekturer

Referensarkitekturen står inte helt för sig själv den "samarbetar/angränsar" och kräver referensarkitekturer inom andra områden så som IaM (Identitets och åtkomsthantering). Referensarkitekturen för hantering av data nyttjar och kan påverka nya behov hos andra områdets referensarkitekturer. Angränsade referensarkitekturer är inte omnämnda i den övergripande kartbilden och omnämns inte vidare annat än som (möjliga) referenser i denna skrift.





5. Komponenter

De funktionsinnehållande delarna i referensarkitekturen benämner vi som komponenter. Komponenterna är till för att modularisera funktionalitet samt underlätta val som en organisation kan göra vid en implementering av referensarkitekturen. Komponenterna kan vara realiserade eller realiseras med olika tekniker exempelvis klassisk programmerade eller med hjälp av modeller och AI. Komponenterna kan finnas framtagna av olika organisationer, möjlig att inköpa eller enbart kravställd inför en framtida utveckling. Digg tillhandahåller ett antal byggblock som kan likställas med ett antal komponenter. Vilka komponenter som ska användas avgör varje organisation som ska införa en hantering av data. I tabellen nedan så exemplifieras några komponenter samt sätter dessa i sitt sammanhang i förhållande till huvudförmågorna. Komponenterna kan återfinnas i en eller flera av de tre huvudfrågorna, nedan är ett exempel på hur det kan se ut.

Komponenter:	Insamla	Bearbeta	Tillgängliggöra	Beskrivning
Integrationstjänster	X	X	X	Anpassade gränssnitt för att hämta data/dataprodukter från externa källor och integrera dem i det interna systemet. Kan vara tidstyrda eller händelsestyrda samt ha förmåga att hantera strömmande realtidsdata.
Händelseförmedling	X	X	X	Samla och publicera händelser som andra komponenter kan konsumera och agera på vid behov.
Schemaläggning	X		X	Tidsstyrning av aktiviteter och processer.
Kontrakt		X		Tydliga avtal och definitioner för insamlade data, format och användningsvillkor.
Validering	X	X		Kontroll av format och syntax av data för att säkerställa att de följer tekniska specifikationer. Valideringen omfattar också standardiserade format och datastrukturer som används för att säkerställa att data kan förstås och användas korrekt av användare och system.
Notifikationstjänster	X		X	Skicka meddelanden eller aviseringar om status och händelser relaterade till datainsamlingen.



Normalisering av data	X	X		Omvandling av (rå-)data till en gemensam datastruktur för att bland annat säkerställa kompatibilitet samt eliminera duplicerade poster (dubletter av data) i en dataprodukt.
Transformationskomponenter		X		Omvandlar och rengör data när de samlas in för att säkerställa att de är i rätt format och innehåller nödvändig information i enhetlig form.
Pseudonymisering		X		Ersätta känslig information med en pseudonym (alias) för att skydda den verkliga identiteten hos individer i datasamlingar. Se kapitel 7, Referenser
Anonymisering		X		Tar helt bort eller irreversibelt omvandla data för att göra dem oigenkännliga, vilket säkerställer att personlig information inte kan härledas från dem.
Maskning av data		X		Filtrerar eller ta bort specifika känsliga data från dataprodukt för att minska riskerna för oavsiktlig exponering.
Sammanslagning av data	X	X		Kombinera data från olika källor eller tabeller för att skapa en mer omfattande och sammanhängande dataprodukt.
Gallring av data		X		Eliminerar eller ta bort data som är oönskad eller överflödigt enligt författning, föreskrift eller beslut. samt för att förbättra databasens kvalitet Gallring är ofta en del av livscykelhantering av data.
Informationssäkerhetsklassificering av data		X		Att kategorisera dataprodukter baserat på dess känslighetsnivå/säkerhetsnivå, vilket är användbart för att tillämpa lämpliga säkerhetsåtgärder och accesskontroller.
Datakatalog			X	En centraliserad förteckning över dataprodukter innehållande information om tillgängliga datamängder. Se kapitel 7, Referenser



Begreppskatalog			X	Tillgång till begreppskatalog samt administrativ hantering så som tillgängliggörande, import och export av kataloger, samt synonymhantering. Se kapitel 7, Referenser
Referensdatakatalog			X	Tillgång till referensdatakatalog samt administrativ hantering så som tillgängliggörande, import och export av kataloger.
API-tjänster			X	Application Programming Interfaces som möjliggör programmatisk åtkomst till data och tillhandahåller standardiserade metoder för att hämta och manipulera data.
Attributtjänst			X	En attributtjänst kan vara ett bra stöd vid olika tillämpningar. Exempelvis stöd för identitet och åtkomsthantering. Attributtjänsten kan då tillhandahålla förslagsvis kategori, säkerhetsklass m.m för en datamängd vilket kan matchas mot en behörighetsprofil som använder sig av dessa attribut för att medge åtkomst till data.
Översättning		X		Tjänst för översättning av olika språk.
Ontologi			X	Möjliggör för organisationer att förstå sina data bättre. En ontologi fokuserar på datats relationer. I kontexten av referensarkitekturen är en ontologi ett stöd till en datakatalog, i den meningen kan ontologin hjälpa en applikation att förstå en organisations tillgängliggjorda data satt i sitt sammanhang. Med hjälp av AI kan en ontologi byggas upp och hållas aktuell över tiden.
Datapipeline	X	X	X	Datapipeline är tjänsten att bygga och koordinera databearbetningsprocesser för att uppnå en hög grad av automation i hela dataflöde.



5.1 Återanvändning av komponenter och standards

Det är lätt att man fastnar i sin egen organisations behov och ser den som helt unik utifrån ett datahanterings perspektiv. Vid utbyte av data mellan olika organisationer så är likheterna stora. Mer likheter än olikheter finns även inom organisationers sätt att insamla, bearbeta och tillgängliggör data. Återanvändning av komponenter kan göras mellan organisationer. Likheterna mellan olika organisationers behov är stor vad gäller komponenter exempelvis för pseudonymisering, anonymisering, översättning och integrationstjänster vilka bör kunna återanvändas. Flertalet komponenter och bearbetningar kan också med fördel nyttjas som gemensamma tjänster där även körmiljö tillhandahålls.

Fördel med återanvändning av komponenter är minskad kostnader samt likartade resultat vid bearbetning av data, exempelvis nyttjande av samma komponent vid översättning av text bör ge samma översättning.

En organisation står inför att antal val att göra när man ska välja komponenter, exempelvis:

- Används eller finns en komponent tillgänglig hos annan organisation eller inom den egna organisationen
- Finns det samordningsbehov mellan organisationer att använda samma komponenter
- Ska en komponent egenutvecklas, utvecklas i samråd med andra organisationer eller ska den köpas in. Att egenutveckla generiska komponenter exempelvis översättningskomponenter rekommenderas inte
- Ska en komponent exekveras OnPrem, eller externt exempelvis hos annan organisation eller i molnet

Återanvändning rör inte bara komponenter, många principer och standards är återanvändbara inom och mellan olika organisationer. Återanvändbarheten utesluter inte att nya principer och standards kan tillkomma under ett långsiktigt införande, vilket innebär att samordning, harmonisering av lösningar och gemensamt uppfyllande av krav mellan organisationerna är ett långsiktigt samarbete. Återanvändning och standardisering leder till således till interoperabilitet mellan organisationer.



5.2 Nationellt stöd

Det finns redan idag en hel del nationellt stöd framtaget som kan återanvändas. Bland annat Diggs byggblock. Dess byggblock ska gå att använda av många organisationer för att stödja digital utveckling, såsom att bygga digitala tjänster mot slutanvändare eller utbyta data med en annan aktör. Där ett byggblock kan vara realiseringen av en komponent.

Komponent	Beskrivning
Metadatahantering	<p>Byggblock framtaget i DIGGs regi där man har tagit fram en metadata-profil som kallas DCAT-AP-SE vilket är den övergripande specifikation som ska användas för lagstadgad leverans av metadata till den nationella förteckningen på Sveriges dataportal.</p> <p>Mer information: https://www.Digg.se/ledning-och-samordning/ena--sveriges-digitala-infrastruktur/byggblock/byggblock.beskrivningar/metadata</p>
DCAT-AP-SE-Processor	<p>Verktyg för att kunna generera en dataproduktspecifikation baserat på DCAT-AP-SE ur en API-specifikation.</p> <p>Mer information: https://github.com/DIGGSweden/DCAT-AP-SE-Processor</p>
Sveriges dataportal (datakatalog)	<p>Sveriges dataportal samordnar och tillgängliggör datamängder som tillhandahålls av offentliga och privata organisationer.</p> <p>Mer information: https://www.Digg.se/digitala-tjanster/sveriges-dataportal</p>
Rest-API-profil	<p>Det finns en rest-api profil framtagen i DIGGS:s regi som kan användas vid framtagande av API:er.</p> <p>Mer information: https://dev.dataportal.se/rest-api-profil https://dev.dataportal.se/api-playbook</p>
eDelivery	<p>Byggblock baserat på EU:s CEF eDelivery. Möjliggör säkert datautbyte mellan organisationer.</p>



Mer information:

<https://www.Digg.se/digitala-tjanster/edelivery>

6. Generella styrande principer

eSam och Digg har tagit fram ett antal generella och vägledande principer som applicerar till området datahantering och utbyte av data.

Se kapitel 7, referens 7 *eSam: Vägledande principer för digital samverkan* och referens 8 *Digg: Nationella principer för att tillgängliggöra information*.



7. Referenser

Nummer	Titel
1	ES2024-04 Styrning och ledningsfrågor av data (eSam)
2	ES2024 -10 Beskrivning för datakatalog
3	ES2024-12 Gemensamma begrepp och kodlistor
4	ES2024 -11 Beskrivning och paketering av data som produkt
5	ES2023-01 AI –Utvecklingsprocessen och data (eSam)
6	ES2022-01 Vägledning Pseudonymisering av personuppgifter
7	eSam: Vägledande principer för digital samverkan https://www.Digg.se/kunskap-och-stod/svenskt-ramverk-for-digital-samverkan
8	Digg: Nationella principer för att tillgängliggöra information https://www.Digg.se/kunskap-och-stod/oppna-och-delade-data/offentliga-aktorer/nationella-principer-for-att-tillgangliggora-information
9	Övrigt om Sveriges dataportal: Strategier - Sveriges Dataportal. Dela-data - Sveriges Dataportal Stöd och verktyg - Sveriges Dataportal - Sveriges Dataportal. Lagstiftning - Sveriges Dataportal



8. Appendix

8.1 Införande och skalbarhet

Vid ett införande rekommenderas att ta fram en långsiktig målbild, likväl en målbild per förmåga, detta utifrån organisationens strategiska behov. Målbilden kan omfatta ett antal övergripande användarfall för ökad tydlighet. Målbilden blir tongivande för skalbarhet, val av komponenter och dess orkestrering, samt återanvändning. En mindre organisation har ofta mindre kapacitet att ta ett stort kliv på engång i jämförelse med en stor organisation vid ett införande av datahantering. Även för en stor organisation är det sällan görligt eller lämpligt att ta ett stort kliv på en gång. Vid ett införande så har själva resan också värden, exempelvis organisatorisk och teknisk mognad. Oavsett om man är en liten eller stor organisation så finns över tiden behov av att kunna skala upp sin datahantering, utveckla och vidare automatisera den. Förmågorna i datahanteringen har var för sig möjligheter att skalas upp efter behov. De sammanhållande faktorerna är ramverk, standards och orkestrering samt hur komponenter inskaffas och implementeras. Utöver själva referensarkitekturen för datahantering ska man också lägga perspektiv som ledning och styrning och angränsande referensarkitekturer.

- Vid *insamling* av internt data så har man normalt sett kontroll på sin organisations data. Däremot är det inte säkert att data är produktifierat och standardiserat. Det är viktigt att tydligt beskriva hur data är åtkomligt och i vilka format data finns. Dessa frågor måste lösas på ett enhetligt sätt inom organisationen. Oavsett datakälla så gäller det att se helheten av dataflödet genom alla de fyra förmågorna för att göra ett effektivt val av hur lång man ska likrikta (bearbeta) data innan man kommer till bearbetningsförmågan och därmed kunna välja lämplig komponent(er). Flödet av data och bruket av valda komponenter ska orkestreras. Beroende på målbild, så måste man utforma en eller flera pipelines hanterande valda komponenter. Här är av största vikt att interna ramverk likväl som gemensamma ramverk används och följs. En aspekt att ta hänsyn till är om data är realtidsdata eller ej, det påverkar kapacitetsbehov likväl som orkestrering av dataströmmen. Realtidshantering av data kan påverka lagringsbehov för dataprodukten i fråga, men också behov av temporär lagring vid kapacitetskrävande bearbetning av data innan tillgängliggörandet.
- Vid *bearbetning* av data så hamnar fokuset på organisationens behov av hur data ska te sig inför ett tillgängliggörande. Det innebär att paketering och



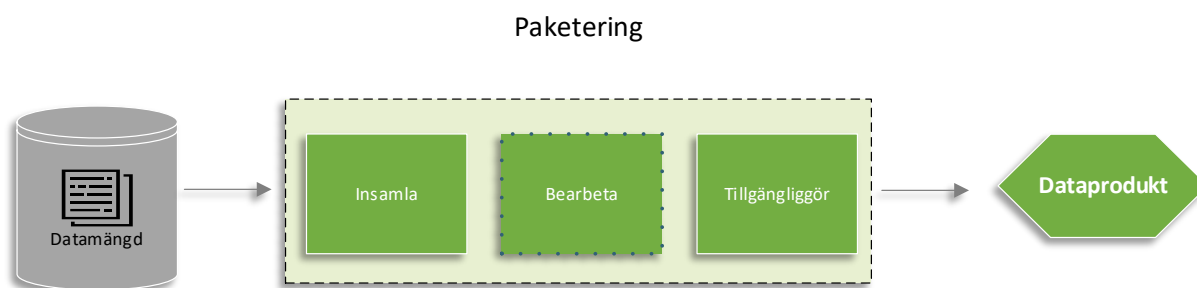
beskrivning av data också påverkas av ett eventuellt planerat tillgängliggörande till andra organisationer. Målbilden viktigt för att kunna välja komponent(er) för bearbetningen av data. Som i de övriga förmågorna så ska valda komponenter orkestreras med hjälp av en eller flera pipelines. Exempel på en viktig komponent att beakta är informationssäkerhetsklassificering av data, detta för nyttjande för identitets och åtkomsthantering av dataprodukter i tillgängliggörande förmågan.

- Vid *tillgängliggörande* av data så hamnar fokuset på hur applikationer/konsumenter hittar data och får tillgänglighet till data på ett säkert sätt. De primära komponenterna är datakatalog, API och attributtjänster för att stöda säker åtkomst av data. Som i de övriga förmågorna så ska valda komponenter orkestreras med hjälp av en eller flera pipelines. För att möjliggöra en aktuell katalogisering över organisationens dataprodukter så omfattar denna förmåga också administration av själva datakatalogen. Dataprodukterna tillgängliggörs i första hand internt inom den egna organisationen. Dataprodukter kan också tillgängliggöras till andra organisationer. Då måste de uppfylla gemensamma och överenskomna standards, så som standards för metadata, APIer, åtkomstsäkerhet, etc. Extern publicering av dataprodukter kan omfatta ett tillgängliggörande via en dataportal eller ett tillgängliggörande hos den egna organisationen.

8.1.1 Exempel på införande

8.1.1.1 Enkelt införande

Ett första steg är att inventera organisationens datatillgångar och identifiera lämpliga kandidater för datadelning. Paketerar man dessa på ett standardiserat sätt så kan man senare skala upp och på ett enklare sätt dela data internt eller externt. Själva katalogiseringen och dataproduktspecifikationen kan initialt hållas i ett enklare vertyg som exempelvis excel eller websida som i ett senare skede får vidareutvecklas eller inhandlas.



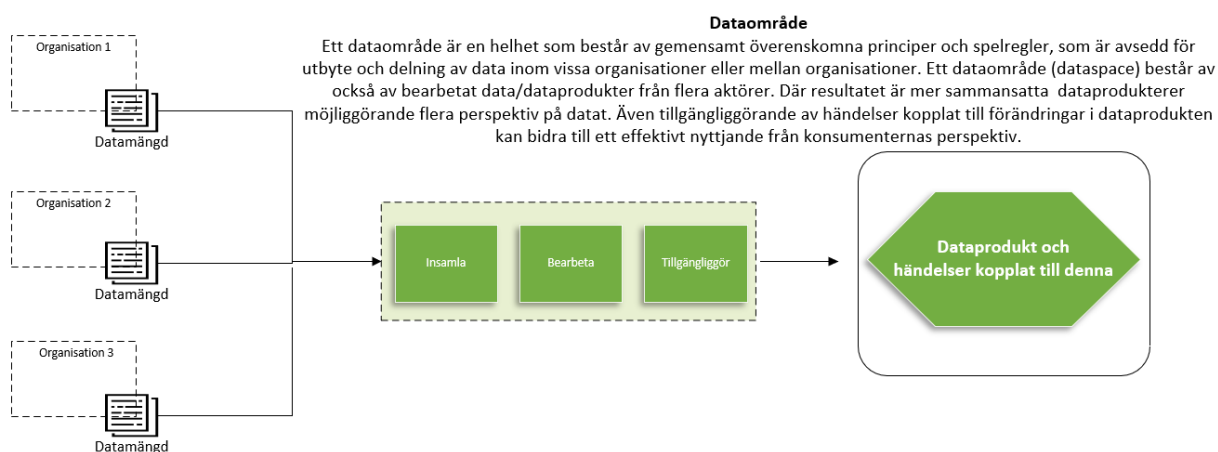


De datatillgångar som är föremål för att tillgängliggöras som öppna data registreras på Sveriges dataportal.

8.1.1.2 Avancerat införande

Man kan även se referensarkitekturen i ett större sammanhang där flera aktörer bidrar med data (insamling) till ett dataområde. Datat struktureras, normaliseras, bearbetas samt katalogiseras för det större sammanhanget. I det här scenariot så kan AI algoritmer och kunskapsmodeller (kapitel 7, referens 5) samt tekniker som datavirtualisering vara ett effektivt stöd för att klassificera data och för att åstadkomma mer avancerade katalogiseringar, bearbetningar samt möjliggöra en differentierad tillgång till data. Man kan se att de olika EU-dataområdes initiativen (dataspaces) är exempel på någonting som faller innanför ramen för ett sånt här införande som bygger på hög grad av gemensamma lösningar, samverkan och delning av data mellan organisationer. Ett nationellt exempel för scenariot skulle kunna vara landets kommuner som samlar in data till dataområdet och tillgängliggör det för varandra samt till myndigheter och andra intressenter.

Införande av dataområdes koncepten kommer driva behovet av samverkan och gemensam infrastruktur, delade komponenter och delat data. Ett dataområde kommer i många fall producera händelser som konsumenter av data kan agera på. Som stöd finns olika lösningsmönster att tillgå vilka rymms innanför denna referensarkitektur. Ett dataområde med gemensamt data behöver regleras centralt utifrån de olika



organisationernas behov samt rättighet till data.



8.1.1.3 Avancerat katalogiserings samt informationssäkerhetsklassificeringsstöd

Arbetet med att identifiera, beskriva och informationssäkerhetsklassificera en organisations datatillgångar kan vara tidskrävande och kräver kunskap. Arbetet sker till största delen manuellt där man delegerat arbetsuppgiften med stöd av rutiner och handledningar. Här finns stora möjligheter att nyttja AI-baserad teknik (kapitel 7, referens 5) eftersom tekniken är gjord för att kunna bearbeta stora datamängder och med stöd av algoritmer och kunskapsmodeller ge ett effektivt stöd för informationsklassificering och dokumentation. Fördelen förutom att man bygger in kunskap och erhåller enhetlighet, är att man kontinuerligt kan utföra arbetsuppgiften med reducerade arbetsinsatser. Stöd av den här typen tas med fördel fram gemensamt och i samverkan mellan organisationer.

8.2 Vidareförädling med stöd av referensarkitekturen

Referensarkitekturen är också tänkt att återanvändas som ett mönster efter ett tillgängliggörande av förädlade, paketerade och beskrivna dataprodukt. En konsument kan återanvända referensarkitekturens mönster för att skapa sin egen händelsekedja anpassade för sina behov och utformning av nya dataprodukt baserade på andra dataprodukt. Observera att bilden visar på en logisk kedja av vidareförädling och inte en implementation, vidareförädlingen kan mycket väl ske inom samma instans av dataplattformen.



8.3 Digg REST API profil

I samverkan med andra myndigheter har Digg tagit fram en REST API profil som gör det möjligt att bygga REST API:er på ett effektivare sätt utifrån bästa praxis och en uppsättning identifierade SKA/BÖR/KAN kravformuleringar. Denna gemensamma profil förenklar processen för både konsumenter och producenter av REST API:er. REST API är ett sätt att bygga ett API baserat på den funktionalitet som finns i HTTP (HyperText Transfer Protocol). REST, står för Representational State Transfer och är i grund och botten en standardiserad mjukvaruarkitekturstil, som består av en speciell typ av API:er, som för branschen är känd och använd. REST förlitar sig på ett tillståndslöst



klient-serverprotokoll och i nästan alla fall kommer det att vara HTTP. Tidigare förlitade sig utvecklare främst på SOAP för att implementera API:et i webbtjänster, men på senare år har REST blivit utvecklarens val på grund av dess enkelhet och skalbarhet.

REST skapades för att behandla objekt på serversidan som resurser som i sin tur kan skapas, uppdateras och raderas. REST kan användas av praktiskt taget alla programmeringsspråk. REST profilen är främst fokuserad till extern åtkomst, där graden av interoperabilitet i den offentliga sektorn behöver stärkas, men det går lika bra att applicera profilen för enbart intern åtkomst. Här är det lämpligt att låta verksamhetsbehovet styra. Beroende på hur REST API:et designats så kan man erhålla hela eller utvalda delar av datamängden. Nedan ges några exempel på hur anrop kan se ut:

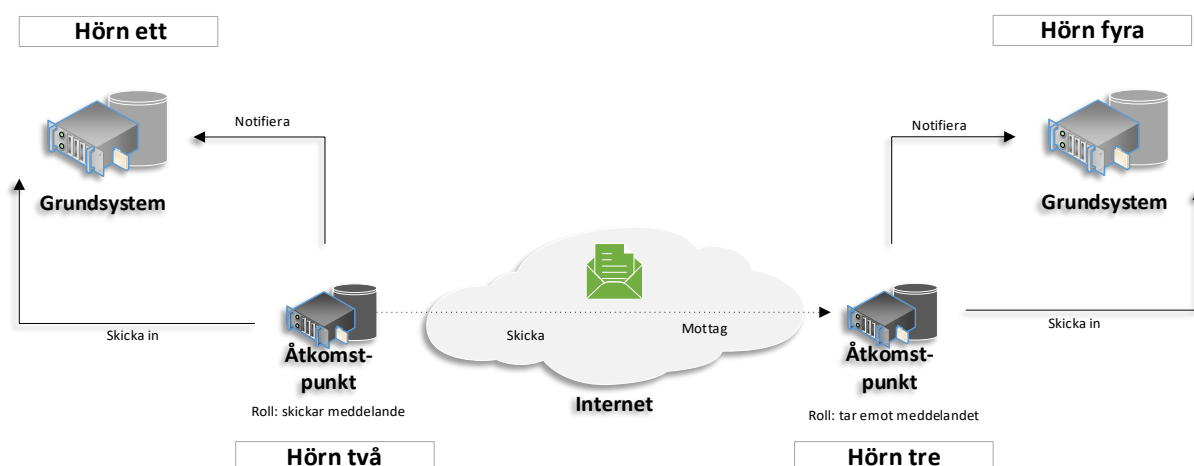
- Lista organisationer
 - GET <https://gw.api.bolagsverket.se/foretagsinformation/v2/organisationer>
- Filtreringsfråga
 - GET <https://gw.api.bolagsverket.se/foretagsinformation/v2/organisationer?filter=bildat%20efter%202020-01-01>
- Hämta information om en enskild organisation
 - GET <https://gw.api.bolagsverket.se/foretagsinformation/v2/organisationer/2021005489>
- Hämta endast enstaka fält för en organisation
 - GET <https://gw.api.bolagsverket.se/foretagsinformation/v2/organisationer/2021005489?fields=postadress,firmateckning>
- Mer information om REST API-profil:
 - [API Playbook](#)
 - <https://dev.dataportal.se/rest-api-profil>

8.4 eDelivery

eDelivery är ett byggblock framtaget av Connecting Europe Facility (CEF) och är en paketering av OASIS standard ebMS3 och AS4 specifikationen. Lite förenklat kan man säga att eDelivery används för att skapa ett säkert distribuerat nätverk för asynkron kommunikation mellan anslutna organisationer/deltagare.



Lösningen är uppbyggd enligt något man kallar en 4-hörns modell där man har accesspunkter en sändande samt en mottagande. Sedan ansluter sig systemen som är godkända medlemmar i den så kallade federationen mot accesspunkterna för att skicka och ta emot data. Inget system är direkt kopplat mot varandra utan dataöverföringen sker genom accesspunkterna. Modellen och de funktioner som eDelivery tillhandahåller borgar för säkert och skalbart datautbyte. Många datautbyten på EU nivå sker med eDelivery.



Jämfört med Digg REST API profil så kan man säga att eDelivery har ett mer utvecklat stöd för att kunna hantera kritiska tillämpningar på ett säkert sätt med inbyggt stöd för detta i standarden. Exempel på funktionalitet är

- Möjlighet att skapa federationer dvs slutna grupper som har behov av att utbyta data med varandra på ett säkert sätt
- Möjlighet att signera och kryptera dataöverföringarna för att säkerställa konfidentialitet och integritet
- Garanterad dataleverans även om någon mottagande del i lösningen skulle drabbas av ett driftavbrott (omsändning)
 - Garanti att man endast erhåller data en gång, dvs man riskerar inte att få samma meddelande flera gånger
 - Funktionalitet som säkerställer att endast avsedd mottagare erhåller data från sändaren i överföringen
 - Hög skalbarhet i modellen

Digg har lanserat en nationell plattform som baseras på eDelivery och stödjer digital samverkan och datautbyte mellan organisationer inom den offentliga sektorn. Det



inkluderar även privata aktörer med uppdrag inom den offentliga sektorn. Mer information om eDelivery och den nationella plattformen finns under länken <https://www.Digg.se/digitala-tjanster/edelivery>

8.5 Data mesh och Data fabric

Data Mesh och Data Fabric är oberoende utvecklade koncept som ofta används omväxlande och kompletterande till varandra. De båda koncepten delar samma mål vilket är att underlätta enkel åtkomst till data. Data fabric är ett framväxande koncept som huvudsakligen är designat som ett teknologiskt mönster, drivet av användning av metadata för att automatisera datamanagement. Data mesh är en arkitektonisk ansats driven av decentraliserad fördelning av ansvar för datahantering och styrning. Data mesh benämner det data som delas som dataprojekter. En dataprojekt är en dokumenterad och konsumtionsfärdig datamängd som underhålls och tillgängliggörs av dataprojektens ägare. Data mesh har i dagsläget ingen eller liten ansats till en mer utvecklad dataplattform det hittar vi däremot i Data fabric, här beskrivs på ett tydligare sätt funktioner som åtkomst och säkerhet, databearbetning, datakatalog mm.

Datafabric är primärt rekommendationer samt metadata för analys och bearbetning. De primära styrkorna är:

- Datakatalog som stöd för förståelse och tillgängliggörande av data
- Metadata, för beskrivande av data, samt möjliggöra för bland annat bearbetning av data
- Dataplattform - infrastruktur för att bereda data för användning.
- Orkestrering av dataprojekter
- Fokuserar på en decentraliserad datalagring inom och utom organisationen organisering av data med en holistisk syn med olika vyer över datatillgångarna.

Datamesh är ingen teknisk lösning utan är i sak konceptuell, definierad av fyra stycken grundpelare så som domändrivet ägandeskap av data, produktsyn på data, självhjälp (self-service) för nyttjande av datatillgångar samt federerad ledning och styrning av data (Federated Governance). De primära styrkorna är:

- Decentraliserat ägandeskap och lagring av data, ägandeskapet finns där kunskapen om data finns
- Dataprojekter
- Samordnad ledning och styrning



- Skalbar och distribuerad arkitektur
- I många fall utvecklings och påbyggnadsbar mot en tidigare dataarkitektur

9. Begreppskatalog

Detta dokument är en bilaga till *Referensarkitekturen för databehandling och utbyte av data mellan myndigheter* och innehåller de centrala begreppen för att beskriva referensarkitekturen.

Begreppskatalogen innehåller begreppet, en kort definition samt källa. I de fall källan är angiven eSam* innebär det att det är en definition framtagen inom arbetet med referensarkitekturen. Dessa definitioner har oftast tagit fram genom att titta på flera olika källor och definitioner för att hitta den definition som är mest relevant för uppdraget.

Begrepp	Beskrivning /definition	Källa
Attributtjänst (stödjer identitet och åtkomst)	En attributtjänst är en tjänst som hanterar och tillhandahåller säkerhetsrelaterad information om användares behörigheter och attribut för att reglera åtkomst till resurser	eSam*
Anonymisering	Anonymisering är processen att omvandla eller ta bort all data som kan identifiera en individ från data, vilket resulterar i att data inte längre är kopplad till någon specifik person. Syftet är att göra data okänslig för personliga identifikationer.	eSam*



API	Application Programming Interface (API) är specifikation av hur olika applikationsprogram kan använda, kommunicera och/eller överföra information mellan varandra.	eSam*
Arkivera	Tillföra data till ett arkiv.	eSam*
Begrepp	Mental föreställning om ett objekt eller om en grupp av objekt med likartade egenskaper. idé; betydelse; föreställning	Rikstermbanken
Begreppskatalog	En samling av definierade termer och begrepp som används inom en organisation eller mellan organisationer för att skapa gemensam förståelse och kommunikation.	eSam*
Begreppsmodell	Grafisk representation av begrepp inom en viss domän och hur de relaterar till varandra	eSam*
Data	Data är råa, opolerade fakta och symboler som representerar	eSam*



	<p>händelser, objekt eller mätvärden.</p>	
Data fabric	<p>Metoder och design för datahantering som syftar till att göra all data i en organisation tillgänglig och sökbar på liknande sätt.</p>	<p>Data fabric IDG:s ordlista</p>
Data lake	<p>En data lake är en stor mängd data som sparas i oförändrad form i ett gemensamt förråd. Alltså utan någon form av normalisering eller annan anpassning. Det kan vara både strukturerade och ostrukturerade data.</p>	<p>IDG</p>
Data mesh	<p>Data mesh är en decentraliserad dataarkitektur som organiserar data efter en specifik affärsdomän</p>	<p>IBM</p>
Data pipeline	<p>Övergripande begrepp för alla processer där data flyttas mellan olika steg exempelvis vid bearbetning av data. Orkestrering är en delmängd av data pipeline.</p>	<p>eSam*</p>
Data som produkt	<p>”Data som produkt” är resultatet av att</p>	<p>eSam*</p>



	tillämpa produkttänkande gentemot datamängder.	
Datadriven	Datadrivet handlar om att använda data för att fatta faktabaserade och proaktiva beslut.	eSam*
Datakatalog	En datakatalog är en produkt som lagrar metadata om en informationsmängd.	eSam*
Datakonsument	Aktör som mottar eller använder tjänst eller information.	eSam*
Datakvalitet	Datakvalitet avser tillståndet för data. Hög datakvalitet innebär att datan är lämpad för dess avsedda användningsområden i drift, beslutsfattande och planering.	eSam*
Datakälla	Ursprunget för datan, var datat har skapats.	eSam*
Datalager	Databas som innehåller aktuella data från ett företags verksamhet, och som är utformad för avancerad analys, men som är åtskilt från det IT-system som	IDG (informationslager)



	företaget använder i verksamheten.	
Datamaskning	Datamaskning är processen att skydda känslig information genom att ersätta eller dölja den med irrelevanta eller pseudonyma data.	eSam*
Datamodell	Grafisk beskrivning av de data som lagras i ett IT-system	eSam*
Datamängd	Identifierbar samling data	ISO 19115-1:2014 Metadata – Part 1
Dataområde	Översättning av EU-kommissionens begrepp data space. Infrastruktur och stöd för att kunna samla och dela data mellan organisationer och andra aktörer.	eSam*
Dataproducent	Aktör som tillhandahåller tjänst eller information.	eSam*
Dataprodukt	Datamängd eller serie av datamängder som erbjuds en konsument.	DIGG
Dataproduktspecifikation	specifikation av en dataprodukt tillsammans med ytterligare information	ISO 19131:2020 Data product specifications



	som möjliggör att densamma kan skapas, distribueras och användas av tredje part	
Dataproduktspecifikation för tillgängliggörande av data	specifikation av en dataproduct tillsammans med ytterligare information som möjliggör att data kan distribueras och användas av konsument	Lantmäteriet
Dataproduktspecifikation för underhåll av data	specifikation av en dataproduct tillsammans med ytterligare information som möjliggör att data kan skapas, underhållas och distribueras till eller inom en dataproductent	Lantmäteriet
Dataproductägare	Roll som äger och ansvar för förvaltningen av en eller flera dataproducter	eSam*
Distribuerad arkitektur	En distribuerad arkitektur innebär att dess IT-system organiseras som en struktur av kommunicerande tjänster	Wikipedia



Domändriven design (Domain-driven design)	tillvägagångssätt för mjukvarudesign	eSam*
Dynamiska data	Data som uppdateras ofta eller i realtid. Text data som genereras av sensorer betraktas vanligtvis som dynamiska data. Dynamiska data ändras över tid men kan även vara oföränderlig under vissa perioder.	Wikipedia
Gallring av data	Gallring av data är processen att medvetet och systematiskt ta bort eller radera onödig, föråldrad eller oanvändbar data från en datalagring för att minska onödig upprepning och frigöra utrymme samt förbättra dataskydd och efterlevnad av regelverk.	eSam*
Grunddata	Uppgifter, inom offentlig förvaltning, som flera aktörer har behov av men som inte beslutats vara nationella grunddata (än).	eSam*
Händelseförmedling	Händelseförmedling är processen att hantera dataflöde genom att	eSam*



	<p>motta, bearbeta och publicera händelser inom och mellan textplattformar och system.</p>	
<p>Identitet och åtkomst (IAM)</p>	<p>System för att reglera vilka användare som ska ha tillgång till en organisations it-nätverk och vilka resurser varje enskild användare ska få komma åt</p>	<p>IDG</p>
<p>Information</p>	<p>Information är resultatet av bearbetning eller tolkning av data. Det är data som har organiserats, analyserats eller förståtts på ett sätt som ger det en meningsfull och användbar betydelse för människor eller system.</p>	<p>eSam*</p>
<p>Informationsförsörjning</p>	<p>Processen från insamling av data till användning av information. Syftet med informationsförsörjning är att ”underlätta för informationssökare att nå den information</p>	<p>Rikstermbanken</p>



	som är relevant för dem”.	
Informationsförvaltare	Den person som utför det dagliga arbetet med en informationsmängd.	eSam*
Informationsmodell	Grafisk beskrivning av de informationsobjekt en viss verksamhet behöver och hur de relaterar till varandra.	eSam*
Informationsobjekt	Sammanhållen och strukturerad informationsmängd som kan hanteras och utväxlas som en enhet mellan olika personer och/eller system.	eSam*
Informationsområde	Större mängder av information som hör ihop versamhetstekniskt.	eSam*
Informationssäkerhetsklassning (informationsklassning)	Innebär att man värderar organisationens informationstillgångar utifrån de interna och externa krav på konfidentialitet, riktighet och tillgänglighet	Informationssäkerhet.se - klassningsmodell
Informationsutbytesmodell	Specialisering av informationsmodell	Ramverk för nationella grunddata - bilaga



	med fokus på information som delas utbyts mellan flera aktörer.	informationsarkitektur 2.0 (Digg.se)
Informationsägare	Den person som äger informationsmängden och är ansvarig för ett datas innehåll samt för att fatta beslut om informationsmängden och de krav som ställs på den.	eSam*
Integrationstjänst	Integrationstjänster, även kallade integrationsadaptorer, är programvarukomponenter eller verktyg som möjliggör kommunikation och datautbyte mellan olika system, applikationer eller plattformar genom att hantera och omvandla data så att de kan samverka effektivt och sömlöst.	eSam*
Kontrakt	Överenskommelse mellan flera parter som reglerar användning och tillvägagångssätt.	eSam*
Masterdata	Datamängd som innehåller en organisations centrala data och som behöver användas av flera	Rikstermbanken



	system, applikationer och processer.	
Metadata	Data om data. Metadata används för att beskriva innehållet och/eller strukturen för en viss datasamling. Det kan göras ur mer än ett perspektiv.	.SE Stiftelsen för internetinfrastruktur
Nationella grunddata	Nationella grunddata är uppgifter, inom offentlig förvaltning, som flera aktörer har behov av, som är viktiga i samhället och som uppfyller överenskomna egenskaper, principer och riktlinjer.	DIGG
Normalisering av data	Syftar till att göra värden jämförbara, till exempel att se till att allt material räknas om till samma måttenheter och att tidsperioder (till exempel räkenskapsår) sammanfaller. Även formatanpassning är en form av normalisering.	IDG
Notifikationstjänster	Används för att skicka meddelanden eller aviseringar om status och händelser	eSam*



	relaterade till datainsamlingen.	
Orkestrering	Automatiserad samordning av program och tjänster för utförande av en uppgift.	IDG
Ostrukturerad data	Data som inte har någon känd struktur.	Rikstermbanken
Personuppgift	Varje upplysning som avser en identifierad eller identifierbar fysisk person. Avgörande är att uppgiften, enskilt eller i kombination med andra uppgifter, kan knytas till en levande person.	eSam*
Pseudonymisering	Pseudonymisering är processen att ersätta identifierande information med en pseudonym, vilket gör det svårt att koppla data till en specifik individ utan ytterligare data. Det används för att skydda integriteten och sekretessen av personlig information.	eSam*
Referensarkitektur	En referensarkitektur är en modell eller ramverk som beskriver en rekommenderad	eSam*



	<p>struktur och organisering för ex system inom ett specifikt område. Den fungerar som en vägledning och standardiseringsgrund.</p>	
Referensdata	<p>Referensdata är data som används för att klassificera eller kategorisera andra data. Vanligtvis är de statiska eller förändras långsamt över tiden. Exempel på referensdata är måttenheter, landskoder, företagskoder, fasta omvandlingsfrekvenser, t.ex. vikt, temperatur och längd.</p>	Wikipedia
Schemaläggning	<p>Schemaläggning är planeringen och tidsstyrningen av aktiviteter och processer för att automatisera och koordinera uppgifter enligt en förutbestämd tidtabell.</p>	eSam*
Skörda data	<p>Process som kopierar datamängder och dess metadata mellan en eller flera datakataloger.</p>	eSam*



Strukturerad data	Data i strukturerat format är organiserad på ett sätt som gör den ingående informationen lätt att identifiera.	Rikstermbanken
Synonym	Ord eller uttryck som har samma betydelse som ett annat ord eller uttryck.	Rikstermbanken
Tillgängliggöra	Ge tillgång till information/data. Oavsett om det görs frivilligt eller på grund av en skyldighet i annan författning.	Innovation genom information, SOU 2020:55
Transformera	Omvandling (transformering) av data från ett format till ett annat format.	eSam*
Validering	Validering av data är processen att kontrollera och säkerställa att data uppfyller de fördefinierade regler och kriterier, och är giltig och användbar för dess avsedda ändamål.	eSam*
Öppna data	Data i öppna format som kan utnyttjas, vidareutnyttjas och delas fritt av vem som	EU, Öppna data-direktivet (skäl 16)

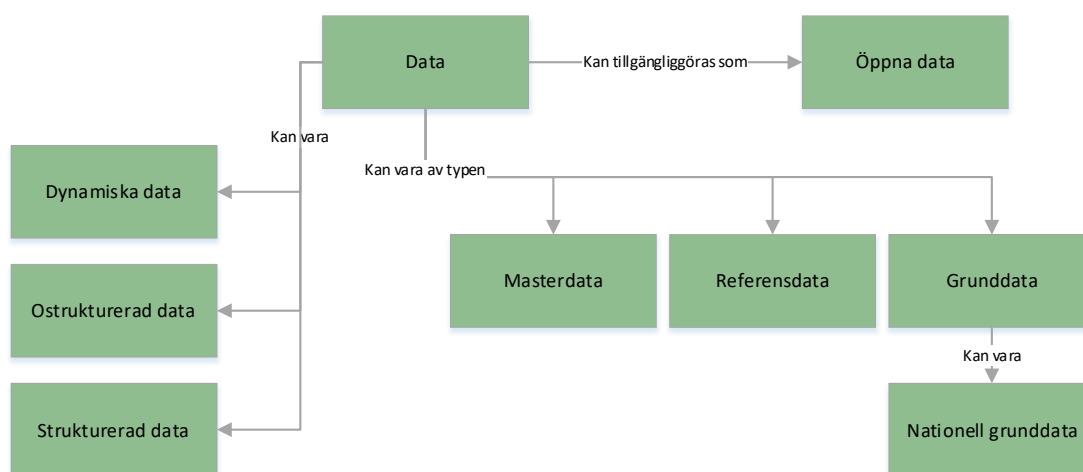


	helst för valfritt ändamål.	
Överföringsformat	En teknisk beskrivning över hur informationen är strukturerad vid ett utbyte.	eSam*

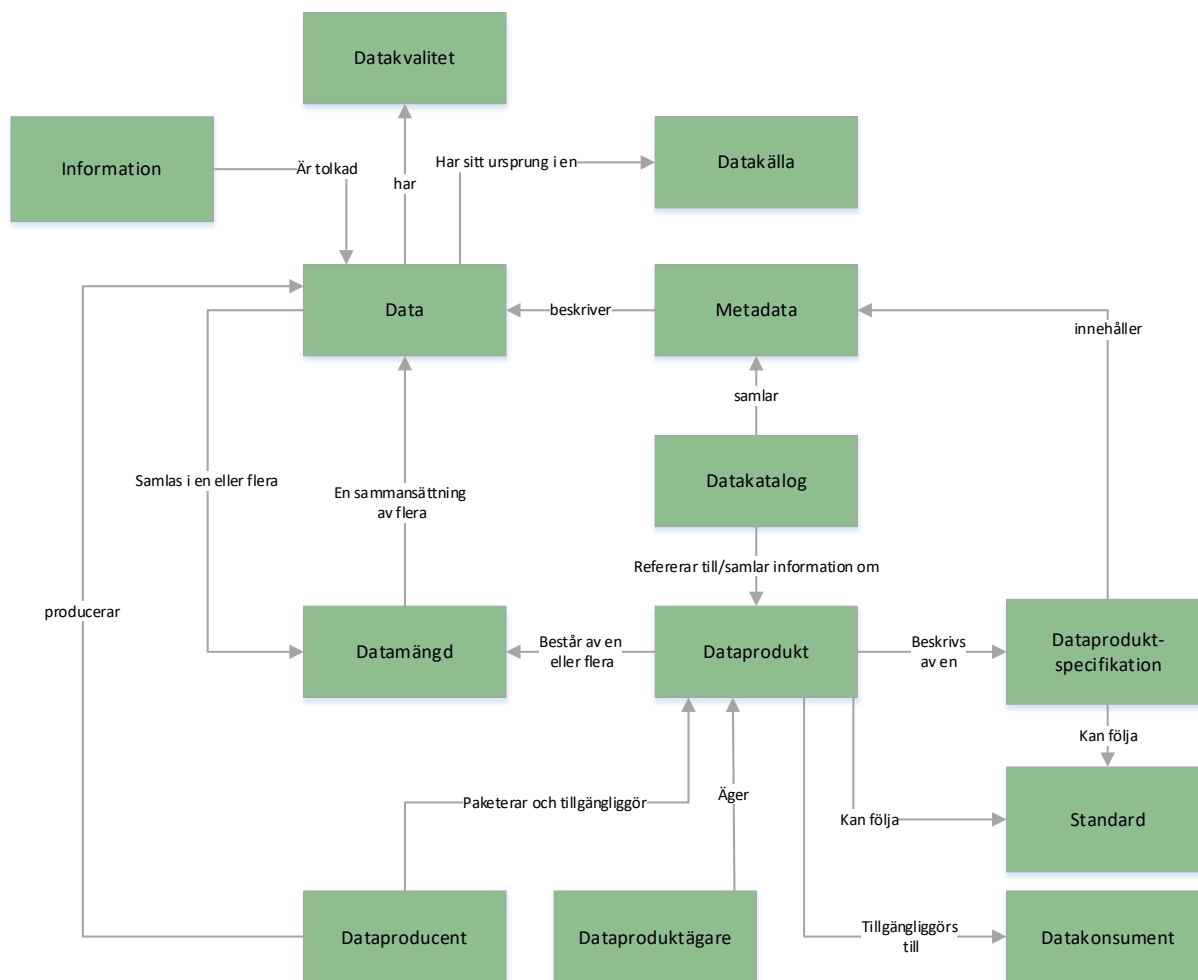
9.1 Begreppsmodeller

Två begreppsmodeller finns framtagna för de delar av referensarkitekturen där de behövs för att konkretisera och beskriva begreppen i relation till varandra. Det är dels en för att förtydliga olika typer/kategorier av data och en modell för att beskriva ett större sammanhang som referensarkitekturen omfattar.

9.1.1 Begreppsmodell för olika typer av data



9.1.2 Begreppsmodell för data och relationer till dataproduct mm



eSam är ett medlemsdrivet program för samverkan mellan myndigheter för att underlätta och påskynda digitaliseringen inom det offentliga. eSam bildades 2015 som en frivillig fortsättning på E-delegationen. En viktig uppgift för eSam är att ta fram stöd och vägledningar som ger förutsättningar för att öka den digitala samverkan inom offentlig förvaltning.

Alla stöddokument finns på [eSamverka.se](https://www.esamverka.se)

I eSam ingår Arbetsförmedlingen, Arbetsmiljöverket, Bolagsverket, Boverket, Centrala Studiestödsnämnden, Domstolsverket, E-hälsomyndigheten, Ekonomistyrningsverket, Finansinspektionen, Folkhälsomyndigheten, Försäkringskassan, Havs- och vattenmyndigheten, Inspektionen för vård och omsorg, Jordbruksverket, Kemikalieinspektionen, Kriminalvården, Kronofogdemyndigheten, Kustbevakningen, Lantmäteriet, Länsstyrelserna, Migrationsverket, Naturvårdsverket, Patent- och Registreringsverket, Pensionsmyndigheten, Riksarkivet, Rättsmedicinalverket, Sida, Skatteverket, Skolverket, Statens institutionsstyrelse, Statens servicecenter, Statens tjänstepensionsverk, Statens veterinärmedicinska anstalt, Statistiska centralbyrån, Tillväxtverket, Trafikverket, Transportstyrelsen, Tullverket och Universitets- och högskolerådet (februari 2024)

